

On Product Codes with Probabilistic Amplitude Shaping for High-Throughput Fiber-Optic Systems

Alireza Sheikh *Member, IEEE*, Alexandre Graell i Amat, *Senior Member, IEEE*,
and Alex Alvarado, *Senior Member, IEEE*

Abstract—Probabilistic amplitude shaping (PAS) can flexibly vary the spectral efficiency (SE) of fiber-optic systems. In this paper, we demonstrate the application of PAS to bit-wise hard decision decoding (HDD) of product codes (PCs) by finding the necessary conditions to select the PC component codes. We show that PAS with PCs and HDD yields gains up to 2.7 dB and SE improvement up to approximately 1 bit/channel use compared to using PCs with uniform signaling and HDD. Furthermore, we employ the recently introduced iterative bounded distance decoding with combined reliability of PCs to improve performance of PAS with PCs and HDD.

Index Terms—Coded modulation, hard-decision decoding, probabilistic shaping, product codes, staircase codes.

I. INTRODUCTION

TO keep up with the current trend on data demand and realize high spectral efficiencies (SEs), coding in combination with a high order modulation, a scheme known as coded modulation (CM), has become indispensable in fiber-optic communications. For instance, a recently commercially available optical transponder employs up to 64 quadrature amplitude modulation (QAM) to provide 400 Gbit/s and 600 Gbit/s throughput on a single frequency channel, i.e., 400 Gbit/s/λ and 600 Gbit/s/λ, respectively [1].

In practice, there is always a gap between the performance of CM and the corresponding theoretical limit. This gap is mainly due to two reasons: (i) the employed codes have finite length and suboptimal decoders are usually used in order to constrain the receiver complexity; (ii) equidistant signal constellation points (e.g., square QAM) with uniform signaling are usually employed, yielding the asymptotic 1.53 dB gap to the Shannon limit for transmission over the additive white Gaussian noise (AWGN) channel.

Product-like codes such as product codes (PCs) [2] and staircase codes (SCCs) [3] with hard decision decoding (HDD) are particularly suitable for very high-throughput applications such as next generation fiber-optic systems [4], [5]. To improve the performance of CM with product-like codes, one approach is to boost the decoding performance. As HDD yields a considerable performance loss with respect to soft decision decoding

(SDD), recently, several hybrid decoding schemes have been proposed in order to partially close the gap between HDD and SDD with limited additional complexity to that of HDD [6]–[13]. In particular, iterative bounded distance decoding with combined reliability (iBDD-CR) recently introduced in [6] is a low-complex hybrid decoding scheme for PCs which utilizes the channel reliabilities to improve the performance of HDD.

The performance of CM can also be improved by means of probabilistic shaping [14]–[16]. Probabilistic amplitude shaping (PAS) was proposed in [16] and attracted a significant interest in the fiber-optic community [17]–[20]. The original PAS architecture employs binary low-density parity check (LDPC) codes. However, PAS with PCs and HDD/hybrid decoding can provide a solution for high-throughput fiber-optic systems.

In this paper, we apply PAS to bit-wise HDD of PCs. The contribution is twofold: (i) We define an interleaver in the PAS architecture [16] that distributes the different bit-levels equally among the PC component codewords. Due to the structure of PCs, this can be accomplished by a random interleaver. Further, the structure of PCs imposes some constraints on the PCs to be used with PAS. We find the necessary conditions on the parameters of the PC component codes in order to be used in combination with PAS; (ii) we employ iBDD-CR decoding of PCs [6] to improve the performance of PAS with PCs and (conventional) iterative bounded distance decoding (iBDD). We show that PAS with PCs and iBDD provides gains up to 2.7 dB and a spectral efficiency (SE) improvement of up to 1 bit/channel use (bpcu) compared the baseline scheme with uniform signaling. Furthermore, we show that PAS with PCs and iBDD-CR decoding can close the gap between PAS with iBDD of SCCs and PCs.¹

II. PRELIMINARIES

Without loss of generality, we consider ASK modulation. However, in Sec. IV we report results for square QAM, which can be seen as the Cartesian product of two ASK constellations. The transmitted constellation points are chosen from $\mathcal{X} \triangleq \{-2^m + 1, \dots, -1, 1, \dots, 2^m - 1\}$, where m is the number of bits per symbol. We consider an AWGN channel.² The channel output at time instant i is given as

$$Y_i = X_i + Z_i \quad i = 1, 2, \dots, n \quad (1)$$

¹**Notation:** $P_X(\cdot)$ denotes the probability mass function (PMF) of the RV X . Boldface letters denote vectors and matrices. x_i is the i -th element of \mathbf{x} . $\mathbb{E}_X(\cdot)$ and $H(X)$ stand for the expectation and entropy of the RV X .

