Probabilistically Shaped Rate-Adaptive Polar-Coded 256-QAM WDM Optical Transmission System

Iqbal, Shajeel; Kaminski, Pawel Marcin; Klejs, Frederik; Yankov, Metodi Plamenov; Porto da Silva, Edson; Da Ros, Francesco; Oxenløwe, Leif Katsuo; Forchhammer, Søren

Published in:
Journal of Lightwave Technology

Link to article, DOI:
10.1109/jlt.2019.2959829

Publication date:
2020

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
Probabilistically Shaped Rate-Adaptive Polar-Coded 256-QAM WDM Optical Transmission System

Shajeel Iqbal, Pawel M. Kaminski, Frederik Klejs Member, OSA, Metodi P. Yankov Member, IEEE, Edson P. da Silva Senior Member, IEEE, Member, OSA, Francesco Da Ros Member, IEEE, Leif K. Oxenløwe Member, IEEE, Fellow, OSA, and Søren Forchhammer Member, IEEE

Abstract—In this paper, a rate-adaptive coded modulation (CM) system combining polar codes and many-to-one probabilistic shaping is constructed and experimentally demonstrated. We propose to control the polar codes using a fraction of bits referred to as pre-set bits. This not only allows to offset the puncturing loss of rate-adaptive polar codes but also provides shaping gains compared to the non-punctured polar codes. Preset bits and many-to-one shaping are combined to form a rate-adaptive bit-interleaved CM system. We experimentally evaluate the system performance for 10×8-Gbd dual polarization 256 quadrature amplitude modulation (QAM) wavelength division multiplexed (WDM) system with various input data rates ranging from 61 Gbps to 91 Gbps per carrier. The experimental results demonstrate a 200 km reach increase over a wide range of distances compared to the non-punctured polar codes and punctured polar codes.

Index Terms—Polar codes, Punctured polar codes, Rate-adaptation, Probabilistic shaping, wavelength division multiplexing, Experimental demonstration.

I. INTRODUCTION

M ODERN optical communication systems must adapt their bit rate (and spectral efficiency (SE)) based on channel conditions for robustness [1]–[3]. Future optical mesh networks and radio-over-fiber (RoF) networks both demand transceivers to operate with different signal qualities. Therefore, a coded modulation (CM) scheme with a possibility to adapt its transmission rate to the varying channel conditions is needed in these networks. Forward-error-correction (FEC) coding and constellation shaping (CS) can be employed in a bit-interleaved coded modulation (BICM) system to design a rate-adaptive CM scheme (Fig. 1).

The CS can be distinguished into two approaches; geometric shaping (GS) and probabilistic shaping (PS). In GS, the quadrature amplitude modulation (QAM) symbols are geometrically placed in space to obtain a non-uniform distribution of symbols [4]–[6]. The PS is based on optimizing the probability of occurrence of the constellation symbols where, e.g., for Gaussian channels, low-amplitude symbols should appear more frequently [7] [8]. One example of PS is the probabilistic amplitude shaping (PAS) [7], which concatenates a constant-composition distribution matcher (CCDM) [9] with a systematic FEC encoder. In the CCDM, a binary input data sequence to a systematic FEC encoder is first transformed into a sequence of symbols with a Maxwell-Boltzmann (MB) distribution. A similar technique was adopted in [10], where list decoding of polar codes is used to realize the MB distribution of QAM symbols. The schemes in [7] and [10] both require an additional shaping encoder at the transmitter and/or shaping decoder at the receiver which increases the complexity at the transmitter and/or receiver. A many-to-one PS method was presented as a low complexity solution in [11] combined with turbo codes. However, the system in [11] suffers high latency due to the turbo decoder and might need iterative de-mapping for low signal-to-noise (SNR) ratio.

