Probabilistic Amplitude Shaping with Hard Decision Decoding and Staircase Codes

Downloaded from: https://research.chalmers.se, 2023-11-01 10:10 UTC

Citation for the original published paper (version of record):
Sheikh, A., Graell i Amat, A., Liva, G. et al (2018). Probabilistic Amplitude Shaping with Hard Decision Decoding and Staircase Codes. Journal of Lightwave Technology, 36(9): 1689-1697. http://dx.doi.org/10.1109/JLT.2017.2787342

N.B. When citing this work, cite the original published paper.
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. 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...
snr = \frac{u\text{ power}}{\text{scaling parameter}} (i.i.d.) \text{ Gaussian RVs with zero mean and unit variance.} \text{ The spontaneous emission noise [25].}

\text{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].}

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., \( x \) and \( X \), respectively. Expectation with respect to the pmf of RV \( X \) is denoted by \( E_X(\cdot) \). \( H(X) \) and \( I(X; Y) \) 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 \( i \)

\[ Y_i = \Delta X_i + Z_i \quad i = 1, 2, \ldots, n \]

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, \( \Delta \) 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 \( \Delta \) is defined to attain an average transmit power \( P \) according to

\[ E[(\Delta X)^2] = P. \]

According to the definitions above, the SNR is given by \( \text{SNR} = \frac{P}{\Delta}. \) We consider a block-wise transmission system where \( u \) denotes the transmitted information block and \( \hat{u} \) 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 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 \( Y \)

\[ I(X; Y) \triangleq \mathbb{E} \left[ \log_2 \left( \frac{p_{Y|X}(Y|X)}{\sum_{x' \in \mathcal{X}} p_{Y|X}(Y|x') P_X(x')} \right) \right] \quad (1) \]
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 \( R < I(X; Y) \). The ultimate limit is given by the channel capacity, obtained maximizing the MI over all possible input distributions, \( C \triangleq \sup_{P_X} I(X; Y) \).

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 \( X \in \mathcal{X} \) 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

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

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

\[ q(x, \hat{x}) = \varepsilon^{d_H(L(x), L(\hat{x}))} \]

where \( \varepsilon \) is an arbitrary constant in \((0, 1)\), \( L(x) \) is the \( m \)-bit labeling associated with constellation symbol \( x \), and \( d_H(L(x), L(\hat{x})) \) 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]

\[ R_{\text{HDD}} = [H(X) - m H_b(p)]^+ \]

where \( H_b(p) = -p \log_2 p - (1 - p) \log_2 (1 - p) \) 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 \( R_{\text{HDD}} \) 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
maximized, i.e., in this section, the CM system with HDD can operate at a
orders. The figure shows that using the shaping described
rate and spectral efficiencies. In Fig. 2, we depict the achievable
summarizes the shaping gain for different modulation orders
For each SNR, we select $\lambda$ such that the achievable rate is
maximized, i.e.,
$$\lambda^* = \argmax_{\lambda} R_{\text{HDD}}.$$  
(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
$$\hat{x} = \argmax_{x \in \mathcal{X}} P_{Y|X}(y|x)P_X(x)$$
yielding the conditional pmf $P_{Y|X}$ 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 $R_{\text{HDD}}$ 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], $P_{X|\lambda}^\lambda$ is symmetric, i.e., $P_{X|\lambda}^\lambda(x) = P_{X|\lambda}^\lambda(-x)$ for $x \in \mathcal{X}$. Therefore, the random variable $X$ can be factorized as
$$X = A \cdot S$$
(5)
where $A \triangleq |X|$ and $S \triangleq \text{sign}(X)$ are the absolute value and sign of the RV $X$, respectively. It results that $S$ is uniformly distributed,
$$P_S(1) = P_S(-1) = \frac{1}{2}$$
(6)
while the distribution of $A$ satisfies
$$P_A^\lambda(a) = 2P_{X|\lambda}^\lambda(a)$$
(7)
where $a \in A \triangleq \{1, \ldots, 2^m - 1\}$.

The idea of the PAS [9] is then to split the information sequence $u$ into two sequences $u^x$ and $u^s$. The sequence $u^x$ 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 $u^s$, are then encoded using a binary code with a systematic encoder. The parity bits generated by the systematic encoder and $u^x$ 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.