²The fiber-optic channel can be modeled as an AWGN channel using the Gaussian noise model [21].

The work of A. Sheikh and A. Alvarado has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 757791).

A. Sheikh and A. Alvarado are with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands (emails: {a.sheikh,a.alvarado}@tue.nl).

A. Graell i Amat is with the Department of Electrical Engineering, Chalmers University of Technology, SE-41296 Gothenburg, Sweden (email: alexandre.graell@chalmers.se).

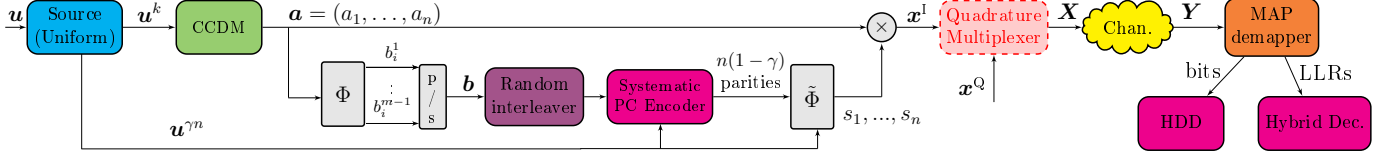


Fig. 1: Block diagram of the CM scheme with PAS and PC under consideration. The parameters shown in the system model corresponds to encoding and decoding of one PC code array. The HD demapper outputs bits and log-likelihood ratios (LLRs) corresponding to the PC code bits for HDD and hybrid decoding, respectively.

where n is the block length, Z_i are independent and identically distributed (i.i.d.) Gaussian RVs with zero mean and unit variance, and X_i is the constrained channel input, hence, $\text{SNR} = \mathbb{E}_X[X^2]$. We consider a block-wise transmission system where \mathbf{u} stands for the transmitted information block and $\hat{\mathbf{u}}$ denotes the corresponding decoded block.

Similar to [18], in this paper we consider the PAS scheme of [16] with a binary code for transmission and bit-wise HDD, i.e., bit-wise Hamming metric decoding, at the receiver side. The achievable information rate of such a system, denoted by R_{HDD} , is derived in [22, eq. (62)]. Similar to [16], [18], we also consider the Maxwell-Boltzmann distribution with shaping parameter λ , where the PMF of a constellation point $x \in \mathcal{X}$ is given as $P_X^\lambda(x) = \frac{\exp(-\lambda x^2)}{\sum_{\tilde{x} \in \mathcal{X}} \exp(-\lambda \tilde{x}^2)}$. For each SNR, λ is optimized such that R_{HDD} is maximized, i.e.,

$$\lambda^* = \underset{\lambda > 0}{\operatorname{argmax}} R_{\text{HDD}}. \quad (2)$$

We use a maximum a-posteriori (MAP) demapper, which employs $P_X^{\lambda^*}$ for the detection.

III. CODED MODULATION WITH PAS AND PCs

The idea of PAS is to shape the constellation amplitudes and then utilize the parities of a systematic code for the sign bits. Let us denote by $\mathcal{A} \triangleq \{1, 3, \dots, 2^m - 1\}$ the set of ASK amplitudes, of size 2^{m-1} . For a target $P_X^{\lambda^*}$, the PMF of an amplitude $a \in \mathcal{A}$ is $P_A^{\lambda^*}(a) = 2P_X^{\lambda^*}(a)$, as the sign bits are uniformly distributed. The schematic of PAS considered in this paper is shown in Fig. 1. In the following, we briefly review the different components of Fig. 1. In what follows, the PAS input is a vector of bits of length $k + \gamma n$, where k bits are employed for amplitude shaping. Furthermore, the output of PAS is a vector of shaped symbols of length n , where each symbol is taken from an ASK constellation of size 2^m .

