

# Multilevel Polar-Coded Modulation

Mathis Seidl, Andreas Schenk, Clemens Stierstorfer, and Johannes B. Huber

Institute of Information Transmission, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Email: {seidl, schenk, clemens, jbhuber}@LNT.de.

**Abstract**—A framework is proposed that allows for a joint description and optimization of both binary polar coding and the multilevel coding (MLC) approach for  $2^m$ -ary digital pulse-amplitude modulation (PAM). The conceptual equivalence of polar coding and multilevel coding is pointed out in detail. Based on a novel characterization of the channel polarization phenomenon, rules for the optimal choice of the bit labeling in this coded modulation scheme employing polar codes are developed. Simulation results for the AWGN channel are included.

## I. INTRODUCTION

Polar codes [1] are known as a low-complexity binary coding scheme that provably approaches the capacity of arbitrary symmetric binary-input discrete memoryless channels (B-DMCs). The generalization to  $q$ -ary channels ( $q > 2$ ) has been the subject of various works, cf., e.g. [2]. Though, the topic of *polar-coded modulation*, i.e., the combination of  $2^m$ -ary digital PAM modulation and binary polar codes for increased spectral efficiency, has hardly been addressed so far. In [3], a transmission scheme for polar codes with bit-interleaved coded modulation (BICM) [4] has been proposed, focussing on the interleaver design.

In this paper, we consider the multilevel coding (MLC) construction [5], [6] for memoryless channels like the AWGN channel (no fading).

It has been observed (e.g., [7]) that the MLC approach is closely related to that of polar coding on a conceptual level. Based on these similarities, we propose a framework that allows us to completely describe both polar coding and  $2^m$ -ary modulation in a unified context as certain channel transforms. This unified description enables us to design optimized constellation-dependent coding schemes for MLC.

The paper is organized as follows: In Sec. II, the framework for a joint description of polar coding and  $2^m$ -ary PAM modulation is developed. This framework is then used for describing the polar coding construction in Sec. III, leading to a novel interpretation of the polarization phenomenon. The optimum combination of binary polar coding and multilevel coding is discussed in Sec. IV, followed by simulation results for the AWGN channel in Sec. V.

## II. CHANNEL TRANSFORMS

### A. Sequential Binary Partitions

Let  $W : \mathcal{X} \rightarrow \mathcal{Y}$  be a discrete, memoryless channel (DMC) with input symbols  $x \in \mathcal{X}$  (alphabet size  $|\mathcal{X}| = 2^k$ ), output symbols  $y \in \mathcal{Y}$  from an arbitrary alphabet  $\mathcal{Y}$ , and mutual

information  $I(X; Y)$ .<sup>1</sup> We define an order- $k$  *sequential binary partition* ( $k$ -SBP)  $\varphi$  of  $W$  to be a channel transform

$$\varphi : W \rightarrow \{B_\varphi^{(0)}, \dots, B_\varphi^{(k-1)}\} \quad (1)$$

that maps  $W$  to an *ordered* set of  $k$  binary-input DMCs (B-DMCs) which we will refer to as *bit channels*. For any given  $W$ , such a  $k$ -SBP is characterized by a binary labeling rule  $\mathcal{L}_\varphi$  that maps binary  $k$ -tuples bijectively to the  $2^k$  input symbols  $x \in \mathcal{X}$ :

$$\mathcal{L}_\varphi : [b_0, b_1, \dots, b_{k-1}] \in \{0, 1\}^k \mapsto x \in \mathcal{X}. \quad (2)$$

The number of possible labelings equals  $(2^k!)$ .