Most of the rate-adaptive CM systems for optical transmission have been tested with low-density parity-check (LDPC) codes, e.g., [3] [12] [13]. However, polar codes [14] have interesting characteristics. They are known to asymptotically achieve the capacity of any discrete memory-less symmetric channel [14]. The successive cancellation (SC) decoder for polar codes does not display an error floor [15], as compared to LDPC codes which display an error floor if they are not carefully designed for each SE [16]. Short polar codes have been adopted for the fifth-generation (5G) wireless standard [17] with low encoding and decoding algorithm complexity outperforming, e.g., new radio LDPC codes for short code lengths [18]. Hardware implementations of short polar codes...
have shown similar performance benefits as the other FEC codes such as LDPC and turbo codes [19]. Such short 5G polar codes could also be applied to optical systems incorporating wireless technologies such as RoF systems [20]. Another potential use for short polar codes can be as an inner soft-decision code in a concatenated FEC scheme similar to that composed by an outer staircase code and a soft-decision Hamming code specified in [21] targeting 400G data center interconnect. Hamming codes can be replaced by short polar codes in high-end optical communication for even better performance.

Rate-adaptive polar codes can be constructed using two techniques, namely; puncturing [22] or shortening [23]. Both techniques are based on assigning known values to indices of the uncoded message which are later not transmitted, and the frozen set is co-designed with the punctured/shortened bits. An overview of puncturing and shortening techniques for polar codes is provided in [24]. In puncturing, the log-likelihood ratios (LLRs) of punctured bits, called incapable bits, are set to zero at the decoder, while in shortening the LLR of shortened bits, called over-capable bits, are set to infinitely high values at decoder. However, both methods come at the cost of increased bit-error rate (BER) [25] due to the incapable or over-capable bits in the decoder. As a part of our previous work [25], we experimentally demonstrated how many-to-one PS could be used to offset the performance loss related to puncturing of polar codes.

As the main contribution of this paper, we jointly optimize the many-to-one shaper with the redundant bits in the polar codes [26]. More precisely, we pre-set some of the bits in the polar coded codeword in a controlled manner inspired by the shortening of polar codes [23]. The pre-set bits are carefully placed in the mapping function so as to eliminate the effect of the ambiguities caused by the many-to-one mappings. The pre-set bits are known at the receiver, which leads to higher shaping gains than reported in our previous work [25]. In this work, we show by simulations and experimental results that the shaping with pre-set bits offsets the puncturing loss as well as providing shaping gains compared to the non-punctured polar codes.

The rest of the paper is organized as follows: First, we introduce polar codes and many-to-one CS method in Sections II and III, respectively. A low-complexity rate-adaptive CM scheme based on many-to-one PS and polar codes is described in Section IV. In Section IV, we also discuss the performance of the proposed system. The performance of our proposed system is numerically compared with other rate-adaptive schemes in Section V. We extend our simulated system with pre-set bits in [26] and experimentally investigate the proposed scheme in back-to-back and optical transmission on a fiber link in Sections VII and VIII, respectively. We show that the proposed polar-coded modulation scheme increases the transmission reach of the system w.r.t. the punctured polar codes and non-punctured polar codes. The system design not only provides shaping gains, but at the same time, very fine granularity in rate step sizes can be realized. Section IX concludes this paper.

II. POLAR CODES

We consider polar codes [14], [27] as a FEC code for the BICM model in Fig. 1. Polar coding is based on the channel polarization [14] operation, not to be confused with the polarization of the optical field. The channel polarization operation converts a physical channel into virtual channels with very high reliabilities and with very low reliabilities. The reliabilities can be calculated by the Bhattacharya parameter [14], Monte-Carlo simulation of bit-channels or density evolution under a Gaussian approximation. An overview and comparison of such methods can be found in [29]. In this work, we use the method proposed in [14]. Once the channel reliabilities are computed the polar code can be constructed. For a specified block length $N$ and $K$ information bits, the polar encoded codeword $x^N_1$ can be defined as:

$$x^N_1 = u^N_1 G^N_2$$  \hspace{1cm} (1)$$

where $n = \log_2(N)$, $\otimes$ denotes the Kronecker product, and

$$G_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$ \hspace{1cm} (2)$$

The vector $u^N_1$ contains the $K$ information bits at the most reliable bit positions and the $N-K$ least reliable bit positions are set to the pre-known values called as frozen bits. The frozen bits are the added redundant bits in the polar code. The set $I \subseteq 1, 2, ..., N$ contains the indices of information bits and the frozen bits. The rate $R$ of the code is defined as $R = K/N$ and can be easily changed by selecting or deselecting the frozen bits.

