Chalmers Publication Library

(article starts on next page)

Performance Metrics for Systems with Soft-Decision FEC and Probabilistic Shaping

Tsuyoshi Yoshida, *Member, IEEE,* Magnus Karlsson, *Senior Member, IEEE,*
and Erik Agrell, *Senior Member, IEEE*

Abstract—High-throughput optical communication systems utilize binary soft-decision forward error correction (SD-FEC) with bit interleaving over the bit channels. The generalized mutual information (GMI) is an achievable information rate (AIR) in such systems and is known to be a good predictor of the bit error rate after SD-FEC decoding (post-FEC BER) for uniform signaling. However, for probabilistically shaped (nonuniform) signaling, we find that the normalized AIR, defined as the AIR divided by the signal entropy, is less correlated with the post-FEC BER. We show that the information quantity based on the distribution of the single bit signal, and its asymmetric log-likelihood ratio, are better predictors of the post-FEC BER. In simulations over the Gaussian channel, we find that the prediction accuracy, quantified as the peak-to-peak deviation of the post-FEC BER within a set of different modulation formats and distributions, can be improved more than 10 times compared with the normalized AIR.

Index Terms—Bit error rate, bitwise decoding, forward error correction, generalized mutual information, modulation, mutual information, optical fiber communication, probabilistic shaping.

I. INTRODUCTION

Performance metrics are key in the design, evaluation, and comparison of communication schemes. In optical systems, the bit error rate (BER) of the received data is traditionally the most important metric to quantify the performance. In modern optical communications, the BER requirement is typically $< 10^{-15}$. After forward error correction (FEC) was introduced, it became desirable to find other performance metrics than the received BER after FEC decoding, the so-called *post-FEC BER*. This is because measurements or simulations of very low BER after FEC decoding are time-consuming and of less general significance as they only apply to the chosen FEC code. In addition, FEC decoder hardware is usually not available in most laboratories.

For this purpose, the BER before the decoder, the *pre-FEC BER*, or the Q-factor derived from this BER value, became common to predict the post-FEC BER [1]. These metrics work reasonably well with hard-decision FEC decoding and binary

modulation in each dimension, such as on–off keying, binary phase-shift keying, or quaternary phase-shift keying.

The deployment of coherent detection with digital signal processing [2] made modulation formats more complex and diverse, and soft-decision (SD) FEC became widely utilized [3]. In such systems, the pre-FEC BER is insufficient to predict the post-FEC BER, especially for the purpose of comparing different modulation formats. The mutual information (MI) as a metric was introduced in optical communications in [4] and was considered for other channels in, e.g., [5], [6]. The MI is well suited for optical systems using coded modulation with nonbinary FEC codes or bit-interleaved coded modulation (BICM) with iterative demapping [7]. However, even if such schemes work well in theory, practical optical systems often use bitwise receivers, i.e., BICM without iterative demapping, due to the simpler implementation, and the MI does not accurately predict the post-FEC BER of BICM schemes [8]. Instead, Alvarado *et al.* proposed using the generalized MI (GMI), which is an achievable information rate (AIR), as a performance metric in optical communications with binary SD-FEC and bitwise decoding [8].

Shaping is now receiving wide interest in the optical communications community, as a means to close the gap between AIRs with common uniform quadrature amplitude modulation (QAM) formats and the capacity-achieving nonuniform input distribution [9]–[15]. In this paper, we investigate, for the first time, how well certain information-theoretic metrics can predict the post-FEC BER in probabilistically shaped systems with binary FEC codes and bitwise SD decoding. It turns out that the normalized AIR with the signal entropy is not as good metric in this case as with uniform signaling. A metric recently proposed in [15] will be discussed as well. Numerical simulations show that two other metrics, based on the distribution of the single bit stream and the conditional distribution of log-likelihood ratios (LLRs or L-values), have significantly better correlation with the post-FEC BER.