Each bit channel  $B_\varphi^{(i)}$  ( $0 \leq i < k$ ) of a  $k$ -SBP is supposed to have knowledge of the output of  $W$  as well as of the values transmitted over the bit channels of smaller indices  $B_\varphi^{(0)}, \dots, B_\varphi^{(i-1)}$ . Thus, we have

$$B_\varphi^{(i)} : \{0, 1\} \rightarrow \mathcal{Y} \times \{0, 1\}^i, \quad 0 \leq i < k. \quad (3)$$

The mutual information between channel input and output of  $B_\varphi^{(i)}$  assuming equiprobable input symbols is therefore given by

$$I(B_\varphi^{(i)}) := I(B_i; Y | B_0, \dots, B_{i-1}) \quad (4)$$

which we will refer to as the (*symmetric*) *bit channel capacity* of  $B_\varphi^{(i)}$ . (If  $W$  is a symmetric channel, this value in fact equals the channel capacity.) The mutual information of  $W$  is preserved under the transform  $\varphi$ , i.e.,

$$\sum_{i=0}^{k-1} I(B_\varphi^{(i)}) = I(X; Y) \quad (5)$$

which directly follows from the well-known chain rule of mutual information.

Considering polar-coded modulation, we show that the code construction can be described by SBPs. We are particularly interested in two properties of SBPs, namely the mean value and the variance of the bit channel capacities, defined respectively as

$$M_\varphi(W) := \frac{1}{k} \sum_{i=0}^{k-1} I(B_\varphi^{(i)}) = \frac{1}{k} I(X; Y) \quad (6)$$

$$V_\varphi(W) := \frac{1}{k} \sum_{i=0}^{k-1} I(B_\varphi^{(i)})^2 - M_\varphi(W)^2. \quad (7)$$

Clearly, from (5) the mean value  $M_\varphi(W)$  in fact depends only on the channel  $W$ , rather than on the transform  $\varphi$ . It represents

<sup>1</sup>A short remark on the notation: Channels are denoted by sans serif fonts, capital roman letters stand for random variables while boldfaced symbols denote vectors or matrices.

the average (symmetric) capacity of  $W$  per transmitted binary symbol.

The variance of an SBP  $\varphi$  is upper-bounded by

$$V_\varphi(W) \leq M_\varphi(W)(1 - M_\varphi(W)) \quad (8)$$

with equality only iff all  $I(B_\varphi^{(i)})$  are either 0 or 1. This follows from

$$\begin{aligned} V_\varphi(W) &= \frac{1}{k} \sum_{i=0}^{k-1} I(B_\varphi^{(i)})^2 - M_\varphi(W)^2 \\ &\leq \frac{1}{k} \sum_{i=0}^{k-1} I(B_\varphi^{(i)}) - M_\varphi(W)^2 \\ &= M_\varphi(W)(1 - M_\varphi(W)) . \end{aligned} \quad (9)$$

and  $0 \leq I(B_\varphi^{(i)}) \leq 1$  for all  $0 \leq i < k$ . Note that this upper bound does not depend on the particular labeling  $\mathcal{L}_\varphi$  but only on the channel  $W$ .

An important subset of  $k$ -SBPs is formed by those transforms whose labeling rules are described by binary bijective linear mappings. Let  $W = (B_0 \times \dots \times B_{k-1})$  be a vector channel of  $k$  independent arbitrary B-DMCs  $B_0, \dots, B_{k-1}$ . Then, we call the  $k$ -SBP

$$\varphi : (B_0 \times \dots \times B_{k-1}) \rightarrow \{B_\varphi^{(0)}, \dots, B_\varphi^{(k-1)}\} \quad (10)$$

a *linear  $k$ -SBP* if its labeling rule is given by

$$\mathcal{L}_\varphi : \mathbf{b} \in \mathbb{F}_2^k \mapsto \mathbf{b} \cdot \mathbf{A}_\varphi \in \mathbb{F}_2^k . \quad (11)$$

with  $\mathbf{b} := [b_0, b_1, \dots, b_{k-1}]$  and  $\mathbf{A}_\varphi$  being an invertible binary  $(k, k)$  matrix. Clearly, the number of possible linear  $k$ -SBPs equals the number of non-singular binary  $(k, k)$  matrices and is significantly smaller than that of general  $k$ -SBPs.

