Probabilistic Amplitude Shaping with Hard Decision Decoding and Staircase Codes

N.B. When citing this work, cite the original published paper.

(article starts on next page)

Probabilistic Amplitude Shaping with Hard Decision Decoding and Staircase Codes

Alireza Sheikh *Student Member, IEEE*, Alexandre Graell i Amat, *Senior Member, IEEE*,
Gianluigi Liva, *Senior Member, IEEE*, and Fabian Steiner, *Student Member, IEEE*

Abstract—We consider probabilistic amplitude shaping (PAS) as a means of increasing the spectral efficiency of fiber-optic communication systems. In contrast to previous works in the literature, we consider probabilistic shaping with hard decision decoding (HDD). In particular, we apply the PAS recently introduced by Böcherer *et al.* to a coded modulation (CM) scheme with bit-wise HDD that uses a staircase code as the forward error correction code. We show that the CM scheme with PAS and staircase codes yields significant gains in spectral efficiency with respect to the baseline scheme using a staircase code and a standard constellation with uniformly distributed signal points. Using a single staircase code, the proposed scheme achieves performance within 0.57–1.44 **dB** of the corresponding achievable information rate for a wide range of spectral efficiencies.

Index Terms—Coded modulation, error correcting codes, hard decision decoding, probabilistic shaping, optical networks, signal shaping, staircase codes.

I. INTRODUCTION

TO MEET the ever increasing data rate demands, next-generation fiber-optic communication systems need to use the available spectrum more efficiently. Therefore, there is currently a great focus in the research community in increasing the spectral efficiency of these systems. In this regard, forward error correction (FEC) in combination with higher order modulation, a scheme commonly referred to as coded modulation (CM), has become a key part of fiber-optic systems.

Traditionally used signal constellations, such as amplitude shift keying (ASK) and quadrature amplitude modulation (QAM), are characterized by equidistant signal points and uniform signaling, i.e., each signal point is transmitted with the same probability. Unfortunately, such constellations result in a gap to the Shannon limit (1.53 dB for an additive white Gaussian noise (AWGN) channel in the high signal-to-noise ratio (SNR) regime). To close this gap and to increase the spectral efficiency, signal shaping may be applied. There are

two main classes of signal shaping, geometric shaping [2]–[6], and probabilistic shaping [7]–[9]. In geometric shaping, the constellation points are arranged in the complex plane in a nonequidistant manner to mimic the capacity achieving distribution. Probabilistic shaping, on the other hand, starts with a constellation with equidistant signal points, e.g., ASK or QAM, and assigns different probabilities to different constellation points.

Both geometric and probabilistic shaping have been considered for fiber-optic communications as a means to increase the spectral efficiency, showing significant gains with respect to conventional constellations [6], [10]–[13]. A significant advantage of probabilistic shaping is that it builds up on off-the-shelf constellations, hence incurring no additional complexity in system design and implementation compared to geometric shaping. In [9], a new CM scheme using probabilistic shaping was proposed. The proposed scheme, dubbed probabilistic amplitude shaping (PAS), was shown in [9] to achieve performance within 1.1 dB from the capacity of the AWGN channel for a wide range of spectral efficiencies using off-the-shelf low-density parity-check (LDPC) codes and soft decision decoding (SDD). More recently, this scheme has been considered for fiber-optic communications in [14], [15].

All these previous works consider shaping in combination with SDD. However, while FEC with SDD yields very large net coding gains, it entails a high decoding complexity, which translates into a large chip area and power consumption [16]. To reduce the decoding complexity, hard decision decoding (HDD) is an appealing alternative. Hard decision decoders consume significantly less power than soft-decision decoders [16]. Despite the rise of FEC with SDD, such as LDPC codes and spatially-coupled LDPC codes [17], the interest in FEC schemes with HDD has experienced a revived attention in the fiber-optic communications research community in the past few years, thanks to the appearance of very powerful FEC-HDD schemes. Constructions such as staircase codes [18]–[20], braided codes [21], and other generalized product codes [22], [23] yield very large net coding gains yet with much lower decoding complexity than FEC-SDD schemes.

In this paper, we consider signal shaping as a means of increasing the spectral efficiency of the fiber-optic system without increasing the launch power. However, contrary to the previous literature on signal shaping, which focuses on SDD, we consider signal shaping with (bit-wise) HDD. In particular, we apply the PAS scheme proposed in [9] to a CM scheme with staircase codes and HDD. Similar to [9], using ASK modulation as the underlying signal constellation, we optimize

shaping such that a given achievable information rate, with bit-wise HDD in our case, is maximized. This paper extends the work presented in [1], where the application of PAS to bit-wise HDD was originally proposed, showing remarkable coding gains. In [1] the design was based on the maximization of the generalized mutual information. Here, we aim at optimizing the input distribution for a rate that is achievable by the characteristic decoding strategy of PAS, derived in [24]. We then discuss the adaptation of PAS to the use of staircase codes and the optimization of the code parameters. Furthermore, we address the selection of the operating point for finite frame length. Finally, we show through simulation results that the probabilistic shaping CM scheme with staircase codes and HDD achieves up to 2.88 dB gain improvement with respect to the system using a staircase code and a conventional, uniform signal constellation.

The remainder of the paper is organized as follows. In Section II, the achievable information rate with bit-wise HDD is discussed. In Section III, the PAS scheme with HDD is introduced. The CM scheme with PAS and staircase codes is further elaborated in Sections IV and V. Finally, simulation results are given in Section VI and some conclusions are drawn in Section VII.

Notation: The following notation is used throughout the paper. We define the sets $\mathbb{N} \triangleq \{1, 2, \ldots\}$ and $\mathbb{N}_0 \triangleq \{0, 1, \ldots\}$. We denote by $P_X(\cdot)$ the probability mass function (pmf) and by $p_X(\cdot)$ the probability density function (pdf) of a random variable (RV) X. We use boldface letters to denote vectors and matrices, e.g., $\boldsymbol{x}$ and $\boldsymbol{X}$, respectively. Expectation with respect to the pmf of RV X is denoted by $\mathbb{E}_X(\cdot)$. $\mathsf{H}(X)$ and $\mathsf{I}(X; Y)$ stand for entropy of the RV X and mutual information between RVs X and Y, respectively.

II. ACHIEVABLE INFORMATION RATE WITH HARD DECISION DECODING AND BIT-WISE DECODING

We consider a discrete-time AWGN channel[1] with input-output relation at time instant i

$$Y_i = \Delta X_i + Z_i \qquad i = 1, 2, \ldots, n$$

where n is the number of channel uses (i.e., the block length), X_i is the input of the channel, Y_i is its output, Δ is a scaling constant, and $\{Z_i\}$ are independent and identically distributed (i.i.d.) Gaussian RVs with zero mean and unit variance. The scaling parameter Δ is defined to attain an average transmit power P according to

$$\mathbb{E}\left[(\Delta X)^2\right] = \mathsf{P}.$$

According to the definitions above, the SNR is given by $\mathsf{SNR} = \mathsf{P}$. We consider a block-wise transmission system where $\boldsymbol{u}$ denotes the transmitted information block and $\hat{\boldsymbol{u}}$ denotes the decoded information block.