II. SYSTEM MODEL AND PERFORMANCE METRICS

In this section, the system model is described and the considered performance metrics are introduced.

A. System model with nonuniform signaling

Fig. 1 shows the system model under consideration, which is the state-of-the-art configuration described in [9]. The source data stream, consisting of independent, uniform bits, is processed in a distribution matcher and converted into a nonuniform bit sequence A. The binary FEC encoder treats A
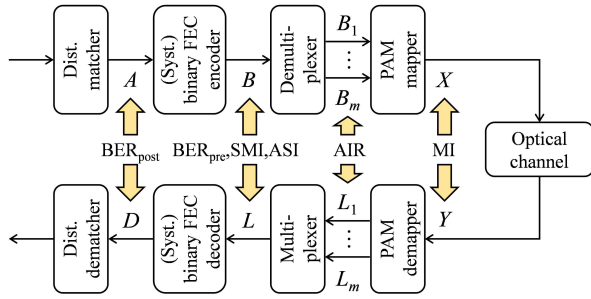
Fig. 1. System model for nonuniform PAM signaling with bit-interleaved binary SD-FEC.

as information bits and creates the bit sequence B by inserting parity bits. When nonuniform (probabilistically shaped) signaling is used, the FEC encoding has to be systematic. Interleaving can be included in the encoder, but the nonuniform distribution must not be changed by encoding or interleaving. The bit sequence B is demultiplexed into m parallel bit tributaries (bit channels) $B_1, B_2, \cdots, B_m$ used for 2^m-PAM mapping to the symbol X. The received PAM symbol Y from the optical channel is demapped to the bit tributaries' LLRs $L_1, L_2, \cdots, L_m$. These are multiplexed into a single LLR sequence L and decoded. The decoder jointly processes all bit tributaries. The decoded bit sequence D is finally recovered by the distribution dematcher. Bit errors have to be eliminated before distribution dematching to avoid error propagation. This system model can apply also to uniform signaling by removing the distribution matcher and the dematcher.

B. Performance metrics for uniform signaling

The BER after FEC decoding, $\mathrm{BER}_{\mathrm{post}}$, and before decoding, $\mathrm{BER}_{\mathrm{pre}}$, are defined as, resp.,

$$\mathrm{BER}_{\mathrm{post}} = \sum_{b \in \{0,1\}} P_{A,D}(b, 1-b), \tag{1}$$

$$\mathrm{BER}_{\mathrm{pre}} = \sum_{b \in \{0,1\}} P_{B,\mathrm{sign}(L)}(b, (-1)^{1-b}), \tag{2}$$

where P denotes the joint probability of the indicated random variables. The normalized GMI, which is a GMI (AIR with uniform signaling) normalized by the sent rate, is defined via the mutual information $I(B_i; L_i)$ between B_i and L_i as

$$I_{\mathrm{n}} = \frac{1}{m} \sum_{i=1}^{m} I(B_i; L_i). \tag{3}$$

where L_i is the *a posteriori* LLR [16, eq. (3.31)], defined as $L_i = \log(L_{i,0}/L_{i,1})$, $L_{i,b} = \sum_{\boldsymbol{b} \in \{0,1\}^m : b_i = b} P_{\boldsymbol{B}}(\boldsymbol{b}) p_{Y|\boldsymbol{B}}(y \mid \boldsymbol{b})$, $\boldsymbol{B} = [B_1, B_2, \cdots B_m]$, and $p_{Y|\boldsymbol{B}}(y \mid \boldsymbol{b})$ is the real-valued channel assumed in the demapper [12, eq. (9)–(11)]. To predict $\mathrm{BER}_{\mathrm{post}}$ in the case of uniform signaling with a binary FEC and bitwise SD decoding, I_{n} is better than $\mathrm{BER}_{\mathrm{pre}}$ [8].