The decoding of the polar codes is based on the SC decoding [14]. The SC decoding estimates the bits sequentially i.e., the bit $u_{i+1}$ is estimated using the previous bit estimates $u_1,...,u_i$ and the channel output observation $y^N_1 = [y_1, y_2, ..., y_N]$. The dependence of the estimated value on previous estimates in the SC decoder leads to propagation of error in the SC tree, once an error occurred in the decoding. Due to this, the SC decoder has poor performance. Polar codes perform best under cyclic-redundancy check (CRC) aided successive cancellation list (CA-SCL) decoder [28], which is considered in this work. The CA-SCL decoder for polar codes is also based on the SC decoder and can be affected by the error propagation if the list size is not large enough.

III. MANY-TO-ONE CONSTELLATION SHAPING

Many-to-one constellation shaping can be designed using the dyadic probability mass function (PMF) approximations [8]. A dyadic PMF has the form $p(x_i) = 2^{-d_i}$, where $d_i$ is an integer and $x_i$ is the $i$-th symbol from the constellation. The optimized PMF for a fiber link can be found by maximizing the achievable information rate (AIR) [11]. The PMF obtained from [11] is found to be robust also for other fiber link distances and is therefore applied here. From the algorithm in [31] the dyadic PMF approximations for 16 pulse amplitude modulation (PAM) can be written as:

$$p(x_i = \pm13, \pm15) = 1/2^5$$

$$p(x_i = \pm3, \pm5, \pm7, \pm9, \pm11) = 1/2^4$$
In this work, we consider the BICM model for polar codes, for which gray mapping is found to be optimal [30]. To ensure that the resulting constellation forms a binary reflected gray code, the algorithm from [31] is followed and the mappings are shown in Table I. The PMF and mapping function for 256-QAM is obtained from the outer product of two 16-PAMs. The proposed many-to-one mapping in [11] is an instance of many-to-one PS as originally proposed by Gallager in [32, p. 208].

It can be seen from Table I that in many-to-one mappings multiple bit sequences are mapped to the same symbol. For example, the symbol ‘-1’ has the label ‘111xx’, where ‘x’ represents an ambiguous bit. The bit sequences that will be mapped to e.g., symbol ‘-1’ will be ‘11100’, ‘11101’, ‘11110’ and ‘11111’. Many-to-one maximum a-posteriori probability de-mapping will result in low certainty at the ambiguous bits, and they will be treated as punctured by the FEC decoder. Since puncturing is data dependent and thus unknown by the decoder, we refer to this type of puncturing as implicit puncturing. This effect can also be seen in Fig. 2, where we plot the bit-wise mutual information (MI) of each bit level for many-to-one mapped 16-PAM on an additive-white Gaussian noise (AWGN) channel. The bit-wise MI is found by [33],

\[ I(L; X) = 1 - \frac{1}{N/m} \sum_{k=1}^{N/m} \log_2(1 + e^{-b_k L_k}), \]

where \( N/m \) is the total number of input symbols \( X \) to the channel, \( b_k \) is the \( k \)-th bit in the symbol and \( L \) represents their LLRs after de-modulation. It can be seen that the MI of bit levels 4 (L4) and 5 (L5) does not reach its maximum value i.e., 1 bit/symbol even when the SNR is very high (Fig. 2a). The bit levels 4 and 5 are the ambiguous bit locations in the many-to-one mapper. Due to the many-to-one mappings, most of the bits in bit-level L5 (and some on bit-level L4) do not influence the choice of constellation point. Nevertheless, the de-mapper attempts to estimate them, resulting in low reliability on those positions, as evident from their bit-wise MI (Fig. 2a). The latter is non-zero due to the fact that the bits in L5 for the outer symbols can be de-mapped with high reliability.

Bit-levels L5 and L4 can also be viewed as suffering from random erasures with probability 0.875 and 0.25, respectively. A combination of many-to-one mappings with polar codes will introduce ambiguities in the polar decoder which can affect the performance of polar codes due to its serial nature (error propagation) as explained in Section II. In the next section, we will explain a new method to avoid ambiguities due to the many-to-one mapper and how to achieve shaping gains compared to the uniform mappings.

**IV. POLAR-CODED BICM MODEL WITH PRE-SET BITS**

To improve the performance of polar codes with many-to-one mappings, we propose to use pre-set bits in the polar codes. Pre-set bits are bits that are set to pseudo-random values. As an example, the symbol ‘-1’ has the label ‘111xx’, where ‘x’ represents a punctured bit. The transmitter and receiver for polar codes
with many-to-one mappings and pre-set bits incorporating the BICM model from Fig. 1 are shown in Fig. 3 for 16 pulse amplitude modulation (PAM).