Let $\mathbf{u} = (u_1, u_2, \dots, u_{k+\gamma n})$ be the information vector of length $k + \gamma n$ bits with uniformly distributed components, where $u_i \in \{0, 1\}$ for $i = 1, 2, \dots, k + \gamma n$. The vector \mathbf{u} is parsed to $\mathbf{u}^{\gamma n}$ and \mathbf{u}^k , of lengths γn and k , respectively. The vector \mathbf{u}^k is the input to the shaping block, which generates a sequence of amplitudes $\mathbf{a} = (a_1, a_2, \dots, a_n)$ with distribution $P_A^{\lambda^*}$. The shaping block employs distribution matching. The block length considered in this paper is large (in the order of 12k–260k), hence, there is no difference between the performance of distribution matchers [23], [24]. Thus, we use the constant composition distribution matching (CCDM) method proposed in [23]. We also consider the binary reflected Gray coding (BRGC) at the mapper Φ . In particular, Φ generates $m-1$ bits corresponding to binary image of shaped amplitudes, yielding sequence $\mathbf{b} = (b_1, b_2, \dots, b_{n(m-1)})$ of length $n(m-1)$

bits, where $\mathbf{b}_i = (b_i^1, b_i^2, \dots, b_i^{m-1})$ for $i = 1, 2, \dots, m-1$. The sequence \mathbf{b} and $\mathbf{u}^{\gamma n}$ are interleaved yielding a sequence of length $n(m-1) + n\gamma$, which is encoded using systematic PCs. The encoding generates parity bits of length $n(1-\gamma)$. By combining $n(1-\gamma)$ parity bits with $\mathbf{u}^{\gamma n}$ and employing the mapping according to $0 \mapsto +1$ and $1 \mapsto -1$ in block $\tilde{\Phi}$, the uniformly distributed sign sequence $\mathbf{s} = (s_1, s_2, \dots, s_n)$ of length n is yielded. The element-wise multiplication of \mathbf{a} by \mathbf{s} generates a sequence $\mathbf{x}^1 = (x_1^1, x_2^1, \dots, x_n^1)$ with the desired distribution for the ASK symbols³. The hard detected (HD) sequence at the receiver is first decoded and then de-interleaved, resulting in the sequences $\hat{\mathbf{a}} = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n)$ and $\hat{\mathbf{u}}^k$. Finally $\hat{\mathbf{a}}$ is de-shaped using an inverse CCDDM and combined with $\hat{\mathbf{u}}^k$ to generate $\hat{\mathbf{u}}$ as the decision on \mathbf{u} . In the following, we discuss the encoding, interleaving, and decoding in more details.

We consider binary PCs with Bose-Chaudhuri-Hocquenghem (BCH) component codes. Let \mathcal{C} be a systematic BCH component code of length \tilde{n} and information length \tilde{k} , in short (\tilde{n}, \tilde{k}) . \mathcal{C} is constructed over the Galois field $\text{GF}(2^v)$ with (even) block length \tilde{n} and information block length \tilde{k} given by

$$\tilde{n} = 2^v - 1 - s, \quad (3)$$

$$\tilde{k} = 2^v - vt - 1 - s \quad (4)$$

where s and t are the shortening length and the error correcting capability of \mathcal{C} , respectively. Therefore, a shortened BCH code is completely specified by the parameters (v, t, s) . A PC with (\tilde{n}, \tilde{k}) component codes is defined as the set of all $\tilde{n} \times \tilde{n}$ arrays $\mathbf{C} = [c_{i,j}]$ such that each row and column of \mathbf{C} is a valid codeword of \mathcal{C} . Fig. 2 shows the code array of a PC. The red part corresponds to the information bits while the parity bits are shown in blue.

In a conventional CM with PCs, for given component parameters v and t , s can be selected without any constraints. However, the PAS architecture imposes a constraint on the parameters of the component codes of the PC, i.e., for a given v and t , only some values of s are feasible. In what follows, we find such constraint. Let us consider a PC with (\tilde{n}, \tilde{k}) BCH component code corresponding to a code rate $R = \frac{\tilde{k}^2}{\tilde{n}^2}$. Using a 2^m -ASK constellation in the PAS, R is given as

$$R = \frac{\tilde{k}^2}{\tilde{n}^2} = \frac{m-1+\gamma}{m}, \quad (5)$$

where $0 \leq \gamma < 1$ is a tuning parameter used to vary the SE of PAS. We assume that the bit sequence corresponding to the output of the interleaver (of size $(m-1+\gamma)n$) is parsed

³For QAM modulation, \mathbf{x}^Q is generated similarly and multiplexed with \mathbf{x}^1 , which yields a sequence with the desired distribution for the QAM symbols.