In this paper, the receiver assumes a Gaussian $p_{Y|\boldsymbol{B}}$. In this case, the AIR of a uniform signal set can be estimated from the ensemble of transmitted bits $B_i(j)$ at time index j and the corresponding LLRs $L_i(j)$ over n_{s} symbols as [8, eq. (30)]

$$\hat{I} = 1 - \frac{1}{m n_{\mathrm{s}}} \sum_{i=1}^{m} \sum_{j=1}^{n_{\mathrm{s}}} f(B_i(j), L_i(j)), \tag{4}$$

$$f(\beta, \lambda) = \log_2(1 + \exp(-(-1)^\beta \lambda)). \tag{5}$$

Once ensemble representations of $B_i(j)$ and $L_i(j)$ are provided, these ensembles can be joined into ensemble representations of $B(k)$ and $L(k)$ simply by disregarding the channel index i and using sample index k. The statistical relations between B, L, and their corresponding tributaries B_i and L_i are $P_B(b) = (1/m) \sum_{i=1}^{m} P_{B_i}(b)$, $p_L(l) = (1/m) \sum_{i=1}^{m} p_{L_i}(l)$, and $p_{B,L}(b,l) = (1/m) \sum_{i=1}^{m} p_{B_i,L_i}(b,l)$. This is because $I_{\mathrm{n}} = I_{\mathrm{s}}$ for uniform signaling [17, Theorem 4], where

$$I_{\mathrm{s}} = I(B; L) \tag{6}$$

is called *single-bit MI (SMI)* in this paper.

C. Performance metrics for nonuniform signaling

The metrics for nonuniform signaling, I_{n} and I_{s}, need to be renormalized in order to span the range $[0, 1]$. To this end, (3) and (6) are replaced by the more general expressions

$$I_{\mathrm{n}} = R_{\mathrm{BMD}}/H(\boldsymbol{B}), \tag{7}$$

$$I_{\mathrm{s}} = I(B; L)/H(B), \tag{8}$$

where R_{BMD} in (7) is an AIR for nonuniform signals [9, eq. (67)] [10, eq. (45)], defined as

$$R_{\mathrm{BMD}} = \max \left\{ 0, H(\boldsymbol{B}) - \sum_{i=1}^{m} (H(B_i) - I(B_i; L_i)) \right\}, \tag{9}$$

where $H(\boldsymbol{B}) = -\sum_{\boldsymbol{b} \in \{0,1\}^s} P_{\boldsymbol{B}}(\boldsymbol{b}) \log_2 P_{\boldsymbol{B}}(\boldsymbol{b})$ is the entropy of the s-dimensional discrete variable $\boldsymbol{\mathcal{B}}$.

We further consider the information quantity from an SD-FEC decoder operation, which works based on the hard-decision value $\mathrm{sign}(L)$ and the reliability $|L|$. If $|L|$ is large, there will be less probability to flip the hard-decision value. Thus, the *asymmetric LLR* $L_{\mathrm{a}} = (-1)^B L$, which is included in (5), is an important quantity, as it accounts for both the sign and the magnitude of the LLR. The distribution

$$p_{L_{\mathrm{a}}}(l) = \sum_{b \in \{0,1\}} P_B(b) p_{L|B}((-1)^b l \mid b), \tag{10}$$

where $p_{L|B}(l \mid b) = (1/P_B(b)) \sum_{i=1}^{m} p_{B_i,L_i}(b,l)$, should be as asymmetric as possible for better decoding performance.[1] Then, we can quantify the asymmetry of L_{a} by the *asymmetric information (ASI)*

$$I_{\mathrm{a}} = 1 - h(L_{\mathrm{a}} \mid |L_{\mathrm{a}}|) = 1 + h(|L_{\mathrm{a}}|) - h(L_{\mathrm{a}}), \tag{11}$$

which is defined in $[0, 1]$, where $h(\mathcal{A}) = -\int_{-\infty}^{\infty} p_{\mathcal{A}}(a) \log_2 p_{\mathcal{A}}(a) \mathrm{d}a$ is the differential entropy of the continuous variable $\mathcal{A}$.