For simplicity, for the analysis and the design of the PAS scheme in this and next section, we consider ASK modulation as the underlying modulation, i.e., the channel input alphabet

[1]The AWGN channel is an accurate model for long-haul coherent fiber-optic communications when the fiber-optic channel is dominated by amplified spontaneous emission noise [25].

is given by $\mathcal{X} \triangleq \{-2^m + 1, \ldots, -1, 1, \ldots, 2^m - 1\}$, where m is the number of bits per symbol, and $M = 2^m$ is the number of signal points. However, the PAS scheme directly extends to square QAM constellations, which can be seen as the Cartesian product of two ASK constellations (see also Section V-B). In Section VI we give results for QAM constellations.

For a given distribution P_X of the channel input, the mutual information (MI) between the channel input X and channel output Y

$$\mathsf{I}(X; Y) \triangleq \mathbb{E}\left[\log_2\left(\frac{p_{Y|X}(Y|X)}{\sum_{x' \in \mathcal{X}} p_{Y|X}(Y|x') P_X(x')}\right)\right] \quad (1)$$

determines the upper limit on the achievable rate. A rate R is achievable, i.e., the probability of error can be made arbitrarily small in the limit of infinitely large block length n, if $\mathsf{R} < \mathsf{I}(X; Y)$. The ultimate limit is given by the channel capacity, obtained maximizing the MI over all possible input distributions, $\mathsf{C} \triangleq \sup_{P_X} \mathsf{I}(X; Y)$.

In this paper, we consider the PAS scheme of [9] with a binary code for transmission and bit-wise HDD (i.e., bit-wise Hamming metric decoding) at the receiver side. Denote by $\hat{X} \in \mathcal{X}$ the RV associated with the detector output (hard decision). An achievable rate for the PAS scheme can be computed by resorting to the approach introduced in [24], yielding

$$\mathsf{R}_{\mathsf{HDD}} = \sup_{s > 0}\left[\mathsf{H}(X) + \mathbb{E}\left[\log_2 \frac{q(X, \hat{X})^s}{\sum_{x' \in \mathcal{X}} q(x', \hat{X})^s}\right]\right]^+ \quad (2)$$

where $(a)^+ = \max(0, a)$, s is the optimization parameter, and $q(X, \hat{X})$ is the (mismatched) decoding metric [26], [27]. For HDD, the decoding metric is the bit-wise Hamming metric, which has the equivalent form

$$q(x, \hat{x}) = \varepsilon^{\mathsf{d}_{\mathsf{H}}(\mathsf{L}(x), \mathsf{L}(\hat{x}))}$$

where ε is an arbitrary constant in $(0, 1)$, $\mathsf{L}(x)$ is the m-bit labeling associated with constellation symbol x, and $\mathsf{d}_{\mathsf{H}}(\mathsf{L}(x), \mathsf{L}(\hat{x}))$ is the Hamming distance between the binary labelings of x and $\hat{x}$. Here, we consider the binary reflected Gray code (BRGC) labeling [28]. By optimizing over s, it is possible to show that (2) reduces to [24, Sec. 6.4]

$$\mathsf{R}_{\mathsf{HDD}} = [\mathsf{H}(X) - m\mathsf{H}_b(p)]^+$$

where $\mathsf{H}_b(p) = -p \log_2 p - (1 - p) \log_2(1 - p)$ is the binary entropy function, and p is the (raw) bit error probability at the output of the hard detector (i.e., the pre-FEC bit error rate). In Fig. 1, we plot $\mathsf{R}_{\mathsf{HDD}}$ for a 16-ASK constellation with uniform distribution. For the sake of comparison, we also plot the capacity of the unconstrained-input AWGN channel.

III. PROBABILISTICALLY-SHAPED ASK CONSTELLATION WITH HARD DETECTION

In this section, we use probabilistic shaping to boost the achievable rate compared to the case where constellation
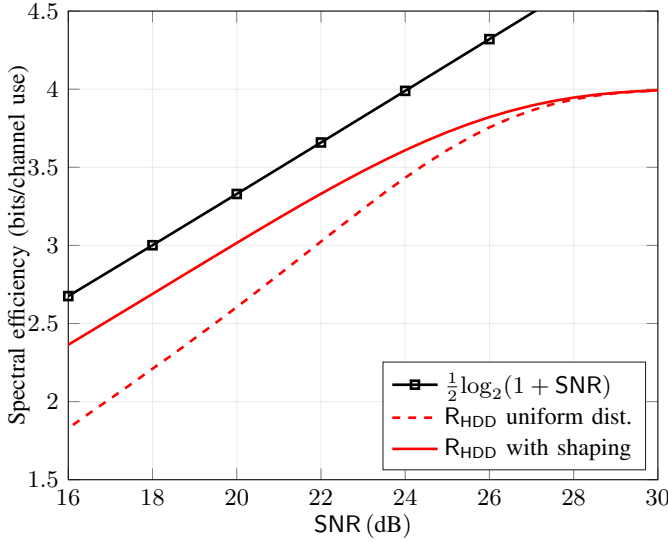
Figure 1. Achievable rates with bit-wise HDD (R_{HDD} given by (2)) with and without shaping for a 16-ASK constellation (for the uniform case, see also [29], [30]). The unconstrained-input AWGN channel capacity is also shown as reference.
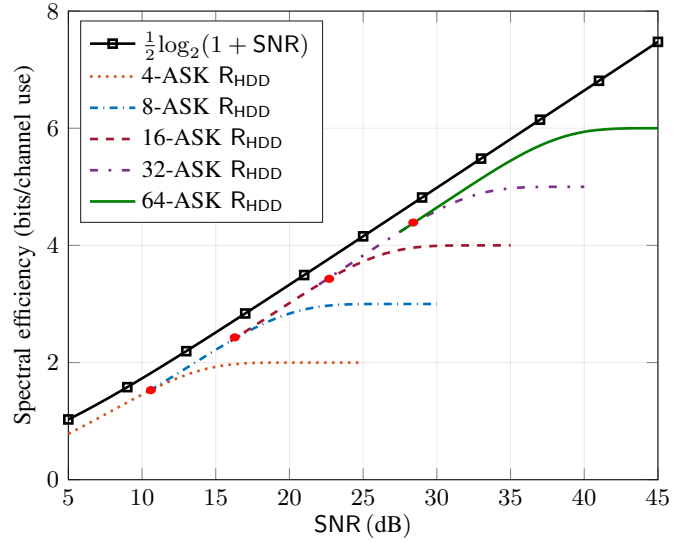

Figure 2. Achievable rates of the probabilistic amplitude shaping with bit-wise HDD. The red points show the SNR values after which a higher order modulation should be used. Below the red dots, the achievable rate curve for a given ASK constellation virtually overlaps with that of the lower-order constellation.