### B. Product Concatenation of SBPs

Under certain conditions, it is possible to concatenate two (or more) SBPs in a product form. Let

$$\varphi : W \rightarrow \{B_\varphi^{(0)}, \dots, B_\varphi^{(k_1-1)}\} \quad (12)$$

be an arbitrary  $k_1$ -SBP and

$$\psi : (B_0 \times \dots \times B_{k_2-1}) \rightarrow \{B_\psi^{(0)}, \dots, B_\psi^{(k_2-1)}\} \quad (13)$$

a  $k_2$ -SBP that takes a vector channel of  $k_2$  independent B-DMCs  $B_0, \dots, B_{k_2-1}$  as an input. Each of the vector channels  $(B_\varphi^{(i)})^{k_2}$  – obtained by taking  $k_2$  independent instances of  $B_\varphi^{(i)}$  – can be partitioned by  $\psi$ . Thus,  $\varphi$  and  $\psi$  may be concatenated by considering the vector channel  $W^{k_2}$ , leading to a product SBP of order  $k_1 k_2$ :

$$\varphi \otimes \psi : W^{k_2} \rightarrow \{B_{\varphi \otimes \psi}^{(0)}, \dots, B_{\varphi \otimes \psi}^{(k_1 k_2 - 1)}\} . \quad (14)$$

Here, the bit channels of  $\varphi \otimes \psi$  are given by

$$B_{\varphi \otimes \psi}^{(k_2 i + j)} : \{0, 1\} \rightarrow \mathcal{Y}^{k_2} \times \{0, 1\}^{k_2 i + j} \quad (15)$$

with symmetric capacities

$$I(B_{\varphi \otimes \psi}^{(k_2 i + j)}) = I(B_{k_2 i + j}; Y_0, \dots, Y_{k_2 - 1} | B_0, \dots, B_{k_2 i + j - 1}) \quad (16)$$

such that

$$\frac{1}{k_2} \sum_{j=0}^{k_2-1} I(B_{\varphi \otimes \psi}^{(k_2 i + j)}) = I(B_\varphi^{(i)}) \quad (17)$$

for all  $0 \leq i < k_1$  and  $0 \leq j < k_2$ . We remark that the product transform  $\varphi \otimes \psi$  is completely determined in a unique way by the individual SBPs  $\varphi$  and  $\psi$  since their bit channels imply a fixed order.

The product concatenation of SBPs does not influence the mean value of the bit channel capacities

$$M_{\varphi \otimes \psi}(W^{k_2}) = M_\varphi(W) \quad (18)$$

due to the chain rule of mutual information. However, the variance of the bit channel capacities increases. It is given by the sum of the variance of the first transform and the averaged variance of the second transform around the bit channel capacities of the first one:

$$V_{\varphi \otimes \psi}(W^{k_2}) = V_\varphi(W) + \frac{1}{k_1} \sum_{i=0}^{k_1-1} V_\psi(B_\varphi^{(i)}) . \quad (19)$$

If  $\varphi$  and  $\psi$  are linear SBPs with labeling rules specified by  $\mathbf{A}_\varphi$  and  $\mathbf{A}_\psi$ , respectively, then their product  $\varphi \otimes \psi$  is again a linear  $k_1 k_2$ -SBP with labeling rule

$$\mathcal{L}_{\varphi \otimes \psi} : \mathbf{b} \in \mathbb{F}_2^{k_1 k_2} \mapsto \mathbf{b} \cdot \mathbf{P}_{k_1, k_2} \cdot (\mathbf{A}_\psi \otimes \mathbf{A}_\varphi) . \quad (20)$$

Here,  $\mathbf{A}_\psi \otimes \mathbf{A}_\varphi$  denotes the Kronecker product of  $\mathbf{A}_\psi$  and  $\mathbf{A}_\varphi$ .  $\mathbf{P}_{k_1, k_2}$  is the  $(k_1 k_2, k_1 k_2)$  permutation matrix that maps the  $(k_2 i + j)$ -th component of the vector  $\mathbf{b}$  to position  $i + k_1 j$  (for all  $0 \leq i < k_1, 0 \leq j < k_2$ ).