Once the bit reliabilities of the virtual channels are calculated, we first assign the $K$ most reliable bits to the data bits, then we select $S$ bits as the pre-set bits and the rest of the bits $N - K - S$ are selected as the frozen bits. The set $I$ now also contains the indices of pre-set bits. The aim of pre-set bits is to mitigate the effect of ambiguities due to the many-to-one mapper as much as possible. The information bits and the pre-set bits are treated as systematic bits in the code of length $N$. An example of combining many-to-one mapping, as specified in Table I, with a polar code using preset bits and 16-PAM, as depicted in Fig. 3, is presented below. It can be seen from Table I that most of the ambiguous bits appear on the bit level 5 (L5). Thus, the heuristic approach applied in this work is to pre-set $S = N/m$ bits and put them on L5, where $m$ is the bit label length (e.g., for 16-PAM $m = \log_2(16) = 4$ in the non-shaped case and $m = 5$ in the shaped case). The number and position of pre-set bits $S$ which results in best BER performance can be optimised. It can be observed from the example considered in Table I that not all the bits on L5 are ambiguous, and some bits on L4 are ambiguous as well. An optimization can be performed to find out the optimal mapping of pre-set bits to ambiguous bit-levels L4 and L5. The optimization for the number of pre-set bits and their mapping is left as future work.

The binary source vector $u^{N}_N$ now contains $K$ data bits, $S$ pre-set bits and $(N - K - S)$ parity bits as shown in Fig. 3. After the systematic polar encoding, the data, parity and pre-set bits are interleaved separately using a pseudo-random interleaver. It is to be noted that during the encoding, the pre-set bits at first are treated as the information bits so that the systematic encoder will not change their values and the pre-set bits can be easily extracted after encoding. Given the pseudo-random number generator (PRNG) defining the pre-set bits (more in Section IV.A), these become similar to the over-capable bits in shortening [23]. The parity bits placed in bit-level L4 are subjected to implicit puncturing due to many-to-one mappings after encoding. Shaping is based on many-to-one mapping combined with primarily shortening (in L5) and a little puncturing (in L4). Rate-adaptivity is thereafter achieved by varying the information bits $K$ (in L3) (Fig. 3). By choosing different values for $K$ and $S$, practically any rate can be achieved with this BICM model. However, for fixed value of $N$, the number of pre-set bits $S$ sets a limit on the data rate.

An example of mapping in each codeword block with pre-set bits is shown in Fig. 3 for 16-PAM. The data bits are first filled in at the most reliable bit levels in the codeword block, i.e., L1-L3, the parity bits are then placed into the bit level 4 and then level 5 is filled with the pre-set bits. As shown in Fig. 2a, levels 4 and 5 are the ambiguous bit levels and the capacity of bit level 5 is the lowest. Thus only few of the parity bits will be implicitly punctured and the shortened bits/pre-set bits appear on the most ambiguous level i.e., L5. In Fig. 2b, we also plot the MI for the bit levels with pre-set bits. It can be seen from Fig. 2b that the MI of L5 with pre-set bits is zero because L5 is not carrying any useful information, as the decoder reads these pre-set bits from the PRNG sequence. The role of pre-set bits is to eliminate the effect of ambiguities at the demapper. Due to this, there is a loss in average MI as can be seen in Fig. 2c. The MI loss from pre-set bits is small, compared to the gain, obtained by decoding with pre-set bits as side information.

For the example considered in Fig. 3, we have placed the pre-set bits only in the bit-level L5. For a more general case, as stated earlier in this section, the placement of pre-set bits in L4 and L5 can be optimised. In this general case, the pre-set bits along with parity bits might be placed on both ambiguous bit-levels i.e., L4 and L5. In both cases, the pre-set bits from the PRNG sequence can still be considered as the bits that are not transmitted and thus pre-set bits shorten the FEC code.