Table I
SHAPING GAIN FOR DIFFERENT SPECTRAL EFFICIENCIES

Modulation	SE (bit/channel use)	Shaping gain (dB)
4-ASK	1	0.78
8-ASK	2	1.56
16-ASK	3	1.98
32-ASK	4	2.22
64-ASK	5	2.37

points are drawn uniformly from $\mathcal{X}$. Similar to [9], we consider the Maxwell-Boltzmann distribution for the channel input X,

$$P_X^\lambda (x) = \frac{\exp\left(-\lambda x^2\right)}{\sum\limits_{\tilde{x}\in\mathcal{X}} \exp\left(-\lambda \tilde{x}^2\right)}. \tag{3}$$

For each SNR, we select λ such that the achievable rate is maximized, i.e.,

$$\lambda^* = \underset{\lambda}{\mathrm{argmax}} \ R_{HDD}. \tag{4}$$

Unlike [9], in this paper we consider HDD at the receiver. In particular, we consider a symbol-wise maximum a-posteriori (MAP) detector that outputs

$$\hat{x} = \underset{x\in\mathcal{X}}{\mathrm{argmax}} \ p_{Y|X}(y|x)P_X(x)$$

yielding the conditional pmf $P_{\hat{X}|X}$ to be used, jointly with P_X, to compute (2).

In Fig. 1, we depict the achievable rate in (2) for the probabilistically-shaped scheme according to (3)–(4) for an underlying 16-ASK. As can be seen, the achievable information rate for the shaped constellation is significantly better than that of the uniform constellation. For an spectral efficiency of 3 bits per channel use, the shaping gain is around 2 dB. Table I summarizes the shaping gain for different modulation orders and spectral efficiencies. In Fig. 2, we depict the achievable rate R_{HDD} of the shaped constellation for different modulation orders. The figure shows that using the shaping described in this section, the CM system with HDD can operate at a

roughly constant gap (around 1.9 dB) to the capacity of the soft-decision AWGN channel.

IV. CODED MODULATION SCHEME WITH PAS AND STAIRCASE CODES

In this section, we apply PAS [9] to HDD and staircase codes. The CM scheme using PAS is depicted in Fig. 3. As observed in [9], $P_X^{\lambda^*}$ is symmetric, i.e., $P_X^{\lambda^*}(x) = P_X^{\lambda^*}(-x)$ for $x \in \mathcal{X}$. Therefore, the random variable X can be factorized as

$$X = A \cdot S \tag{5}$$

where $A \triangleq |X|$ and $S \triangleq \mathsf{sign}(X)$ are the absolute value and sign of the RV X, respectively. It results that S is uniformly distributed,

$$P_S(1) = P_S(-1) = \frac{1}{2} \tag{6}$$

while the distribution of A satisfies

$$P_A^{\lambda^*}(a) = 2P_X^{\lambda^*}(a) \tag{7}$$

where $a \in \mathcal{A} \triangleq \{1, ..., 2^m - 1\}$.

The idea of the PAS [9] is then to split the information sequence $\boldsymbol{u}$ into two sequences $\boldsymbol{u}^{\mathsf{s}}$ and $\boldsymbol{u}^{\mathsf{a}}$. The sequence $\boldsymbol{u}^{\mathsf{a}}$ is used to generate a sequence of amplitudes $a_1, \ldots, a_n$ with the desired distribution. The binary image of the amplitudes and the remaining information bits, i.e., those belonging to $\boldsymbol{u}^{\mathsf{s}}$, are then encoded using a binary code with a systematic encoder. The parity bits generated by the systematic encoder and $\boldsymbol{u}^{\mathsf{s}}$ are used to generate n sign labels $s_1, \ldots, s_n$. Assuming uniform distribution of the information bits and since the parity bits at the output of the encoder tend to be uniformly distributed as well, the sign labels closely mimic the desired (uniform) distribution (6). A detailed description of the CM scheme (see Fig. 3) is provided in the following.
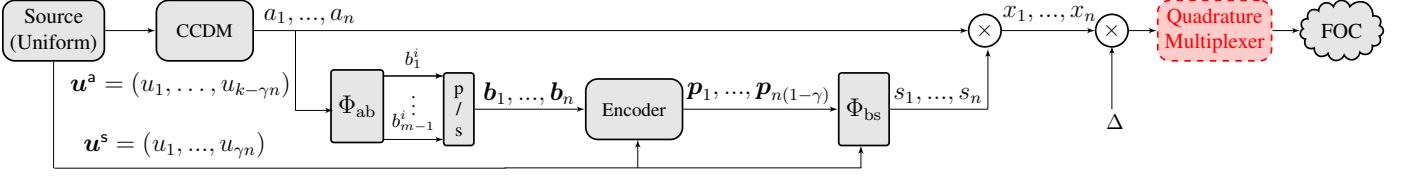
Figure 3. Block diagram of the probabilistically amplitude shaped CM scheme.

A. Amplitude Shaping

Let $\boldsymbol{u} = (u_1, ..., u_k)$ be the information sequence of length k bits, $u_i \in \{0, 1\}$. Information bits are modeled as uniformly-distributed i.i.d. RVs. The vector $\boldsymbol{u}$ is split into two vectors $\boldsymbol{u}^{\mathsf{s}}$ and $\boldsymbol{u}^{\mathsf{a}}$, of lengths γn and $k - \gamma n$, respectively, where γ is a tuning parameter whose meaning will become clear later and it is assumed that $\gamma n \in \mathbb{N}_0$. Vector $\boldsymbol{u}^{\mathsf{a}}$ is used to generate a sequence of amplitudes $\boldsymbol{a} = (a_1, ..., a_n)$ with distribution $P_A^{\lambda^*}$ through a shaping block. The binary interface which generates the sequence of amplitudes with a given distribution (in our case $P_A^{\lambda^*}$) from a uniformly distributed input is called the *distribution matcher*. The distribution matchers proposed in the literature can be categorized in two groups, variable-length [31]–[34] and fixed-length [35] distribution matchers. To limit error propagation, we consider fixed-length distribution matching. In particular, we use the *constant composition distribution matching* (CCDM) method proposed in [35]. This distribution matching algorithm uses arithmetic coding to generate the output amplitudes in an online fashion. Hence, there is no need for a lookup table as required by other algorithms. We refer the interested reader to [9, Sec. V] and [35] for more details. The rate of the shaping block is $(k - \gamma n)/n$ and it approaches $\mathsf{H}(A)$ for large n. We remark that the degree of parallelism in implementing the CCDM can be increased using the product distribution matcher [36] or the streaming distribution matcher [37].

B. Amplitude-to-Bit Mapping

To reduce the number of bit errors associated with a symbol error, we consider the BRGC labeling. In the mapping part, for an M-ASK modulation with $m = \log_2 M$ bits per symbol, the sequence of amplitudes $a_1, ..., a_n$ is transformed into a sequence of bits using the mapper Φ_{ab}. We label each of the amplitudes a_i with $m - 1$ bits using the BRGC labeling to construct a binary string $\boldsymbol{b}_i \triangleq \boldsymbol{b}(a_i) = (b_1^i, \cdots, b_{m-1}^i)$. The sequence $\boldsymbol{b} = (\boldsymbol{b}_1, ..., \boldsymbol{b}_n)$ is of length $\ell = n(m-1)$ bits.