All metrics defined for nonuniform signaling can be applied to uniform signaling having symmetric distributions ($P_B(0) = P_B(1) = 1/2$ and $p_{L|B}(l \mid 0) = p_{L|B}(-l \mid 1)$), because it is just a special case of nonuniform signaling. In this case, $\sum_{i=1}^{m} H(B_i) = m$, $H(B) = 1$, $h(L) = 1 + h(|L_{\mathrm{a}}|)$, and $h(L \mid B) = h(L_{\mathrm{a}})$. Therefore, (8) and (11) are identical and $I_{\mathrm{n}} = I_{\mathrm{s}} = I_{\mathrm{a}}$.

---

[1]For uniform signaling, the distribution (10) can be obtained by setting $b = 0$ in the distribution of "symmetrized" LLRs in [17, eq. (19)].
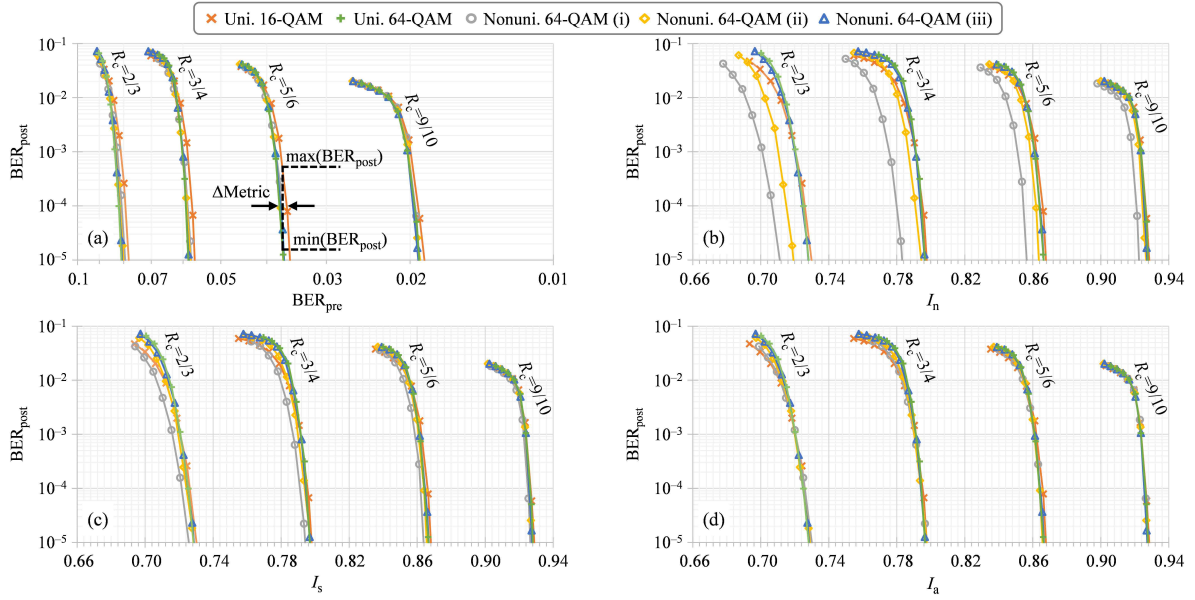
Fig. 2. Simulated post-FEC BER as a function of (a) pre-FEC BER $\mathrm{BER_{pre}}$, (b) normalized AIR I_n, (c) SMI I_s, and (d) ASI I_a.

TABLE I
8-PAM SIGNAL PARAMETERS FOR SIMULATION.