### A. Generation of Pre-Set Bits

The pre-set bits can be generated by a PRNG at both the transmitter and receiver. A PRNG takes as input a fixed value, called a seed, to generate random bits. The generated sequence by a PRNG is not entirely random and strictly depends on the initial value, i.e., the seed. Thus, the sequence can be reproduced later if the seed is known. We use the index number of each codeword in the sequence of codewords as a seed to a PRNG. The PRNG then produces the pseudo-random pre-set bits for each codeword at the transmitter. By sharing the seed between transmitter and receiver, pre-set bits are therefore available at the receiver provided that synchronization is achieved.

---

**Fig. 3.** Polar-coded BICM transmitter and receiver with many-to-one mapper and pre-set bits. An example for 16-PAM.
B. De-mapping and decoding

At the receiver, the demodulator de-maps the received noisy symbol sequence to their LLRs. With the use of the PRNG, the pre-set bits are available at the receiver; therefore, the pre-set bits are fed to the de-mapper as a-priori information akin to an iterative de-mapper [34]. From Fig. 2c, it can be seen that the MI of bit level 4 (L4) is increased due to the fact that the de-mapper utilizes the a-priori information, available for the pre-set bits on L5 to improve the reliabilities of other levels as well. However, the use of a-priori information of pre-set bits does not change the MI of bit levels L1-L3 (Fig. 2c).

At the decoder, the location of pre-set bits can be found by the set \( I \). We set the values of pre-set bits at these locations to a very high absolute LLR value. The pre-set bits were encoded as part of the data bits. Therefore, the decoder also treats the pre-set bits as data bits to decode the codeword. Since systematic encoding was employed at the transmitter, the pre-set bits can be directly extracted from the decoded codeword and are discarded once the decoding is finished. The CRC is performed only on the \( K \) information bits.

C. Performance

We simulated the transmitter and receiver from Fig. 3 over an AWGN channel. We studied the performance of 256 QAM with non-punctured polar codes, the proposed scheme with pre-set bits and punctured polar codes. Punctured polar codes are used as a benchmark since they are the standard approach to rate-adaptivity, which, however, incurs performance loss [25]. Various input data rates can be achieved by controlling the FEC code rate. For example, an input data rate of 5 bits/symbol for 256-QAM with \( m = 8 \) in non-shaped case (i.e., non-punctured and punctured polar codes) can be achieved with a FEC code rate of \( 5/8 = 0.625 \). For the shaped case with 256-QAM and \( m = 10 \), the FEC code rate is \( 5/10 = 0.5 \) to achieve an input data rate of 5 bits/symbol. The simulated input data rates are 4, 5 and 6 bits/QAM symbol. For fixed \( N \), the FEC code rate can be controlled by \( K \). By varying \( K \), practically any rate can be achieved without changing the value of \( N \). The parameters for tested input data rates for all schemes are shown in Table II. The proposed scheme in Table II uses many-to-one mappings (Table I) to achieve shaping. All the evaluated codes are decoded using CA-SCL decoding with 32-bit CRC and list size 32.

The code length considered here is \( N = 8192 \) and for punctured polar codes the code length is \( N_t = 6554 \) with \( l = 1038 \) punctured bits. The total number of transmitted symbols for one code length in case of punctured polar codes and the proposed scheme are the same i.e., 820 symbols. It is to be noted here that we measure the frame-error rate (FER) and BER after actual decoding is performed. The FER and BER are plotted after observing at least 100 simulated frames in error, which is a reasonable statistic for measuring FER and BER.

The results are shown in Fig. 4. For spectral efficiency \( \eta = \{4,5,6\} \) bits/symbol we observe \{0.5, 0.3, 0.45\} dB of shaping gains in BER and \{0.6, 0.4, 0.5\} dB of shaping gains in FER compared to the non-punctured polar codes. It can be seen from Fig. 4 that the shaping gains compared to the punctured polar codes are significantly improved. For FER, the shaping gains with the proposed scheme for \( \eta = \{4,5,6\} \) bits/symbol are \{1.1, 1.3, 1.2\} dB w.r.t the punctured polar codes. Similarly, for BER the shaping gains with the proposed scheme for \( \eta = \{4,5,6\} \) bits/symbol are \{0.9, 1.2, 1.2\} dB compared to the punctured polar codes.