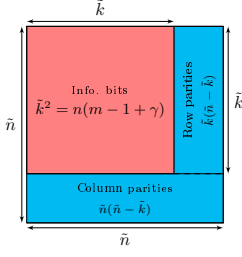


Fig. 2: Code array of a PC.

Table I: Parameters of the designed PCs for $v = 10$, $t = 3$, and 16-ASK modulation

γ	0.7682	0.7503	0.6912	0.6797	0.5942	0.5233	0.4464	0.3645	0.2303	0.1426	0.00854
s	3	77	261	289	447	535	605	661	727	759	797
\tilde{n}	1020	946	762	734	576	488	418	362	296	264	226
\tilde{k}	990	916	732	704	546	458	388	332	266	234	196
n	260100	223729	145161	134689	82944	59536	43681	32761	21904	17424	12769
γn	199800	167869	100341	91549	49284	31156	19501	11941	5044	2484	109
R	0.9420	0.9376	0.9228	0.9199	0.8985	0.8808	0.8616	0.8411	0.8076	0.7856	0.7521

into \tilde{k} sequences of length \tilde{k} , each placed as information bits corresponding to the i -th row ($i = 1, \dots, \tilde{k}$) of the PC code array (see red part in Fig. 2). Therefore, n should be selected such that

$$n = \frac{\tilde{k}^2}{m - 1 + \gamma}. \quad (6)$$

As explained before, the number of parity bits should be $n(1 - \gamma)$. Employing (5) and (6), the total number of parity bits corresponding to a PC codeword is given as

$$\tilde{n}^2 - \tilde{k}^2 = \tilde{k}^2 \left(\frac{1}{R} - 1 \right) = n(1 - \gamma). \quad (7)$$

The component codes of the PC employed in the PAS should then be chosen such that the block length (6) and the number of parity bits (7) are positive and non-negative integers, respectively, i.e.,

$$n \in \{1, 2, \dots\}, \text{ and } \gamma n \in \{0, 1, \dots\}. \quad (8)$$

Therefore, in order to use PCs with component code parameters (v, t) in PAS with 2^m -ASK constellation, the feasible values of s must satisfy (6) and (8). We highlight that a feasible value of s corresponds to a feasible value of γ (by employing (3)–(4) in (5), one can see a one-to-one relation between s and γ).

For 16-ASK modulation ($m = 4$), $v = 10$ and $t = 3$, a subset of all feasible values for s and correspondingly γ , (\tilde{n}, \tilde{k}) , n , and R are given in Table I. The same table for SCCs with 16-ASK modulation ($m = 4$), $v = 10$, and $t = 3$ was given in [18, Table II]. Comparing Table I with [18, Table II] for a given code rate, one can see that the structural difference between PCs and SCCs yields different component code parameters for each code. We highlight that the feasible values for s (and thus γ) in Table I are only a subset of a total of 205 feasible values. One can check that the total number of feasible values for s (and thus γ) for a SCC is only 40 (see [18, Table II]). As it is shown in Sec. IV, by varying γ one can change the SE of PAS. Therefore, PAS with PCs provides a much finer granularity of SEs than PAS with SCCs, due to the larger number of feasible values of γ .

We assume that the sign bits generated by PAS, i.e., (s_1, \dots, s_n) , represent the first bit level in the BRGC label of the transmitted symbol. For instance, the transmitted signal points with the corresponding BRGC labels for 8-ASK modulation are shown in Table II. It is well-known that the reliability of different bit levels varies in CM. For symmetric PCs, i.e., with the same row and column component codes,

Table II: BRGC of the amplitudes for 8-ASK

amplitude	−7	−5	−3	−1	1	3	5	7
label	110	111	101	100	000	001	011	010

the performance of iterative decoding improves if both row and column component decoder observe the same channel. This can be achieved by uniformly spreading the bits of \mathbf{b} and $\mathbf{u}^{\gamma n}$ (which correspond to different bit levels) between the component codes of the PC. With this aim, we employ a random interleaver in the PAS architecture (see Fig. 1). We highlight that there is no random interleaver in [18]. In fact, a simple structured interleaver is considered in [18, Sec. IV-C] for SCCs, which ensures that bits from both \mathbf{b} and $\mathbf{u}^{\gamma n}$ are uniformly distributed between rows in each staircase block.