## III. POLAR CODES

Polar codes, as introduced by Arıkan [1], have been shown to be a channel coding construction that provably achieves the symmetric capacity of arbitrary binary-input discrete memoryless channels (B-DMCs) under low-complexity encoding and successive cancellation (SC) decoding. For sake of simplicity, we focus on Arıkan's original construction in this paper; the generalization to polar codes based on different kernels (as considered, e.g., in [8]) is straightforward. Furthermore, we restrict our considerations to the SC decoding algorithm as in [1]; though, our results regarding the code construction are also valid for other (better performing) decoders that are based on the SC algorithm, as, e.g., list decoding [9].

### A. Code Construction

Let  $B : \{0, 1\} \rightarrow \mathcal{Y}$  be a B-DMC and  $I(B)$  its symmetric capacity, i.e., the mutual information of  $B$  assuming equiprobable binary input symbols. The encoding operation for a polar code of length  $N$  may be described by multiplication of a binary length- $N$  vector  $\mathbf{u}$  – containing the information symbols as well as some symbols with fixed values (so-called *frozen symbols*) that do not carry any information – with a generator matrix  $\mathbf{G}_N$  that is defined by the recursive relation

$$\mathbf{G}_N = \mathbf{B}_N \mathbf{F}_N , \quad \mathbf{F}_{2N} = \mathbf{F}_2 \otimes \mathbf{F}_N , \quad \mathbf{F}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad (21)$$

where  $N$  is a power of two and  $\otimes$  again denotes the Kronecker product.  $\mathbf{B}_N$  denotes the  $(N, N)$  bit-reversal permutation matrix [1]. Encoding takes place in the binary field  $\mathbb{F}_2$ . The resulting codeword  $\mathbf{c} = \mathbf{u}\mathbf{G}_N$  is then transmitted in  $N$  time steps over the channel  $\mathbf{B}$ .

The code construction is based on a channel combining and channel splitting operation [1] that may be represented as a linear 2-SBP

$$\pi : \mathbf{B}^2 \rightarrow \{\mathbf{B}_\pi^{(0)}, \mathbf{B}_\pi^{(1)}\} \quad (22)$$

that partitions the vector channel  $\mathbf{B}^2$ , i.e., two independent and identical instances of  $\mathbf{B}$ , into two bit channels

$$\mathbf{B}_\pi^{(0)} : \{0, 1\} \rightarrow \mathcal{Y}^2 \quad (23)$$

$$\mathbf{B}_\pi^{(1)} : \{0, 1\} \rightarrow \mathcal{Y}^2 \times \{0, 1\}.$$

The labeling rule is given by

$$\mathcal{L}_\pi : \mathbf{u} = [u_0, u_1] \in \{0, 1\}^2 \mapsto \mathbf{u} \cdot \mathbf{G}_2 \in \{0, 1\}^2. \quad (24)$$

Since the average capacity per binary symbol does not change under an SBP, we denote the mean value of the bit channel capacities of  $\pi$  by  $I(\mathbf{B})$  instead of  $M_\pi(\mathbf{B})$  in the following.