V. COMPARISON WITH OTHER SCHEMES

We compare the BER and FER performance of standard polar codes and DVB-S2 LDPC codes [35] in Fig. 5 for \( \eta = 6 \) bits/symbol and 256-QAM. The simulation setup is the same as in Section IV.C, except that the code length used here is 16200 bits for LDPC codes and 16384 bits for polar codes. The number of belief propagation (BP) iterations for LDPC codes is 32 and the list size for polar codes is 32. LDPC codes do not need CRC for decoding. For polar codes, the CRC length is 32, and the overhead of CRC is included in the simulation results.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>( \eta ) [bits/symbol]</th>
<th>Data rate [Gbps/channel]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4, 6, 8, 10</td>
<td>4, 5, 6, 8, 10, 12, 16</td>
</tr>
<tr>
<td>Non-punctured polar code</td>
<td>N=8192, K=4096</td>
<td>N=8192, K=4608</td>
</tr>
<tr>
<td>Punctured polar code</td>
<td>N=6554, K=3277</td>
<td>N=6554, K=3686</td>
</tr>
<tr>
<td>Proposed scheme</td>
<td>S=1640</td>
<td>S=1640</td>
</tr>
</tbody>
</table>

Fig. 4. AWGN Simulation: Performance comparison of non-punctured polar codes and punctured polar codes with the proposed scheme. (a) FER. (b) BER. ⊗ Punctured polar code. ⊚ Non-Punctured polar code. ⊳ Proposed Scheme.

TABLE II

| Scheme | \( \eta \) [bits/symbol], Data rate [Gbps/channel] |
|--------|-------------------------|--------------------------|
|        | 4, 6, 8, 10             | 4, 5, 6, 8, 10, 12, 16   |
| Non-punctured polar code | N=8192, K=4096 | N=8192, K=4608 |
| Punctured polar code | N=6554, K=3277 | N=6554, K=3686 |
| Proposed scheme | S=1640 | S=1640 |
It can be seen in Fig. 5 that the performance of uniform polar codes is very similar to uniform LDPC codes. Similar results for polar codes were reported in [36]. Polar codes perform 0.45 dB better than LDPC codes using the proposed scheme at BER of $10^{-15}$ as shown in Fig. 5(b). It can also be observed that the DVB-S2 LDPC codes suffer from an error floor at high SNR while on the other hand polar codes show no error floor [15]. To overcome the error-floor of the LDPC codes, a hard-decision FEC code at the expense of extra overhead can be used, e.g., [37], [38]. Carefully designed LDPC codes can push the error floor below $10^{-15}$ without the use of outer hard-decision codes [39].

We also plot the performance of DVB-S2 LDPC codes with PAS as in [7]. In the case of the PAS, we use a 5/6 DVB-S2 LDPC code with its parameters given in [7, Table VIII]. The total number of symbols for CCDM are $n_c = N/\log_2(16) = 4050$, which gives $K_{dm} = 11016$ shaped amplitude bits, and $K_a = 1170$ bits are used for sign bits along with parity bits, where $\gamma = 1-(1-R) \times m$. One complex QAM symbol is obtained by mapping two real amplitude shift keying (ASK) symbols. It can be seen that the performance of PAS is 0.6 dB better than the proposed scheme at FER of $10^{-3}$. The complexity of PAS is increased due to the use of CCDM at the transmitter and receiver, whereas the complexity increase due to the proposed scheme with pre-set bits is negligible. We also compare the BER performance of PAS before inverse CCDM (preCCDM) and after inverse CCDM (postCCDM) in Fig. 5(b). It can be seen that the postCCDM BER is orders of magnitude higher than the preCCDM BER due to the error-propagation of the de-matcher, also addressed in [40]. Most modern communication systems rely on FER, and in this perspective, the bit-error propagation due to inverse CCDM is of less concern. Furthermore, new DM architectures exist that operate with shorter lengths and thus less pronounced error-propagation [41]. However, BER is used as a figure of merit in optical communication, where BER of $10^{-15}$ is typically desired. Therefore in this work, we consider both BER and FER as figures of merit.

For many-to-one de-mappers, generally more bits per symbol needs to be de-mapped, which slightly increases the complexity of the de-mapper [11]. The complexity of QAM de-mappers scales linearly with $m$ [42], and is negligible w.r.t. the complexity of powerful FEC decoders. As more bits per symbol are de-mapped in many-to-one mappings, for a specified SE the code requires more parity bits compared to the uniform mappings, when a higher modulation order is chosen. Thus, the decoder needs to decode more parity bits in contrast to the uniform mappings. The receiver does not need to carry out additional iterations between de-mapper and decoder which greatly simplifies the decoder implementation compared to the BICM iterative de-mapping schemes [4] [11].