Condition	Uni.	Nonuni. (i)	Nonuni. (ii)	Nonuni. (iii)
$\sum_{i=2}^{m} N_{B_i}/N_\mathrm{s}$	2	1.788	1.934	1.979
$\sum_{i=1}^{m} H(B_i)$	3	2.841	2.960	2.996
$H(B)$	1	0.9831	0.9967	0.9997
$H(\boldsymbol{B})$	3	2.803	2.951	2.995

III. SIMULATIONS

Here we will compare the above performance metrics as indicators of the post-FEC BER. We apply *probabilistic amplitude shaping (PAS)* [9], which is the current state-of-the-art shaping scheme in optical communications [12]. Three nonuniform distributions based on 64-QAM were simulated by applying probabilistic shaping to Gray-coded 8-PAM in each dimension. To generate these distributions, a constant-composite distribution matcher was employed [11] with output block length $N_\mathrm{s} = 1024$.[2] The probabilities used are the same as in [13, Table I (a)–(c)], and some measures of their nonuniformity are indicated in Table I. B_1 is the sign bit, carrying the uniform information bits of the PAM symbol or FEC parity bits, and $B_2, \cdots B_m$ are amplitude bits. The number of input bits of bit tributary B_i to the distribution matcher per block is denoted by N_{B_i}. As benchmarks, uniform signaling using 16-QAM and 64-QAM was also simulated. As SD-FEC codes, we utilized the DVB-S2 binary low-density parity check codes [18], having a codeword length of 64800. The bit mapper was (3,2,1) for 8-PAM, which is optimum [9, Tab. V], and (2,1) for 4-PAM. The examined code rates R_c were 2/3, 3/4, 5/6, and 9/10. The results were averaged over 500 codewords in each case, and a couple of independent simulations were performed to verify that 500 codewords is indeed sufficient for reliable statistics. The optical channel was assumed to be Gaussian with the signal-to-noise ratio varied with 0.1 dB granularity. The FEC decoder's internal calculations used floating-point precision with 20 decoder

iterations.

The metrics I_n, I_s, and I_a were estimated from (7), (8), and (11), resp. The corresponding differential entropies were calculated as in [19, Sec. 8.3] from discretized versions of the LLRs L_i, L, and L_a, whose distributions were estimated using histograms having 2^5 levels with optimized steps. The estimated error introduced by this discretization is only $< 10^{-3}$ in each information quantity and hence negligible.

The results are shown in Fig. 2, where it is desirable that all curves for the same code rate should be close to each other, if the quantity on the horizontal axis is to be useful to predict the post-FEC BER. For most of the metrics, the difference due to modulation format and nonuniformity tends to increase at lower code rates R_c. According to Fig. 2(a), $\mathrm{BER_{pre}}$ works surprisingly well for these signal sets and will provide guidelines for real system performance. Fig. 2(b) shows that the normalized AIR I_n is not a good metric for nonuniform signaling. The required I_n for nonuniform signaling is clearly less than that for uniform signaling to achieve the same $\mathrm{BER_{post}}$ value such as 10^{-4}. Fig. 2(c) indicates that the SMI I_s is better correlated with $\mathrm{BER_{post}}$ for nonuniform signaling than I_n. This is a benefit of changing the treatment from parallel bit channels into a single bit sequence, which is what the FEC encoder and decoder will see. Fig. 2(d) shows that the ASI performs better than the other metrics by using the asymmetric distribution of L_a.

Intuitively, the lower accuracy of I_n can be understood because (7) does not properly take FEC coding in the PAS scheme into account. The maximum FEC code rate for error-free transmission is $R_{\mathrm{c,max}} = 1 - (H(\boldsymbol{B}) - R_\mathrm{BMD})/m$, which is the code rate at which the information rate $H(\boldsymbol{B}) - (1 - R_\mathrm{c})m$ equals the AIR R_BMD. I_n tends to be lower than $R_{\mathrm{c,max}}$ and the difference becomes larger for a strongly shaped signal. The metric $NGMI = 1 - (H(\boldsymbol{B}) - GMI)/m$ [15], where GMI can be defined as $H(\boldsymbol{B}) - \sum_{i=1}^{m} H(B_i \mid Y) =$