C. Encoding

The sequences $\boldsymbol{b}$ and $\boldsymbol{u}^{\mathsf{s}}$ are multiplexed and encoded by a binary linear block code with a systematic encoder. As binary linear block codes, we use staircase codes with Bose-Chaudhuri-Hocquenghem (BCH) codes as component codes. In particular, we consider $(n_{\mathsf{c}}, k_{\mathsf{c}})$ systematic BCH component codes of code length n_{c} and dimension k_{c}. Staircase codes can be decoded iteratively using bounded-distance decoding (BDD) for the decoding of the component codes and can provide a 0.42 dB coding gain over the best known code from the ITU-T G.975.1 recommendation [18]. Let $\mathcal{C}$ be an $(n_{\mathsf{c}}, k_{\mathsf{c}})$
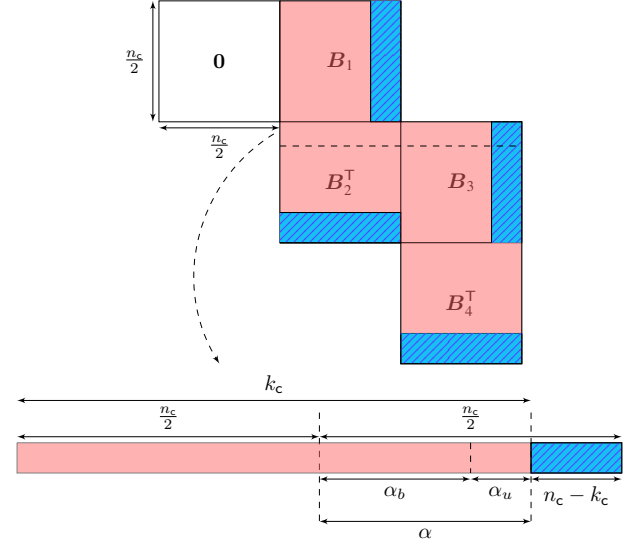

Figure 4. Code array of a staircase code. The red colored area shows the information bits while the area shown with blue hatches corresponds to the parity bits.

shortened BCH code constructed over the Galois field $\mathrm{GF}(2^v)$ with (even) block length n_{c} and information block length k_{c} given by

$$n_{\mathsf{c}} = 2^v - 1 - s,$$
$$k_{\mathsf{c}} = 2^v - vt - 1 - s$$

where s and t are the shortening length and the error correcting capability of the code, respectively. A shortened BCH code is thus completely specified by the parameters (v, t, s). A staircase code with $(n_{\mathsf{c}}, k_{\mathsf{c}})$ component codes is defined as the set of all $\frac{n_{\mathsf{c}}}{2} \times \frac{n_{\mathsf{c}}}{2}$ matrices $\boldsymbol{B}_i$, $i = 1, 2, \ldots$, such that each row of the matrix $[\boldsymbol{B}_{i-1}^{\mathsf{T}}, \boldsymbol{B}_i]$ is a valid codeword of $\mathcal{C}$. Fig. 4 shows the code array of a staircase code. The red parts correspond to the information bits while the parity bits are shown with blue hatches.

The rate of the staircase code with $(n_{\mathsf{c}}, k_{\mathsf{c}})$ BCH component codes is

$$R_{\mathsf{s}} = 1 - \frac{2(n_{\mathsf{c}} - k_{\mathsf{c}})}{n_{\mathsf{c}}}. \tag{8}$$

For a 2^m-ASK constellation, we assume a staircase code with code rate

$$R_{\mathsf{s}} = 1 - \frac{2(n_{\mathsf{c}} - k_{\mathsf{c}})}{n_{\mathsf{c}}} = \frac{m - 1 + \gamma}{m} \tag{9}$$

where $0 \le \gamma < 1$ is a tuning parameter which can be used to select the rate of the staircase code and subsequently the spectral efficiency of the CM scheme.

For $(n_{\mathsf{c}}, k_{\mathsf{c}})$ component codes, each matrix $\boldsymbol{B}_i$ of the staircase code array contains $\alpha = (k_{\mathsf{c}} - n_{\mathsf{c}}/2)$ information

Table II
PARAMETERS OF THE DESIGNED STAIRCASE CODES FOR $v = 10$, $t = 3$, AND 16-ASK MODULATION

γ	0.75	0.6907	0.6809	0.5946	0.5238	0.4444	0.4	0.3617	0.3023	0.2308	0.1429	0
s	63	247	271	431	519	591	623	647	679	711	743	783
n_{c}	960	776	752	592	504	432	400	376	344	312	280	240
k_{c}	930	746	722	562	474	402	370	346	314	282	250	210
n	57600	37636	35344	21904	15876	11664	10000	8836	7396	6084	4900	3600
α_u	90	67	64	44	33	24	20	17	13	9	5	0
R_{s}	0.9375	0.9227	0.9202	0.8986	0.8810	0.8611	0.85	0.8404	0.8256	0.8077	0.7857	0.75

Table III
PARAMETERS OF THE DESIGNED STAIRCASE CODES FOR $v = 10$, $t = 3$, AND 8-ASK MODULATION

γ	0.8125	0.75	0.6	0.5	0.4444	0.25
s	63	303	573	663	669	783
n_{c}	960	720	450	360	324	240
k_{c}	930	690	420	330	294	210
n	76800	43200	16875	10800	8748	4800
α_u	130	90	45	30	24	10
R_{s}	0.9375	0.9167	0.8667	0.8333	0.8148	0.75

bits per row, i.e., each matrix $\boldsymbol{B}_i$ contains a total of $\alpha(n_{\mathsf{c}}/2)$ information bits (see Fig. 4). Consequently, we parse the sequences $\boldsymbol{b}$ and $\boldsymbol{u}^{\mathsf{s}}$ into blocks of length n and γn, respectively, such that

$$n = \frac{\alpha \cdot (n_{\mathsf{c}}/2)}{m - 1 + \gamma}$$

where the parameters $(n_{\mathsf{c}}, k_{\mathsf{c}})$, and hence γ, are chosen such that n and γn are integers. Using (9), it follows that the number of parity bits in each row of $\boldsymbol{B}_i$, $n_{\mathsf{c}} - k_{\mathsf{c}}$, is

$$
\begin{aligned}
n_{\mathsf{c}} - k_{\mathsf{c}} &= \alpha \left(\frac{1}{R_s} - 1 \right) \\
&= \frac{n(m - 1 + \gamma)}{n_{\mathsf{c}}/2} \left(\frac{m}{m - 1 + \gamma} - 1 \right) \\
&= \frac{n(1 - \gamma)}{n_{\mathsf{c}}/2}.
\end{aligned}
\tag{10}
$$