The construction of a polar code of length  $N = 2^n$  may be equivalently represented by the  $n$ -fold product concatenation of  $\pi$  as defined in the preceding section. This follows easily from [1] by comparison of the corresponding permutation matrices. The resulting SBP  $\pi^n$  partitions the vector channel  $\mathbf{B}^N$

$$\pi^n : \mathbf{B}^N \rightarrow \{\mathbf{B}_{\pi^n}^{(0)}, \dots, \mathbf{B}_{\pi^n}^{(N-1)}\} \quad (25)$$

into  $N$  bit channels

$$\mathbf{B}_{\pi^n}^{(i)} : \{0, 1\} \rightarrow \mathcal{Y}^N \times \{0, 1\}^i \quad (26)$$

( $0 \leq i < N$ ) with symmetric capacities

$$I(\mathbf{B}_{\pi^n}^{(i)}) := I(U_i; Y_0, \dots, Y_{N-1} | U_0, \dots, U_{i-1}). \quad (27)$$

Therefore, the transmission of each source symbol  $u_i$  can be described by its own bit channel  $\mathbf{B}_{\pi^n}^{(i)}$ . The output of each channel  $\mathbf{B}_{\pi^n}^{(i)}$  depends on the values of the symbols of lower indices  $u_0, \dots, u_{i-1}$ . Thus, the channels  $\mathbf{B}_{\pi^n}^{(i)}$  imply a specific decoding order.

For data transmission only the bit channels with highest capacity are used, referred to as *information channels*. The data transmitted over the remaining bit channels (so-called *frozen channels*) are fixed values known to the decoder. By this means, the code rate can be chosen in very small steps of  $1/N$  without the need for changing the code construction – a property especially useful for polar-coded modulation (cf., Sec. IV-B).

In order to select the optimal set of frozen channels, the values of the capacities  $I(\mathbf{B}_{\pi^n}^{(i)})$  are required. These can either be obtained by simulation or by density evolution [10].

## B. Successive Decoding

Upon receiving a vector  $\mathbf{y}$  – being a noisy version of the codeword  $\mathbf{c}$  resulting from transmission over the channel  $\mathbf{B}$  – the information bits  $u_i$  can be estimated successively for  $i = 0, \dots, N - 1$ . Here, information combining [11] of reliability

values obtained from the channel output  $\mathbf{y}$  is performed instead of  $\mathbb{F}_2$  arithmetics as in the encoding process.

The successive cancellation (SC) decoding algorithm [1] for polar codes generates estimates on the information symbols  $\hat{u}_i$  (transmitted over  $\mathbf{B}_{\pi^n}^{(i)}$ ) one after another, making use of the already decoded symbols  $\hat{u}_0, \dots, \hat{u}_{i-1}$ . We denote the probability that an erroneous decision is made at index  $i$  given the previous decisions have been correct, by  $p_e(\mathbf{B}_{\pi^n}^{(i)})$ . Thus, the word error rate for SC decoding ( $\text{WER}_{\text{SC}}$ ) is given by

$$\text{WER}_{\text{SC}} = 1 - \prod_{i \in \mathcal{A}} (1 - p_e(\mathbf{B}_{\pi^n}^{(i)})) \quad (28)$$

where  $\mathcal{A}$  denotes the set of indices of the information channels.

## C. Variance of the Bit Channel Capacities

With increasing block length, the set of bit channels  $\mathbf{B}_{\pi^n}^{(i)}$  shows a polarization effect in the sense that the capacity  $I(\mathbf{B}_{\pi^n}^{(i)})$  of almost each bit channel is either near 0 or near 1. The fraction of bit channels not being either completely noisy or completely noiseless tends to zero [1].

In the following, we show that this polarization effect may be represented by the sequence of variances of the respective polar codes' bit channel capacities for increasing block length. The variance of the bit channel capacities of a length- $N$  polar code around their mean value  $I(\mathbf{B})$  is given by

$$V_{\pi^n}(\mathbf{B}^N) = \frac{1}{N} \sum_{i=0}^{N-1} I(\mathbf{B}_{\pi^n}^{(i)})^2 - I(\mathbf{B})^2. \quad (29)$$