---

[2]A smaller N_s, such as $N_\mathrm{s} = 12$ in [14], gives simple implementation but reduces the rate. The rate loss is limited to $< 0.55\%$ in the cases tested here.
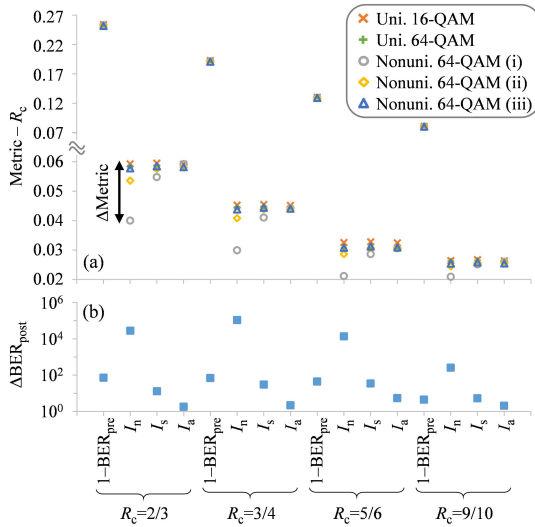
Fig. 3. Comparison of prediction accuracy for (a) the metrics at $\text{BER}_{\text{post}} = 10^{-4}$, and (b) $\Delta\text{BER}_{\text{post}}$ around $\text{BER}_{\text{post}} = 10^{-4}$ among the studied cases.

$H(\boldsymbol{B}) - \sum_{i=1}^{m} H(B_i \mid L_i)$, [3] is identical to $R_{\text{c,max}}$ in the regime where $R_{\text{BMD}} > 0$. On the other hand, one can show that $I_a = NGMI$ by considering [16, eqs. (3.67), (4.81)] and [9, Sec. VI-B]. Note that (4) represents the estimates of $NGMI$ and I_a by Monte-Carlo integration for nonuniform signaling.

Fig. 3 summarizes the prediction accuracy at $\text{BER}_{\text{post}} = 10^{-4}$. Fig. 3(a) compares the information quantities $1 - \text{BER}_{\text{pre}}$, I_n, I_s, and I_a. Relatively large deviations are mainly seen for nonuniform signaling (especially case (i), which is the most nonuniform) with the AIR metric. When considering all signals at the same BER, the maximum metric variations ΔMetric (see Fig. 2(a)) in Fig. 3(a) are 0.0026, 0.019, 0.0048, and 0.0020 for $1 - \text{BER}_{\text{pre}}$, I_n, I_s, and I_a, resp. Note that the prediction accuracy for $1 - \text{BER}_{\text{pre}}$, I_n, and I_s tends to decrease for low R_c or strong shaping, whereas the accuracy for I_a does not. In comparisons of coding, shaping, and modulation schemes, a spectral efficiency difference of several percent tends to be relevant [12], so a prediction variation in I_n (0.019 in this case) may potentially affect such a comparison. The accuracy depends on not only ΔMetric but also the steepness of the relationship of the metric with BER_{post}, so in Fig. 3(b) we show $\Delta\text{BER}_{\text{post}} = \max(\text{BER}_{\text{post}})/\min(\text{BER}_{\text{post}})$ around the specific metric value that gives $\text{BER}_{\text{post}} = 10^{-4}$ on average. Thus $\Delta\text{BER}_{\text{post}}$ is a measure of the correlation between BER_{post} and the chosen metric. The maximum $\Delta\text{BER}_{\text{post}}$ is $10^{1.9}$, $10^{5.0}$, $10^{1.5}$, and $10^{0.73}$ for BER_{pre}, I_n, I_s, and I_a, resp. I_a is clearly the best metric having more than 10 times better accuracy than BER_{pre} and I_n. We tested stronger shaping cases like in [12], and again found that $\Delta\text{BER}_{\text{post}}$ was consistent around 10^{-4} for BER_{pre} and I_a and varied much more ($\geq 10^9$) for I_n and I_s.