In [43], a multi-level polar coding scheme combined with PAS (PC-PAS) has been studied for short packets. However, this combination increases the latency and complexity of the transmitter and receiver due to the multi-level decoding of polar codes and the use of CCDM. A similar scheme to PC-PAS was proposed in [10] for polar codes. The scheme in [10] uses a SCL decoder to generate shaping bits every time a new codeword is required. Thus, this scheme increases the complexity at the transmitter due to the use of the SCL decoder. Both schemes in [43] and [10] might perform better in terms of BER compared to the proposed scheme. Nonetheless, the proposed scheme provides a low-complexity practical solution with comparable performance to the other schemes.

Moreover, the scheme in [10] is inherent to polar codes and cannot be extended to other FEC codes. The proposed scheme can also be used with any other systematic FEC code. We leave the combination of the proposed scheme with other FEC codes such as turbo codes and LDPC codes as a part of our future work.

In the next sections we present an experimental study of the AWGN simulated polar-coded transmitter and receiver with pre-set bits from Fig. 3 for an optical transmission system.

VI. EXPERIMENTAL SETUP

The experimental setup with pre-set bits and many-to-one mappings is shown in Fig. 6. At the transmitter side, 10 optical sub-carriers placed on a 8.5 GHz frequency grid are generated by two frequency-locked 5-line electro-optic frequency combs with a free spectral range of 17 GHz. The two 5-line optical frequency combs are seeded by the split output of an NKT Koheras fiber laser (sub-kHz linewidth). Each output is phase modulated and flattened by a waveshaper (WSS). The comb lines are then modulated with electrical data using two IQ modulators (one for each comb), and interleaved in a 3-dB coupler. At the transmitter output, the modulated carriers are amplified with an Erbium-doped fiber amplifier (EDFA), and polarization-multiplexing is emulated with a polarization delay stage (de-correlation of 953 symbol periods). The data signals
for driving the modulators are generated by encoding pseudo random bit sequences. The encoded bits are interleaved and mapped to 256 QAM symbols. A 5% overhead of 16 QAM pilot symbols is added to assist the adaptive equalization and the carrier phase recovery at the receiver. The symbols are pulse shaped with a root-raised cosine (RRC) filter with 401 taps and a roll-off factor of 0.05. Two de-correlated sequences containing 32 frames of FEC encoded data are loaded in the arbitrary waveform generator (AWG), which then drives each quadrature of the two IQ modulators.

The transmission link under test is based on a recirculating fiber loop composed of four 50 km spans of submarine ultralarge area fiber (OFS SCUBA fiber), with all the propagation losses compensated by Raman amplification in backward pumping configuration. Acousto-optic modulators (AOM) are used as switches for the loop control, and an in-loop EDFA compensates for the internal loop losses due to couplers, isolators, filters and switches. As the wavelength division multiplexed (WDM) signals exit the loop after a set number of turns, they are pre-amplified and received using a coherent setup with a separate fiber laser employed as the local oscillator (LO). All the measurements and the subsequent receiver digital signal processing (DSP) chain is conducted for channel five.

The measured channel is sampled at 40 GS/s with a real-time oscilloscope (5 GHz of receiver bandwidth). In the offline DSP, the acquired signal passes through a front-end compensation stage, resampling, chromatic dispersion (CD) compensation, low-pass filtering, timing recovery, \( T_s / 2 \)-fractionally spaced pilot-based adaptive equalization (85 taps), and pilot-based carrier frequency and phase recovery with a moving average phase estimator. In the end, the de-mapper operates an auxiliary channel as in [11], [44], with variance estimated as \( E_k [y_k - x_k] [y_k] \), and the received data sequence is sent to the FEC decoder. To estimate achievable information rate (AIR), we use an auxiliary function,

\[
q(X|Y) \propto p(X)q(Y|X),
\]

where \( X \) and \( Y \) represent the input and output samples of the channel. The auxiliary channel \( q(X|Y) \) is used to estimate an upper bound on the entropy \( H(X|Y) = -1/K \sum_{k=1:K} \log_2(q(x_k|y_k) \geq H(X|Y)) \) and thereby a lower bound on the mutual information \( I(X; Y) \). This lower bound is the AIR. The auxiliary distribution \( q(Y|X) \) in Eq. (5) is Gaussian with a variance, which is estimated after transmission.