PCs are usually decoded using iBDD of the component codes [25]. In this paper, we also consider the iBDD-CR algorithm introduced in [6] for PCs. In order to apply iBDD-CR in PAS with PCs, the LLR of the hard detected bits at the decoder should be computed. Given the received symbol Y_i (see (1)), one can compute the LLR corresponding to the l -th bit ($l = 1, \dots, m$) of the binary image of Y_i as

$$L_i^l = \ln \left(\frac{\sum_{\tilde{x}_i \in \mathcal{S}_l^0} e^{-\frac{(Y_i - \tilde{x}_i)^2}{2}} \cdot P_X^{\lambda^*}(\tilde{x}_i)}{\sum_{\tilde{x}_i \in \mathcal{S}_l^1} e^{-\frac{(Y_i - \tilde{x}_i)^2}{2}} \cdot P_X^{\lambda^*}(\tilde{x}_i)} \right), \quad (9)$$

where \mathcal{S}_l^0 and \mathcal{S}_l^1 are sets of size 2^{m-1} ASK symbols with 0 and 1 as the l -th bit of the corresponding BRGC label, respectively.

IV. NUMERICAL RESULTS

We evaluate the performance of PAS using PCs with parameters given in Table I for transmission of 256-QAM. Note that the Cartesian product of two 16-ASKs with distribution $P_X^{\lambda^*}$ gives the shaped 256-QAM. The target block error probability (P_e) is set to 10^{-3} . The decoding of the PC is performed using a maximum of 8 decoding iterations employing iBDD with extrinsic message passing [26, Algorithm 1] and iBDD-CR [6].

For the feasible values of s given in Table I, n is large, hence the information rate of PAS for each quadrature is $H(A) + \gamma$. Therefore, the total rate is $2H(A) + 2\gamma$. In Fig. 3, we depict R_{HDD} (both shaped and uniform) for a 256-QAM and $2H(A) + 2\gamma$ for values of γ corresponding to $s = 77, 261, 447, 535$, and 605 . As can be seen, for $\gamma = 0.4464$ ($s = 605$) the curve $2H(A) + 0.8928$ is below the shaped R_{HDD} , hence, employing a single PC with code $R = 0.8616$ (see Table I,

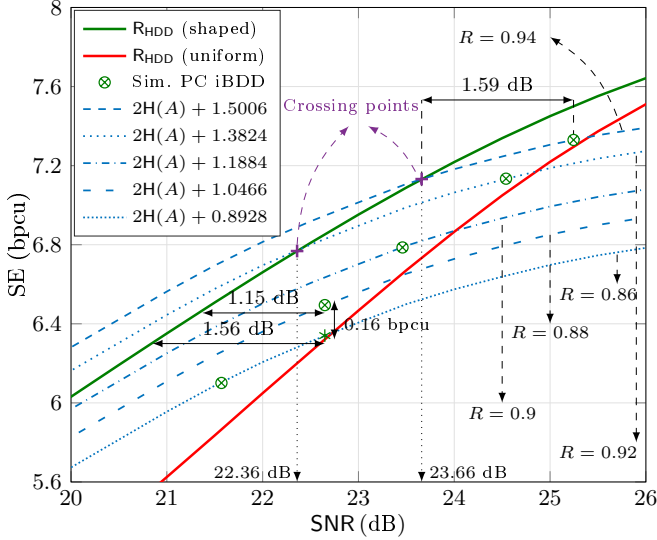


Fig. 3: Simulation results of the PAS based on PCs with $s = 77, 261, 447, 535, 605$, for 256-QAM, and comparison with the corresponding achievable information rates.