The bits $\boldsymbol{b}_{(i-1)\cdot n+1}, \boldsymbol{b}_{(i-1)\cdot n+2}, \ldots, \boldsymbol{b}_{(i-1)\cdot n+n}$ and $\boldsymbol{u}^{\mathsf{s}}_{(i-1)\cdot \gamma n+1}, \ldots \boldsymbol{u}^{\mathsf{s}}_{(i-1)\cdot \gamma n+n}$, $i = 1, 2, \ldots$, are then placed in $\boldsymbol{B}_i$. In particular, let $\tilde{\boldsymbol{b}}_i$ be the string of bits obtained concatenating $\boldsymbol{b}_{i\cdot n+1}, \boldsymbol{b}_{i\cdot n+2}, \ldots, \boldsymbol{b}_{i\cdot n+n}$ and $\tilde{\boldsymbol{u}}^{\mathsf{s}}_i$ the string of bits obtained concatenating $\boldsymbol{u}^{\mathsf{s}}_{i\cdot \gamma n+1}, \ldots \boldsymbol{u}^{\mathsf{s}}_{i\cdot \gamma n+n}$. Then, $\boldsymbol{b}_i$ and $\tilde{\boldsymbol{u}}^{\mathsf{s}}_i$ are divided into $n_{\mathsf{c}}/2$ equal parts, $\boldsymbol{b}_{i,1}, \ldots, \boldsymbol{b}_{i,n_{\mathsf{c}}/2}$ and $\tilde{\boldsymbol{u}}^{\mathsf{s}}_{i,1}, \ldots, \tilde{\boldsymbol{u}}^{\mathsf{s}}_{i,n_{\mathsf{c}}/2}$, of lengths $\alpha_b = n(m - 1)/(n_{\mathsf{c}}/2)$ and $\alpha_u = \gamma n/(n_{\mathsf{c}}/2)$, respectively, and the vector $(\boldsymbol{b}_{i,j}, \boldsymbol{u}^{\mathsf{s}}_{i,j})$ is placed in row j of matrix $\boldsymbol{B}_{i+1}$. Note that $(n_{\mathsf{c}}, k_{\mathsf{c}})$ and γ must be chosen such that

$$\frac{n}{n_{\mathsf{c}}/2} \in \mathbb{N} \tag{11}$$

$$\frac{\gamma n}{n_{\mathsf{c}}/2} \in \mathbb{N}_0. \tag{12}$$

Consider for simplicity the code array blocks $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$. In this case, vector $(\boldsymbol{b}_{0,j}, \boldsymbol{u}^{\mathsf{s}}_{0,j})$ is placed in row j of $\boldsymbol{B}_1$, for $j = 1, \ldots, n_{\mathsf{c}}/2$, and vector $(\boldsymbol{b}_{1,j}, \boldsymbol{u}^{\mathsf{s}}_{1,j})$ is placed in row j of

$\boldsymbol{B}_2$, for $j = 1, \ldots, n_{\mathsf{c}}/2$. Similarly, the rows of matrices $\boldsymbol{B}_i$ for $i > 2$ are filled. Then, encoding of the staircase code is performed as usual, by row/column encoding, using the $(n_{\mathsf{c}}, k_{\mathsf{c}})$ BCH component code (see Fig. 4). We denote by $\boldsymbol{p}_i$, $i = 1, \ldots, n_{\mathsf{c}}(n_{\mathsf{c}} - k_{\mathsf{c}})/2$, the parity bits of matrix $\boldsymbol{B}_i$ (shown in blue hatches in Fig. 4) resulting from the encoding process, and by $\boldsymbol{p}_{i,j}$, $j = 1, \ldots, n_{\mathsf{c}}/2$ the parity bits in the j-th row of $\boldsymbol{B}_i$. Note that each vector $\boldsymbol{p}_{i,j}$ is of length $n_{\mathsf{c}} - k_{\mathsf{c}}$ bits.

If the ASK constellation is fixed, to achieve different spectral efficiencies, the code rate of the staircase code must be changed. For a given modulation order (given m), one can find different feasible solutions for (v, t, s) that lead to a value of γ that satisfies (11), (12), and $n \in \mathbb{N}$.

In Table II and Table III, some feasible values for s (and therefore γ), the parameters $(n_{\mathsf{c}}, k_{\mathsf{c}})$ for the resulting BCH component codes, n, and the rate of the staircase code, R_{s}, are summarized for 16-ASK modulation ($m = 4$) and 8-ASK modulation ($m = 3$), for $v = 10$ and $t = 3$.

D. Bit-to-Sign Mapping

In the demapping block, for each code array block $\boldsymbol{B}_i$, the parity bits of $\boldsymbol{B}_i$ are multiplexed with $\tilde{\boldsymbol{u}}^{\mathsf{s}}_i$ to generate, through the bit-to-sign mapper Φ_{bs}, the sequence of signs $\boldsymbol{s} = (s_1, \ldots, s_n)$, which will be used as the signs of the amplitudes $a_1, \ldots, a_n$. In particular, we combine the block $\boldsymbol{p}_{i,j}$ of $n_{\mathsf{c}} - k_{\mathsf{c}}$ parity bits (corresponding to the parity bits of row j of $\boldsymbol{B}_i$) with the $\gamma n/(n_{\mathsf{c}}/2)$ information bits $\tilde{\boldsymbol{u}}^{\mathsf{s}}_{i,j}$ to form the vector $\boldsymbol{t}_{i,j} = (\boldsymbol{p}_{i,j}, \tilde{\boldsymbol{u}}^{\mathsf{s}}_{i,j})$, of length $n_{\mathsf{c}} - k_{\mathsf{c}} + \gamma n/(n_{\mathsf{c}}/2) = n/(n_{\mathsf{c}}/2)$, where we used (10). Then, the binary vector $\boldsymbol{t}_i = (\boldsymbol{t}_{i,1}, \ldots, \boldsymbol{t}_{i,n_{\mathsf{c}}/2}) = (t_1, \ldots, t_n)$, of length n bits, is used to form the sequence of n signs $\boldsymbol{s}$ by setting $s_i = 2t_i - 1$. Recalling that the distribution of the signs closely mimics the uniform one, according to (5)–(7), the element-wise multiplication of $(a_1, \ldots, a_n)$ by $(s_1, \ldots, s_n)$ generates a sequence $\boldsymbol{x} = (x_1, \ldots, x_n)$ with the desired distribution.

Finally, the sequence $\boldsymbol{x}$ is scaled by Δ and transmitted over the fiber-optic channel (FOC).