### Table I

| 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              |

Figure 1. Achievable rates with bit-wise HDD ($R_{\text{HDD}}$ 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.
A. Amplitude Shaping

Let \( \mathbf{u} = (u_1, \ldots, u_k) \) 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 \( \mathbf{u} \) is split into two vectors \( \mathbf{u}^s \) and \( \mathbf{u}^a \), of lengths \( \gamma n \) and \( k - \gamma n \), respectively, where \( \gamma \) is a tuning parameter whose meaning will become clear later and it is assumed that \( \gamma n \in \mathbb{N}_0 \). Vector \( \mathbf{u}^a \) is used to generate a sequence of amplitudes \( \mathbf{a} = (a_1, \ldots, a_n) \) with distribution \( P_A \) through a shaping block. The binary interface which generates the sequence of amplitudes with a given distribution (in our case \( P_A \)) 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 \( H(A) \) 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 \( a_1, \ldots, a_n \) is transformed into a sequence of bits using the mapper \( \Phi_{ab} \). We label each of the amplitudes \( a_i \) with \( m - 1 \) bits using the BRGC labeling to construct a binary string \( \mathbf{b}_i \equiv \mathbf{b}(a_i) = (b_{i,1}, \ldots, b_{i,m-1}) \). The sequence \( \mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_n) \) is of length \( \ell = n(m - 1) \) bits.

C. Encoding

The sequences \( \mathbf{b} \) and \( \mathbf{u}^a \) 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 \((n_c, k_c)\) 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 \( C \) be an \((n_c, k_c)\) shortened BCH code constructed over the Galois field GF\((2^v)\) with (even) block length \( n_c \) and information block length \( k_c \) given by

\[
\begin{align*}
n_c &= 2^v - 1 - s, \\
k_c &= 2^v - vt - 1 - s
\end{align*}
\]

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 \((n_c, k_c)\) component codes is defined as the set of all \( \frac{n_c}{2} \times \frac{n_c}{2} \) matrices \( \mathbf{B}_i, i = 1, 2, \ldots \), such that each row of the matrix \( \mathbf{B}_{i-1}^T \mathbf{B}_i \) is a valid codeword of \( C \). 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 \((n_c, k_c)\) BCH component codes is

\[
R_s = 1 - \frac{2(n_c - k_c)}{n_c}. \tag{8}
\]

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

\[
R_s = 1 - \frac{2(n_c - k_c)}{n_c} = \frac{m - 1 + \gamma}{m} \tag{9}
\]

where \( 0 \leq \gamma < 1 \) 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 \((n_c, k_c)\) component codes, each matrix \( \mathbf{B}_i \) of the staircase code array contains \( \alpha = (k_c - n_c/2) \) information...
bits per row, i.e., each matrix $B_i$ contains a total of $\alpha (n_e/2)$ information bits (see Fig. 4). Consequently, we parse the sequences $b$ and $u^*$ into blocks of length $n$ and $\gamma n$, respectively, such that

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

where the parameters $(n_e, k_e)$, and hence $\gamma$, are chosen such that $n$ and $\gamma n$ are integers. Using (9), it follows that the number of parity bits in each row of $B_i$, $n_e - k_e$, is

$$n_e - k_e = \alpha \left( \frac{1}{R_s} - 1 \right) = \frac{n(m - 1 + \gamma)}{\gamma n/2} \left( \frac{m}{m - 1 + \gamma} - 1 \right) = \frac{n(1 - \gamma)}{\gamma n/2}. \quad (10)$$

The bits $b_{i(1)\cdot n+n}, b_{i(1)\cdot n+2}, \ldots, b_{i(1)\cdot n+n}$ and $u^*_{i(1)\cdot n+n} \ldots u^*_{i(1)\cdot n+n}$, $i = 1, 2, \ldots$, are then placed in $B_i$. In particular, let $b_i$ be the string of bits obtained concatenating $b_{i1}, b_{i2}, \ldots, b_{in/n}$ and $u^*_i$ the string of bits obtained concatenating $u^*_1, \ldots, u^*_n$. Then, $b_i$ and $u^*_i$ are divided into $n_e/2$ equal parts, $b_{i1}, b_{i1+n/2}$ and $u^*_{i1}, \ldots, u^*_n/2$. Of lengths $a_{\alpha} = n(m - 1)/(n_e/2)$ and $\alpha_u = \gamma n/(n_e/2)$, respectively, and the vector $(b_{i1}, u^*_{i1})$ is placed in row $j$ of matrix $B_{i+1}$. Note that $(n_e, k_e)$ and $\gamma$ must be chosen such that