eighth column), one can achieve all points on $2H(A) + 0.8928$ by only changing the distribution of amplitudes. A similar observation can be made for $\gamma = 0.5233$ ($s = 535$) and $\gamma = 0.5942$ ($s = 447$). However, for $\gamma = 0.7503$ ($s = 77$) and $\gamma = 0.6912$ ($s = 261$) there are crossing points which are shown with plus signs. The crossing point between the curves $2H(A) + 2\gamma$ and shaped R_{HDD} specifies the optimal operating point, which can be achieved with infinity long block length codes and optimal decoding [16]. As suboptimal iterative decoding of PCs is considered in this paper, only the points on the curve $2H(A) + 1.5006$ and $2H(A) + 1.3824$ corresponding to SNRs larger than 23.66 dB and 22.36 dB, respectively, are achievable. Therefore, the SNRs larger than 23.66 dB and 22.36 dB defines the feasible SNR regions for PAS based on employing PCs with $s = 77$ and $s = 261$, respectively.

We have simulated the performance of the PAS with the considered PCs and then find the minimum SNR in the feasible SNR region where P_e can be achieved. In what follows, we refer to the point on the curve $2H(A) + 2\gamma$ corresponding to the minimum SNR achieving P_e as an *operating point*. In Fig. 3, we plot the operating points of the PAS corresponding to PCs with $s = 77, 261, 447, 535$, and 605 , for 256-QAM to achieve $P_e = 10^{-3}$ (green circle with crosses). As it can be seen, for $\gamma = 0.7503$ (corresponding to crossing point with SNR = 23.66 dB), one should back-off roughly 1.59 dB from the optimal operating point (plus sign corresponding to SNR = 23.66 dB) to achieve $P_e = 10^{-3}$. Note that the simulation points are just a subset of size 5 out of 205 feasible operating points (see Sec. III).

We also notice that the slope of $2H(A) + 2\gamma$ is less than that of shaped R_{HDD} , hence, we will switch to another code rate (different γ) in order to operate as close as possible to shaped R_{HDD} . For instance, the point shown with asterisk is also on the achievable rate curve $2H(A) + 0.8928$, corresponding to $\gamma = 0.4464$. However, switching to $\gamma = 0.5233$ (operating on the curve $2H(A) + 1.0466$) reduces the gap between the operating

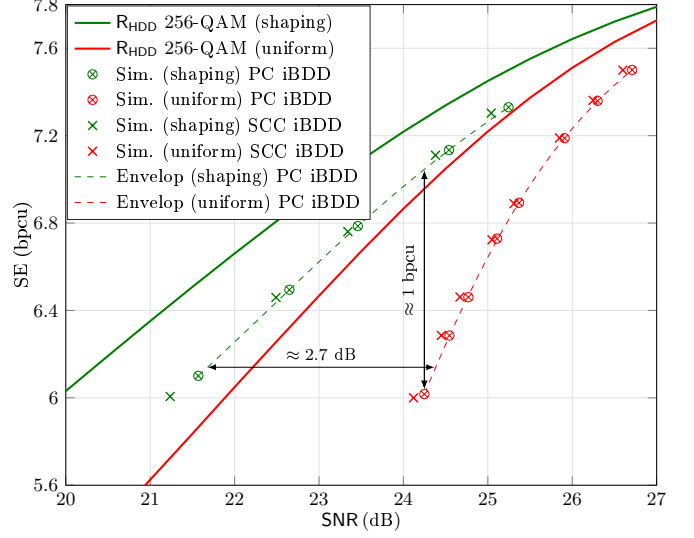


Fig. 4: Simulation results of PAS with PC and SCCs with parameters summarized in Table I and [18, Table. II], respectively, for 256-QAM, and comparison with the corresponding achievable information rates.