**VII. BACK-TO-BACK TRANSMISSION RESULTS**

We tested the input data rates \{4, 4.5, 5, 5.5, 6\} bits/symbol corresponding to \{60.8, 68.4, 76, 83.6, 91.2\} Gbps per WDM channel. All three schemes i.e., non-punctured polar codes, punctured polar codes and proposed scheme are tested experimentally. The parameters of all three schemes are shown in the Table II. We refer to the experimental points with no observed errors after FEC decoding as error-free. We experimentally evaluated a total of 60 frames i.e., \( 5 \times 10^5 \) bits for each channel.

The performance of channel five in optical back-to-back configuration for \( \eta = 5 \) bits/symbol is shown in Fig. 7. We show the AIR versus optical SNR (OSNR) in Fig. 7(a) for all three polar-coded schemes for \( \eta = 5 \) bits/symbol. It can be seen from Fig. 7(a) that the shaping gain w.r.t non-punctured polar codes and punctured polar codes is about 1.5 dB in OSNR.

We also show the BER performance versus OSNR in Fig. 7(b). It can be seen that the punctured polar codes perform about 1.2 dB worse than the non-punctured polar codes at \( \eta = 5 \) bits/symbol for the fixed BER. This puncturing loss can be offset by the combination of polar codes and many-to-one mappings as proposed in [25]. The shaping gains with the proposed scheme in this work are improved to about 1.7 dB at fixed BER compared to the punctured polar codes. The proposed scheme provides shaping gains of up to 0.55 dB compared to the non-punctured polar codes at the same transmission rate. The gains are similar to the AWGN simulation results shown in Fig. 4.

**VIII. TRANSMISSION RESULTS**

The system performance is further investigated over a transmission link using a recirculating loop as explained in the Section VI. We apply all three polar-coded schemes with their parameters given in Table II.

We performed the WDM measurements for the channel five for different rates by varying the distances for each rate. The considered system parameters are given in Table III. The
achieved error-free data rate and the AIR are shown in Fig. 8, at the respective optimal input power, as a function of the link distance. The optimal launch power is determined by varying the input power over a wide range of input powers. We then tested the transmission rates 4 bits/symbol to 6 bits/symbol with a step size of 0.5 for distances of 200 km up to 4000 km. We selected an input data rate and kept on increasing the distance until the error-free transmission was not possible with that data rate. It can be seen from Fig. 8 that the proposed scheme provides a steady increase of 200 km of distance for all tested data rates compared to the non-punctured polar codes. Rate-adaptivity is achieved using implicit puncturing, shortening and by varying the number of information bits in the polar codes and around 1.7 dB of shaping gains are reported compared to the non-punctured polar codes. The proposed method allows one to select any operating point without changing the modulation format and the underlying channel code and at the same time providing performance gains. Thus, the proposed system does not require additional hardware for rate-adaptivity. Higher-order modulation formats and baud rates can also be realized by the direct extension of the proposed system. Robust performance was achieved using a single forward-error-correction code over a range of 200 km to 4000 km. The maximum transmission reach is increased by 200 km compared to the uniform polar codes and by 400 km compared to the conventionally punctured polar codes.

IX. CONCLUSIONS

We presented a low-complexity rate-adaptive probabilistically shaped bit-interleaved coded modulation system and experimentally evaluated its performance on a 10-channel wavelength division multiplexed dual polarization coherent transmission system at 8 GBaud/channel. Improved shaping gains, w.r.t. the non-punctured polar codes, are obtained using many-to-one constellation shaping with pre-set bits in the polar codes. The proposed method provides around 0.55 dB of shaping gains compared to the non-punctured polar codes. Rate-adaptivity is achieved using implicit puncturing, shortening and by varying the number of information bits in the polar codes and around 1.7 dB of shaping gains are reported compared to the conventionally punctured polar codes.

ACKNOWLEDGMENT

The authors would like to thank OFS Denmark for providing the SCUBA fiber.

REFERENCES