Using (19), we notice that the sequence of variances increases monotonously as the block length gets larger, i.e.,

$$V_{\pi^{n+1}}(\mathbf{B}^{2N}) \geq V_{\pi^n}(\mathbf{B}^N). \quad (30)$$

Furthermore, from (9) the sequence  $\{V_{\pi^n}(\mathbf{B}^N)\}_{n \in \mathbb{N}}$  is upper-bounded by

$$V_{\pi^n}(\mathbf{B}^N) \leq I(\mathbf{B})(1 - I(\mathbf{B})) \quad (31)$$

for all  $n \in \mathbb{N}$ . According to (9), this maximum variance can be only achieved iff all bit channel capacities  $I(\mathbf{B}_{\pi^n}^{(i)})$  are either 0 or 1, which obviously corresponds to the state of perfect polarization. As shown by Arıkan [1], the latter is asymptotically approached while the block length  $N$  goes to infinity; therefore, we have

$$\lim_{n \rightarrow \infty} V_{\pi^n}(\mathbf{B}^N) = I(\mathbf{B}) \cdot (1 - I(\mathbf{B})). \quad (32)$$

Although we have not yet been able to establish an explicit relation between bit channel capacity variance and code error performance, one would intuitively expect that increasing the variance by a careful code design should correspond to a sharper polarization of the bit channels and therefore should lead to better performing polar codes in terms of word error rate or bit error rate.

Fig. 1 depicts the variance of the bit channel capacities for polar codes of various block lengths constructed over the BEC channel as a function of its capacity. The converging behaviour for increasing block length  $N$  towards the maximum achievable variance (black line) can clearly be observed.

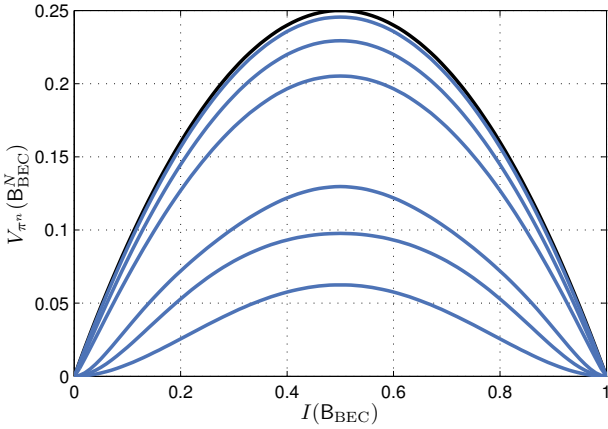


Fig. 1. Bit channel variance for polar codes over a BEC channel  $B_{BEC}$ , block length  $N = 2^n$ ,  $n = 1, 2, 3, 8, 12, 20$ . Black: upper bound on the variance.

#### IV. MULTILEVEL POLAR CODING

We now consider the conventional discrete-time equivalent system model of  $M$ -ary digital pulse-amplitude modulation (PAM) –  $M = 2^m$  being a power of 2 – with signal constellations of real-valued signal points (ASK) or of complex-valued signal points (PSK, QAM etc.) over a memoryless channel  $W$ , e.g., the AWGN channel.

From an information-theoretic point of view, an optimum combination of binary coding and  $M$ -ary modulation follows the multilevel coding (MLC) principle [5], [6].

##### A. Multilevel Coding

In the MLC approach, the  $M$ -ary channel  $W$  is partitioned into  $m$  bit channels (also called *bit levels*) by means of an  $m$ -SBP

$$\lambda : W \rightarrow \{B_{\lambda}^{(0)}, \dots, B_{\lambda}^{(m-1)}\}. \quad (33)$$

The mapping from binary labels to amplitude coefficients is specified by the labeling rule  $\mathcal{L}_{\lambda}$ .