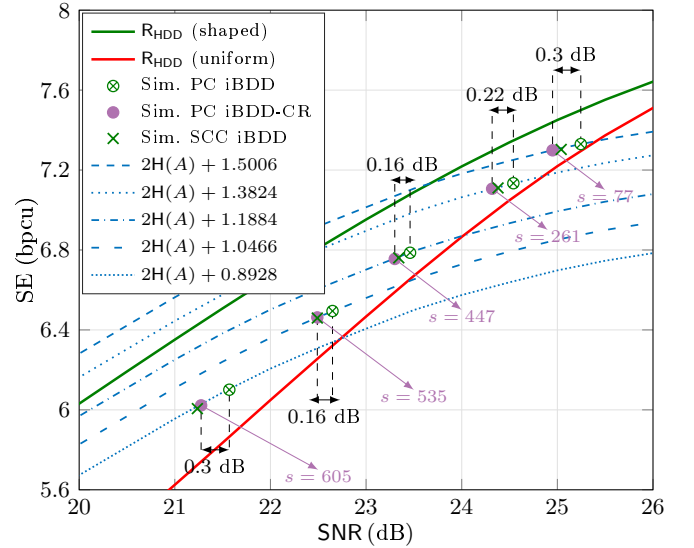


Fig. 5: Performance of the PAS with PC and iBDD-CR algorithm, PC with iBDD, and SCC with iBDD. Starting from top right, the shortening parameter for PC is 77, 266, 447, 535, and 605, and shortening parameter for SCC is 63, 247, 431, 519, and 591.

point and shaped R_{HDD} by 0.41 dB, which corresponds to 0.16 bpcu higher SE for the same SNR. Finally, Fig. 3 shows that the achieved rates by the PAS with PCs are larger than the achievable rate of the (conventional) CM with uniform signaling (compare green circles with crosses and red curve).

For the sake of comparison, we consider the same scenario for PAS and SCCs using a window decoder of size 7 stair-case blocks and 8 decoding iterations, as considered in [18, Sec. VI]. In particular, to have a fair comparison between SCCs and PCs, we consider SCCs with $(v, t) = (10, 3)$ and $s = 63, 274, 431, 519$, and 591, which gives virtually the same code rates as those of the PCs considered in Fig. 4. Comparing green crosses with green circles with crosses, one can see that for PCs the back-off SNR is larger than for SCCs, hence, PAS with PCs yields higher SEs than PAS with SCCs, at the cost of higher operating SNR.

In Fig. 4, we compare the performance of PAS with PCs and SCCs with uniform signaling. In particular, we consider PCs with $s = 77, 289, 447, 605, 661, 727, 759$, and 797 . These PCs have roughly the same code rate as SCCs with $s = 63, 271, 431, 591, 647, 711, 743$, and 783 that are considered in [18, Fig. 8]. As can be seen, the performance of both PAS with PCs and uniform signaling varies on an envelope (see green and red dashed curves in Fig. 4), where the performance improvement of PCs with shaping over the uniform signaling reaches up to 2.7 dB. Furthermore, comparing the SEs of PAS with PCs and uniform signaling, one can see that employing shaping provides up to 1 bpcu SE improvement. Comparing the simulation points with the corresponding achievable information rate (AIR) curve, it is clear that by reducing the code rate for both PCs and SCCs, the SNR gap between operating points and the AIR increases. It is worth to point out that this gap is always lower for the PAS with PCs compared to PCs with uniform signaling.

In Fig. 5, we compare the performance of PAS with PCs decoded using iBDD-CR [6], with that of PAS and both PCs and SCCs decoded using iBDD. Comparing the operating points of PAS with iBDD-CR and iBDD of PCs, an SNR gain of up to 0.3 dB is observed. The improvement yielded by iBDD-CR is at the cost of computing and storing the channel LLRs compared to iBDD [6]. Interestingly, PCs with iBDD-CR decoding and $s = 77, 261$, and 447 provide better operating points than that of SCCs with iBDD and same code rate ($s = 63, 247$, and 431). Furthermore, the PC with iBDD-CR and $s = 519$ provides the same performance as the corresponding SCC ($s = 519$) with iBDD. Also, the PC with iBDD-CR and $s = 605$ provides slightly worse performance compared to the corresponding SCC ($s = 591$) with iBDD.

V. CONCLUSION

We considered probabilistic amplitude shaping with binary PCs and HDD based on iBDD of the component codes. In particular, we addressed the parameter selection of the PC component codes and found the corresponding SE operating points. We showed that the performance of PAS with PCs is up to 2.7 dB better than that of the standard CM scheme with PCs and uniform signaling. Furthermore, we showed that PAS combined with the iBDD-CR decoding algorithm for PCs outperforms PAS with conventional iBDD for PCs, where the performance improvement is up to 0.3 dB. Interestingly, we observed that PAS with PCs and iBDD-CR can close the gap between PAS with iBDD of SCCs and PCs. Future work includes extending of iBDD-CR to SCCs, and investigating the performance of PAS with iBDD-CR of SCCs.