V. DESIGNING THE OPERATING POINT

A. The Effect of Shaping on the Operating Point

For each staircase block $\boldsymbol{B}_i$, the transmitted data is $\boldsymbol{u}^{\mathsf{s}}_{(i-1)\cdot \gamma n+1}, \ldots, \boldsymbol{u}^{\mathsf{s}}_{(i-1)\cdot \gamma n+n}$ and the information bits in $\boldsymbol{u}^{\mathsf{a}}$, which are embedded in the amplitudes $a_{(i-1)\cdot n+1}, a_{(i-1)\cdot n+2}, \ldots, a_{(i-1)\cdot n+n}$. Therefore, the rate is
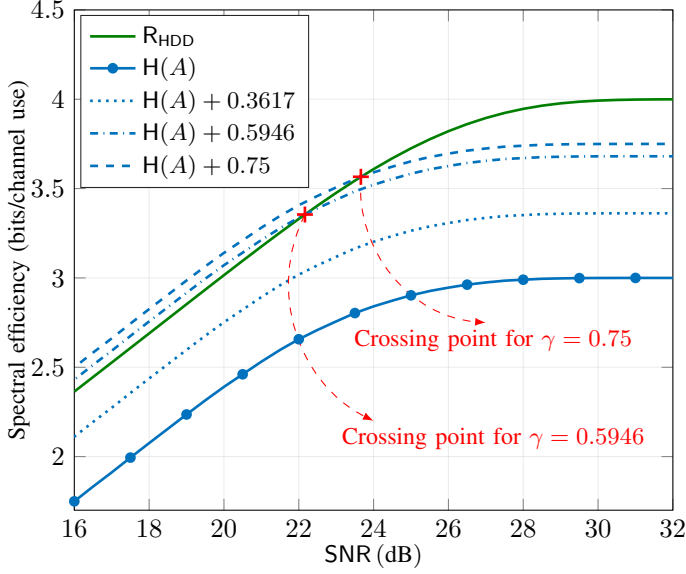
Figure 5. An example for determining the feasible SNR region depending on the selected γ corresponding to some of the designed codes for 16-ASK.
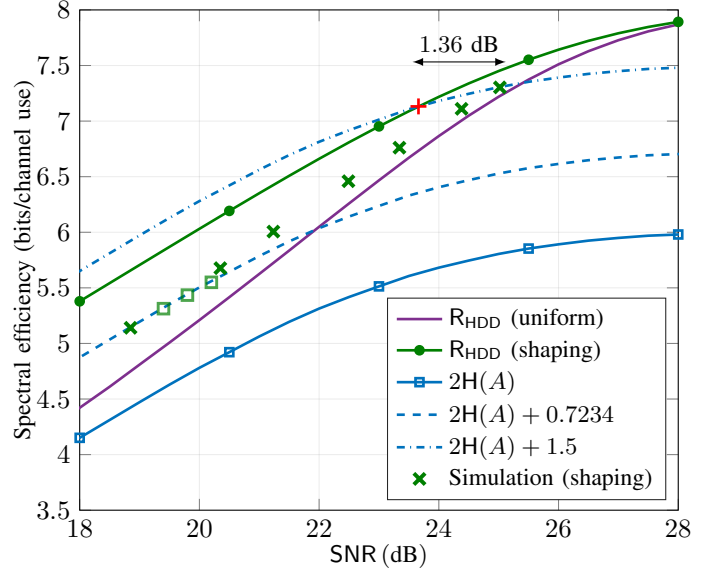

Figure 6. Simulation results of the CM scheme for 256-QAM using the staircase codes summarized in Table II, and comparison with the corresponding achievable information rates.

$R = H(A) + \gamma$ [bits/channel use]. An arbitrarily low error probability can be achieved if

$$R = H(A) + \gamma < R_{HDD}. \tag{13}$$

The crossing point between the curves $H(A) + \gamma$ and R_{HDD} determines the optimal operating point [9], which could be achieved by a capacity achieving code with infinitely large block length. In Fig. 5, we depict R_{HDD} and $R = H(A) + \gamma$ for different values of γ corresponding to some of the designed codes summarized in Table II. As can be seen, for $\gamma = 0$ and $\gamma = 0.3617$, all points on the curve $H(A) + \gamma$ corresponding to the SNRs in the displayed SNR range are achievable (the curve $H(A)+\gamma$ is below R_{HDD}). However, for $\gamma = 0.5946$ and $\gamma = 0.75$ only the points on the curve $H(A)+\gamma$ corresponding to SNRs larger than 22.16 dB and 23.66 dB, respectively, are achievable. Note that below these values, (13) is not satisfied. We remark that, in principle, the points on the curve $H(A)+\gamma$ in the feasible SNR region are only achievable using codes with infinite code length. In practice, codes operate at finite length. For finite code length and a given γ, one can simulate the performance of the designed system and find the minimum SNR in the feasible SNR region required to achieve the desired block error probability.

B. Parameters of the Staircase Code

For a given ASK constellation and a given spectral efficiency, i.e., fixed γ and hence fixed staircase code rate R_s, one can find several component BCH codes (v, t, s) that satisfy (8) and conditions (11), (12), and $n \in \mathbb{N}$, i.e., one can find several staircase codes that yield the desired spectral efficiency. Among them, we may then choose the one that yields the best decoding threshold, i.e., the lowest SNR at which the probability of error goes to zero for infinite block length, using the density evolution derived in [19]. This approach leads, for each spectral efficiency, to the best possible staircase code and therefore to the best performance. However, changing code

for each spectral efficiency may not be desirable in practice. An alternative approach is to fix the parameters (v, t) of the component BCH codes and then tune the shortening parameter s, which leads to different γ and thus to different spectral efficiencies. With this approach one can cover a wide range of spectral efficiencies with a single staircase code.

Note that since one can see square QAM constellations as the Cartesian product of two ASK constellations, the design of the PAS scheme described above readily extends to QAM constellations. In this case, the rate is twice that of the scheme with ASK constellation, i.e., $R = 2H(A) + 2\gamma$. In Fig. 3, the dashed block multiplexes two ASK modulated sequences corresponding to generate the real and imaginary components of the QAM constellation.

VI. NUMERICAL RESULTS

We assess the performance of the CM scheme with PAS and HDD using the codes summarized in Table II and Table III for transmission using square QAM constellations. The target block error probability is set to $P_e = 10^{-3}$. Here, by block error probability we mean the probability that a block of bits of $\boldsymbol{u}^a$ and $\boldsymbol{u}^s$ corresponding to a staircase block is in error. By means of simulations, we find the minimum SNR for which the target P_e is achieved. At the receiver, a symbol-wise MAP detector is used and the decoding of the staircase code is performed using a sliding-window decoder with window size of 7 staircase blocks and a maximum of 8 decoding iterations within the window based on the BDD with extrinsic message passing [21, Algorithm 1].

To minimize the number of the operating modes, we consider the use of a single staircase code (according to the second approach described in Section V-B). In particular, we consider a staircase code with BCH component codes with parameters $(v, t) = (10, 3)$. To achieve different code rates, i.e., different spectral efficiencies, we then find suitable shortening parameters s which lead to a value of γ that satisfies

(11), (12). Some of the values of s and γ are summarized in Table II and Table III.