IV. CONCLUSIONS

We compared performance metrics to predict the post-FEC BER with binary SD-FEC in systems with probabilistic shaping. The normalized AIR has a large prediction variation

---

of 10^5 around $\text{BER}_{\text{post}} = 10^{-4}$ with a low-density parity check code for nonuniform signaling, because it does not properly take the FEC coding into account. By employing the SMI and ASI metrics, the variation can be reduced to 30 and 5, resp. While the prediction variation of normalized AIR and SMI increases in the case of lower R_c or stronger shaping, the variation by ASI does not seem to depend thereon. Thus, once we know the ASI limit for a certain FEC code, we can infer the practical achievable rate by only calculating the ASI without FEC decoding.

REFERENCES

[1] ITU-T, "Forward error correction for submarine systems," 2000. [Online]. Available: www.itu.int/rec/T-REC-G.975

[2] K. Roberts *et al.*, "Performance of dual-polarization QPSK for optical transport systems," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3546–3559, Aug. 2009.

[3] F. Chang *et al.*, "Forward error correction for 100 G transport networks," *IEEE Commun. Mag.*, vol. 48, no. 3, pp. S48–S55, Mar. 2010.

[4] A. Leven *et al.*, "Estimation of soft FEC performance in optical transmission experiments," *IEEE Photon. Technol. Lett.*, vol. 23, no. 20, pp. 1547–1549, Oct. 2011.

[5] L. Wan, *et al.*, "A fading-insensitive performance metric for a unified link quality model," in *IEEE Wireless Communications and Networking Conf.*, Las Vegas, NV, Apr. 2006.

[6] M. Franceschini *et al.*, "Does the performance of LDPC codes depend on the channel?" *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 2129–2132, Dec. 2006.

[7] L. Schmalen *et al.*, "Performance prediction of nonbinary forward error correction in optical transmission experiments," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 1015–1027, Feb. 2017.

[8] A. Alvarado *et al.*, "Replacing the soft FEC limit paradigm in the design of optical communication systems," *J. Lightw. Technol.*, vol. 33, no. 20, pp. 4338–4352, Oct. 2015.

[9] G. Böcherer *et al.*, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.

[10] G. Böcherer, "Achievable rates for probabilistic shaping," [Online]. Available: www.arxiv.org/pdf/1707.01134v3

[11] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.

[12] F. Buchali *et al.*, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: an experimental demonstration," *IEEE J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.

[13] T. Fehenberger, *et al.*, "Sensitivity gains by mismatched probabilistic shaping for optical communication systems," *IEEE Photon. Technol. Lett.*, vol. 28, no. 7, pp. 786–789, Apr. 2016.

[14] J. Cho *et al.*, "Low-complexity shaping for enhanced nonlinearity tolerance," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Düsseldorf, Germany, Sep. 2016, p. W.1.C.2.

[15] J. Cho *et al.*, "Normalized generalized mutual information as a forward error correction threshold for probabilistically shaped QAM," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Göteborg, Sweden, Sep. 2017, p. M.2.D.2.

[16] L. Szczecinski and A. Alvarado, *Bit-Interleaved Coded Modulation: Fundamentals, Analysis, and Design*. New York, NY, USA: Wiley, 2015.

[17] M. Ivanov *et al.*, "On the information loss of the max-log approximation in BICM systems," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3011–3025, June 2016.

[18] European Telecommunications Standards Institute, "Digital Video Broadcasting (DVB)...; Part 1 (DVB-S2)," ETSI Standard EN 302 307-1 V1.4.1, Nov. 2014. [Online]. Available: www.dvb.org/standards

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

---

[3]$H(\boldsymbol{B})$ minus the conditional entropy due to suboptimal reception is a natural extension of the GMI [8, eq. (21)] to nonuniform signaling. Several kinds of rates in PAS are discussed in [10], too.