We highlight that in this paper we used a pragmatic approach to evaluate the performance of PAS with PCs and iBDD-CR, by maximizing the AIR of the PAS with HDD. In principle, hybrid decoders (such as iBDD-CR) operate beyond the Hamming metric. Therefore, the operating point shown in this paper can be improved further by first deriving the AIR of the hybrid system and then optimizing the AIR. This analysis is left as future work.

REFERENCES

- [1] Fujitsu, "1FINITY T600 Transport Blade," Online: <https://www.fujitsu.com/us/Images/1FINITY-T600-Data-Sheet.pdf>.
- [2] P. Elias, "Error-free coding," *Trans. IRE Professional Group on Inf. Theory*, vol. 4, no. 4, pp. 29–37, Sep. 1954.
- [3] 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.
- [4] "Forward error correction for high bit-rate DWDM submarine systems," ITU-T Recommendation G.975.1, 2004.
- [5] O. I. Forum, "Implementation agreement 400ZR," 2018.
- [6] A. Sheikh, A. Graell i Amat, G. Liva, and A. Alvarado, "Refined reliability combining for binary message passing decoding of product codes," *arXiv*, 2020. [Online]. Available: <https://arxiv.org/abs/2006.00070>
- [7] A. Sheikh, A. Graell i Amat, and G. Liva, "Binary message passing decoding of product-like codes," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8167–8178, Dec. 2019.
- [8] —, "Binary message passing decoding of product codes based on generalized minimum distance decoding," in *Proc. 53rd Annu. Conf. Inf. Sciences and Systems (CISS)*, Baltimore, MD, Mar. 2019.
- [9] C. Fougstedt, A. Sheikh, A. Graell i Amat, G. Liva, and P. Larsson-Edefors, "Energy-efficient soft-assisted product decoders," in *Proc. Optical Fiber Commun. Conf. (OFC)*, San Diego, CA, USA, Mar. 2019.
- [10] G. Montorsi, "Low complexity two-stage decoders for bawgn," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, Shanghai, China, May 2019.
- [11] Y. Lei, A. Alvarado, B. Chen, X. Deng, Z. Cao, J. Li, and K. Xu, "Decoding staircase codes with marked bits," in *Proc. Int. Symp. Turbo Codes and Iterative Inf. Processing (ISTC)*, Hong Kong, Dec. 2018.
- [12] M. Barakatain and F. R. Kschischang, "Low-complexity concatenated LDPC-staircase codes," *IEEE/OSA J. Lightw. Technol.*, vol. 36, no. 12, pp. 2443–2449, Jun. 2018.
- [13] G. Liga, A. Sheikh, and A. Alvarado, "A novel soft-aided bit-marking decoder for product codes," in *Proc. European Conf. Optical Communications (ECOC)*, Dublin, Ireland, Sep. 2019.
- [14] 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.
- [15] G. D. Forney, Jr., "Trellis shaping," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 281–300, Mar. 1992.
- [16] 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.
- [17] 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.
- [18] A. Sheikh, A. Graell i Amat, G. Liva, and F. Steiner, "Probabilistic amplitude shaping with hard decision decoding and staircase codes," *IEEE/OSA J. Lightw. Technol.*, vol. 36, no. 9, pp. 1689–1697, May 2018.
- [19] G. Böcherer, P. Schulte, and F. Steiner, "Probabilistic shaping and forward error correction for fiber-optic communication systems," *IEEE/OSA J. Lightw. Technol.*, vol. 37, no. 2, pp. 230–244, Jan. 2019.
- [20] J. Cho and P. J. Winzer, "Probabilistic constellation shaping for optical fiber communications," *IEEE/OSA J. Lightw. Technol.*, vol. 37, no. 6, pp. 1590–1607, Mar. 2019.
- [21] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "A simple and effective closed-form GN model correction formula accounting for signal non-gaussian distribution," *IEEE/OSA J. Lightw. Technol.*, vol. 33, no. 2, pp. 459–473, Jan. 2015.
- [22] G. Böcherer, "Achievable rates for probabilistic shaping," V5, May, 2017. [Online]. Available: <http://arxiv.org/abs/1707.01134>.
- [23] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.
- [24] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, "Multiset-partition distribution matching," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1885–1893, Mar. 2019.
- [25] J. Justesen, "Performance of product codes and related structures with iterated decoding," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 407–415, Feb. 2011.
- [26] 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, Dec. 2013.