In Fig. 6, we plot the *practical operating points* of the PAS scheme for 256-QAM (green crosses), corresponding to the minimum SNR required to achieve $P_e = 10^{-3}$ using the optimal distribution $P_X^{\lambda^*}$ for an underlying 16-ASK constellation (the 256-QAM is then obtained as the Cartesian product of two ASKs with distribution $P_X^{\lambda^*}$). The shortening parameters are 63, 274, 431, 519, 591, 623, and 647 (starting from the cross at the top right). The corresponding values of γ are given in the first row of Table II. Note that we can vary the spectral efficiency from 5.14 to 7.3 bits/channel use using a single code and decoder by simply changing the shaping distribution and shortening of the component code. Remarkably, the rates achieved by the probabilistically-shaped CM scheme are larger than the achievable rate of the CM scheme with uniform distribution (purple curve).

As an example, by finding the crossing point between the achievable rate curve and R_{HDD} for $\gamma = 0.75$ ($s = 63$), marked with a red plus sign in the figure, one can see that for all SNRs larger than 23.66 dB reliable transmission can be achieved. However, when using a practical, finite-length staircase code, one should back-off roughly 1.36 dB to meet the target P_e. We also remark that all points on the curves $2H(A)+2\gamma$ with SNR larger than the minimum required SNR can also meet the target performance. However, since by increasing the SNR the curve $2H(A)+2\gamma$ flattens out, at some point one needs to switch to another code rate (another γ) in order to operate as close as possible to the curve R_{HDD}. For instance, for $\gamma = 0.3617$ ($s = 647$), the other possible operating points are shown with green squares. For 20.35 dB, one should switch to $\gamma = 0.4$ ($s = 623$) to approach R_{HDD} more closely. Furthermore, to have a fine granularity for the achievable spectral efficiencies, we can consider several operating points on the same achievable rate curve, e.g., squares on the achievable rate curve $2H(A)+0.7234$ corresponding to $\gamma = 0.3617$.

In Fig. 7, we plot the performance of the CM scheme with 256-QAM modulation (green crosses) and 64-QAM modulation (blue circles). For 256-QAM, the shortening parameters are the same as in Fig. 6. For 64-QAM, the shortening parameters are 63, 303, 573, and 663. We observe that the performance of the CM scheme with 64-QAM for spectral efficiencies below 5.52 bits/channel use is better than that of the scheme with 256-QAM, showing that, depending on the spectral efficiency region, one should change the underlying constellation to achieve better performance. Also, interestingly, the performance for 64-QAM is closer to the corresponding achievable rate curve than for 256-QAM. We believe that this is due to the fact that for a given spectral efficiency that is achievable with both 64-QAM and 256-QAM, the codes used for 64-QAM have higher rate than those for 256-QAM and it is well known that high-rate staircase codes perform better than lower-rate staircase codes [10], [18]. If the complexity of the system is of significant importance, one may disregard the additional gain of switching from 256-QAM to 64-QAM, and consider only 256-QAM.

Finally, in Fig. 8 we compare the performance of the probabilistically-shaped CM scheme with 64-QAM and 256-
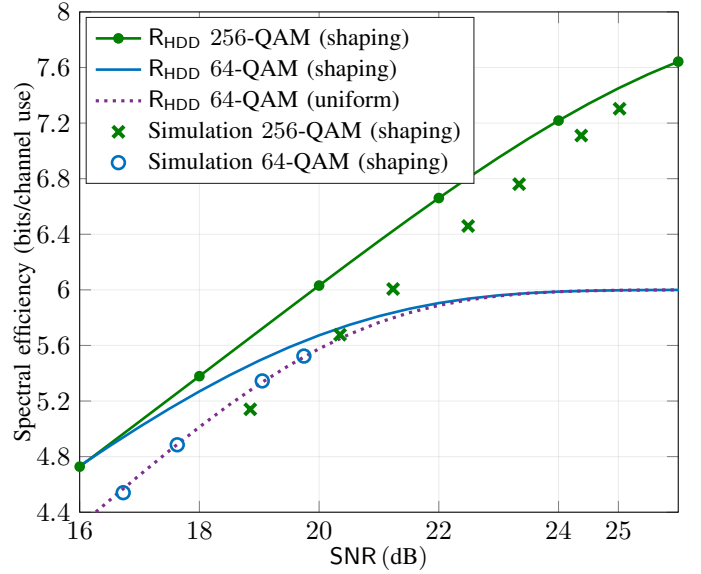

Figure 7. Simulation results of the CM scheme for 256-QAM and 64-QAM using the staircase codes summarized in Table II and the first four columns of Table III, and comparison with the corresponding achievable information rates.
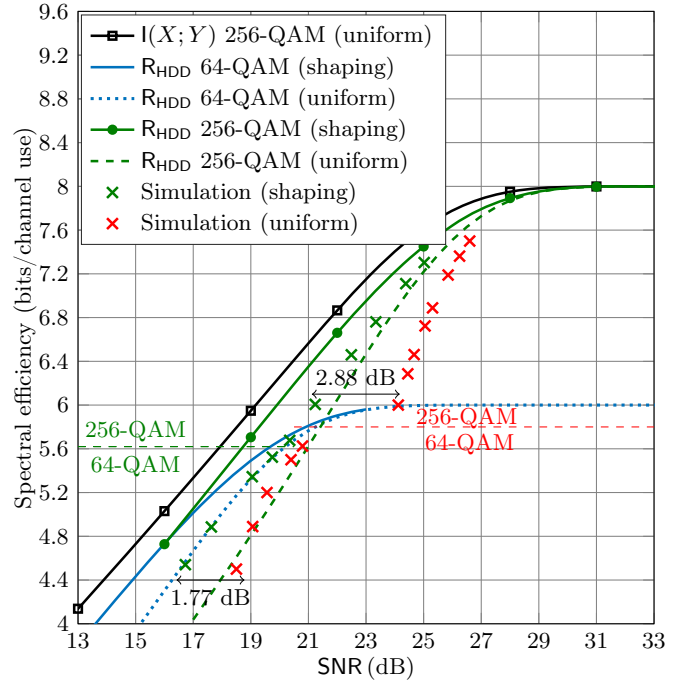

Figure 8. Performance of the probabilistically-shaped CM scheme using $P_X^{\lambda^*}$ and the CM scheme with uniform distribution.

QAM (green crosses) with that of the scheme with uniform distribution (red crosses). For the sake of comparison, we also depict the achievable rate R_{HDD} of the CM scheme with the shaped distribution and the mutual information of the scheme with uniform distribution and symbol-wise SDD, given by $I(X;Y)$ in (1), both for 256-QAM. For the system with uniform distribution, the shortening parameters are 63, 271, 431, 591, 647, 711, 743, and 783 for 256-QAM, and 63, 303, 573, 669, and 783 for 64-QAM. The probabilistic amplitude shaping CM scheme achieves significantly better performance compared to that of the baseline scheme with uniform distribution. Gains up to 2.88 dB and 1.77 dB are achieved

for 256-QAM and 64-QAM, respectively. Furthermore, the probabilistically-shaped CM scheme achieves performance within 0.57–1.44 dB of the corresponding achievable rate for a wide range of spectral efficiencies.