$$\frac{n}{n_e/2} \in \mathbb{N} \quad (11)$$

$$\frac{\gamma n}{n_e/2} \in \mathbb{N}_0. \quad (12)$$

Consider for simplicity the code array blocks $B_1$ and $B_2$. In this case, vector $(b_{1j}, u^*_{1j})$ is placed in row $j$ of $B_1$, for $j = 1, \ldots, n_e/2$, and vector $(b_{1j}, u^*_{1j})$ is placed in row $j$ of $B_2$, for $j = 1, \ldots, n_e/2$. Similarly, the rows of matrices $B_i$ for $i > 2$ are filled. Then, encoding of the staircase code is performed as usual, by row/column encoding, using the $(n_e, k_e)$ BCH component code (see Fig. 4). We denote by $p_{ij}, i = 1, \ldots, n_e(n_e - k_e)/2$, the parity bits of matrix $B_i$ (shown in blue hatches in Fig. 4) resulting from the encoding process, and by $p_{ij}, j = 1, \ldots, n_e/2$ the parity bits in the $j$-th row of $B_i$. Note that each vector $p_{ij}$ is of length $n_e - k_e$ 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 $\gamma$ that satisfies (11), (12), and $n \in \mathbb{N}$.

In Table II and Table III, some feasible values for $s$ (and therefore $\gamma$), the parameters $(n_e, k_e)$ 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 $B_i$, the parity bits of $B_i$ are multiplexed with $u^*_i$ to generate, through the bit-to-sign mapper $p_{\gamma}$, the sequence of signs $s = (s_1, \ldots, s_n)$, which will be used as the signs of the amplitudes $a_1, \ldots, a_n$. In particular, we combine the block $p_{ij}$ of $n_e - k_e$ parity bits (corresponding to the parity bits of row $j$ of $B_i$) with the $\gamma n/(n_e/2)$ information bits $u^*_{ij}$ to form the vector $t_{ij} = (p_{ij}, u^*_{ij})$, of length $n_e - k_e + \gamma n/(n_e/2) = n/(n_e/2)$, where we used (10). Then, the binary vector $t_i = (t_{i1}, \ldots, t_{in/e/2}) = (t_{11}, \ldots, t_{n})$, of length $n$ bits, is used to form the sequence of $n$ signs $s$ by setting $s_i = 2t_i - 1$. Recalling that the distribution of the signs closely mimics the uniform one, according to (5)-(7), the elementwise multiplication of $(a_1, \ldots, a_n)$ by $(s_1, \ldots, s_n)$ generates a sequence $x = (x_1, \ldots, x_n)$ with the desired distribution.

Finally, the sequence $x$ is scaled by $\Delta$ 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 $B_i$, the transmitted data is $u^*_{i(1)\cdot n+n} \ldots u^*_{i(1)\cdot n+n}$ and the information bits in $u^*$, which are embedded in the amplitudes $a_{i(1)\cdot n+n}, a_{i(1)\cdot n+2}, \ldots, a_{i(1)\cdot n+n}$. Therefore, the rate is
A. Parameters of the Staircase Code

For a given ASK constellation and a given spectral efficiency, i.e., fixed $\gamma$ 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.

B. Parameters of the Staircase Code

For a given ASK constellation and a given spectral efficiency, i.e., fixed $\gamma$ 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

The crossing point between the curves $H(A) + \gamma$ and $R_{\text{HDD}}$ 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 $R_{\text{HDD}}$ and $R = H(A) + \gamma$ for different values of $\gamma$ 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 $R_{\text{HDD}}$). 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 $\gamma$, 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.

For a given ASK constellation and a given spectral efficiency, i.e., fixed $\gamma$ 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
(11), (12). Some of the values of $s$ and $\gamma$ 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 $P_X^*$. For an underlying 16-ASK constellation (the 256-QAM is then obtained as the Cartesian product of two ASKs with distribution $P_X^*$). The shortening parameters are 63, 274, 431, 519, 591, 623, and 647 (starting from the cross at the top right). The corresponding values of $\gamma$ 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 $R_{\text{HDD}}$ 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 $\gamma$) in order to operate as close as possible to the curve $R_{\text{HDD}}$. 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 $R_{\text{HDD}}$ 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-QAM (green crosses) with that of the scheme with uniform distribution (red crosses). For the sake of comparison, we also depict the achievable rate $R_{\text{HDD}}$ 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.

ACKNOWLEDGMENT

The authors would like to thank Dr. Georg Böcherer for bringing to our attention the achievable rate for probabilistic amplitude shaping in [24] and useful comments.

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 Gbit/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, “Nondeterministic 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, F. S. Buchali, 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.

[11] 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.