VII. CONCLUSION

We applied probabilistic amplitude shaping to binary staircase codes with hard decision decoding for high-speed fiber-optic communications. We optimize the input distribution to maximize a relevant achievable rate of the CM system with PAS and bit-wise HDD. Outstandingly, the performance of the CM scheme with PAS is significantly better than that of the standard CM scheme with uniform distribution. The probabilistically-shaped CM scheme achieves performance within 0.57–1.44 dB of the achievable rate for a wide range of spectral efficiencies using only a single staircase code, which greatly reduces the decoder complexity.

REFERENCES

[1] A. Sheikh, A. Graell i Amat, and G. Liva, "Probabilistically-shaped coded modulation with hard decision decoding for coherent optical systems," in *Proc. European Conf. Optical Communications (ECOC)*, Gothenburg, Sweden, Sep. 2017.

[2] M. F. Barsoum, C. Jones, and M. Fitz, "Constellation design via capacity maximization," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Nice, 2007, pp. 1821–1825.

[3] Z. H. Peric, I. B. Djordjevic, S. M. Bogosavljevic, and M. C. Stefanovic, "Design of signal constellations for Gaussian channel by using iterative polar quantization," in *Proc. Mediterranean Electrotechnical Conf., (MELECON)*, vol. 2, Tel-Aviv, 1998, pp. 866–869.

[4] I. B. Djordjevic, H. G. Batshon, L. Xu, and T. Wang, "Coded polarization-multiplexed iterative polar modulation (PM-IPM) for beyond 400 Gb/s serial optical transmission," in *Proc. Optical Fiber Commun. Conf. (OFC)*, San Diego, CA, 2010, pp. 1–3.

[5] T. Liu and I. B. Djordjevic, "Multidimensional optimal signal constellation sets and symbol mappings for block-interleaved coded-modulation enabling ultrahigh-speed optical transport," *IEEE Photon. J.*, vol. 6, no. 4, pp. 1–14, Aug. 2014.

[6] O. Geller, R. Dar, M. Feder, and M. Shtaif, "A shaping algorithm for mitigating inter-channel nonlinear phase-noise in nonlinear fiber systems," *IEEE/OSA J. Lightw. Technol.*, vol. 34, no. 16, pp. 3884–3889, Aug. 2016.

[7] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the gaussian channel," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 726–740, Jul. 1990.

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

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

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

[11] F. Buchali, G. Böcherer, W. Idler, L. Schmalen, P. Schulte, and F. Steiner, "Experimental demonstration of capacity increase and rate-adaptation by probabilistically shaped 64-QAM," in *Proc. European Conf. Optical Communications (ECOC)*, Valencia, 2015, pp. 1–3.

[12] C. Pan and F. R. Kschischang, "Probabilistic 16-QAM shaping in WDM systems," *IEEE/OSA J. Lightw. Technol.*, vol. 34, no. 18, pp. 4285–4292, Sep. 2016.

[13] T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik, "On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel," *IEEE/OSA J. Lightw. Technol.*, vol. 34, no. 21, pp. 5063–5073, Nov. 2016.

[14] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration," *IEEE/OSA J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.

[15] A. Ghazisaeidi, I. Fernandez de Jauregui Ruiz, R. Rios-Müller, L. Schmalen, P. Tran, P. Brindel, A. C. Meseguer, Q. Hu, F. Buchali, G. Charlet, and J. Renaudier, "Advanced C+L-band transoceanic transmission systems based on probabilistically shaped PDM-64QAM," *IEEE/OSA J. Lightw. Technol.*, vol. 35, no. 7, pp. 1291–1299, Apr 2017.

[16] B. S. G. Pillai, B. Sedighi, K. Guan, N. P. Anthapadmanabhan, W. Shieh, K. J. Hinton, and R. S. Tucker, "End-to-end energy modeling and analysis of long-haul coherent transmission systems," *IEEE/OSA J. Lightw. Technol.*, vol. 32, no. 18, pp. 3093–3111, Sep. 2014.

[17] C. Häger, A. Graell i Amat, F. Brännström, A. Alvarado, and E. Agrell, "Terminated and tailbiting spatially-coupled codes with optimized bit mappings for spectrally efficient fiber-optical systems," *IEEE/OSA J. Lightw. Technol.*, vol. 33, no. 7, pp. 1275–1285, Apr. 2015.

[18] B. P. Smith, A. Farhood, A. Hunt, F. R. Kschischang, and J. Lodge, "Staircase codes: FEC for 100 Gb/s OTN," *IEEE/OSA J. Lightw. Technol.*, vol. 30, no. 1, pp. 110–117, Jan. 2012.

[19] C. Häger, A. Graell i Amat, H. D. Pfister, A. Alvarado, F. Brännström, and E. Agrell, "On parameter optimization for staircase codes," in *Proc. Optical Fiber Commun. Conf. (OFC)*, Los Angeles, CA, 2015, pp. 1–3.

[20] A. Sheikh, A. Graell i Amat, and M. Karlsson, "Nonbinary staircase codes for spectrally and energy efficient fiber-optic systems," in *Proc. Optical Fiber Commun. Conf. (OFC)*, Los Angeles, CA, 2017.

[21] Y. Y. Jian, H. D. Pfister, K. R. Narayanan, R. Rao, and R. Mazahreh, "Iterative hard-decision decoding of braided BCH codes for high-speed optical communication," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Atlanta, GA, 2013, pp. 2376–2381.

[22] C. Häger, H. D. Pfister, A. Graell i Amat, and F. Brännström, "Density evolution for deterministic generalized product codes on the binary erasure channel at high rates," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4357–4378, Jul. 2017.

[23] C. Häger, A. Graell i Amat, H. D. Pfister, and F. Brännström, "Density evolution for deterministic generalized product codes with higher-order modulation," in *Proc. Int. Symp. Turbo Codes & Iterative Inf. Processing (ISTC)*, Brest, 2016, pp. 236–240.

[24] G. Böcherer, "Achievable rates for probabilistic shaping." [Online]. Available: http://arxiv.org/abs/arXiv:1707.01134.

[25] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3857–3879, Dec. 2012.

[26] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEÜ. Archiv für Elektronik und Übertragungstechnik*, 1993.

[27] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1439–1452, Sep. 1996.

[28] F. Gray, "Pulse code communication," US Patent 2 632 058, 1953.

[29] A. Sheikh, A. Graell i Amat, and G. Liva, "On achievable information rates for coherent fiber-optic systems with hard decision decoding," in *Proc. European Conf. Optical Communications (ECOC)*, Gothenburg, Sweden, Sep. 2017.

[30] ——, "Achievable information rates for coded modulation with hard decision decoding for coherent fiber-optic systems," *IEEE/OSA J. Lightw. Technol.*, vol. 23, no. 35, pp. 5069–5078, Dec. 2017.

[31] 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.

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

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

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

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

[36] F. S. G. Böcherer and P. Schulte, "High throughput probabilistic shaping with product distribution matching." [Online]. Available: http://arxiv.org/abs/arXiv:1702.07510.

[37] ——, "Fast probabilistic shaping implementation for long-haul fiber-optic communication systems," in *Proc. European Conf. Optical Communications (ECOC)*, Gothenburg, Sweden, Sep. 2017.