Channel coding is implemented in the MLC setup by using binary *component codes* [6] for each of the bit levels  $B_{\lambda}^{(i)}$  individually with correspondingly chosen code rates  $R_i$ . The overall rate (bits per transmission symbol) is given as the sum  $R = \sum_{i=0}^{m-1} R_i$ . The receiver then performs multi-stage decoding (MSD), i.e., it computes reliability information for decoding of the first bit level which are passed to the decoder of the first component code. The decoding results are used for demapping and decoding of the next bit level, and so on.

The mutual information between the channel input and channel output of  $W$  assuming equiprobable source symbols is also referred to as the coded modulation [6], or constellation-constraint, capacity  $C_{cm}(W)$ . It is related to the average capacity per binary symbol (6) of  $W$  by

$$C_{cm}(W) := I(X; Y) = \sum_{i=0}^{m-1} I(B_{\lambda}^{(i)}) = m \cdot M_{\lambda}(W). \quad (34)$$

Since  $\lambda$  is an SBP, the coded modulation capacity does not depend on the specific labeling rule  $\mathcal{L}_{\lambda}$ .

A potential drawback of the MLC approach for practical use lies in the necessity for using several (relatively short)

component codes with varying code rates for the particular bit levels. According to the capacity rule [6], the code rate  $R_i$  for the  $i$ -th level should match the bit level capacity  $R_i = I(B_{\lambda}^{(i)})$ . Since these capacities vary significantly for the different levels, for MLC channel codes are preferred, that allow for a very flexible choice of the code rate.

##### B. Multilevel Polar Coding

A *multilevel polar code* of length  $mN$ , i.e., a multilevel code with length- $N$  component polar codes over an  $M$ -ary constellation, is obtained by the order- $mN$  concatenation of the  $m$ -SBP  $\lambda$  of MLC and the  $N$ -SBP  $\pi^n$  of the polar code:

$$\lambda \otimes \pi^n : W^N \rightarrow \{B_{\lambda \otimes \pi^n}^{(0)}, \dots, B_{\lambda \otimes \pi^n}^{(mN-1)}\} \quad (35)$$

as defined in (14). The encoding process for this multilevel polar code is described by the generator matrix

$$P_{m,N} \cdot (G_N \otimes I_m) \quad (36)$$

with  $P_{m,N}$  as in (20), followed by labeling and mapping to the  $N$  transmit symbols as defined by  $\lambda$ . Here,  $I_m$  denotes the  $(m, m)$  identity matrix.

We remark that the selection of frozen channels – and thus, the rate allocation – is done in exactly the same way as for a usual binary polar code by determining the symmetric capacities  $I(B^{(i)})$  ( $0 \leq i < mN$ ) and choosing the most reliable bit channels for data transmission. Therefore, the explicit application of a rate allocation rule to the particular component codes – like considered in the original MLC approach [6] – is not needed in case of multilevel polar codes. However, it has been shown [12] that the rate allocations obtained by this method basically equal those obtained from the capacity rule.

According to (19), the variance of the bit channels of a multilevel polar code with length- $N$  component codes is given by

$$V_{\lambda \otimes \pi^n}(W) = V_{\lambda}(W) + \frac{1}{m} \sum_{i=0}^{m-1} V_{\pi^n}(B_{\lambda}^{(i)}). \quad (37)$$

Thus, the SBP  $\lambda$  – that represents the modulation step – may be seen as the first polarization step of a multilevel polar code. From this representation, it is clear that  $\lambda$  should be chosen such that it maximizes the term (37).

In this approach, both binary coding and  $2^m$ -ary modulation are represented in a unified form as a sequential binary channel partition of the vector channel  $W^N$ . Both should be designed according to the polarization principle, i.e. the maximization of the variance of the bit channel capacities (37) under successive cancellation – or, equivalently, multi-stage – decoding by careful choice of the labeling rule.

##### C. Influence of the Labeling Rule

Here, we focus on the set-partitioning labeling approach (corresponding to  $\lambda_{SP}$ ) by Ungerboeck [13] and Gray labeling  $\lambda_G$  that aims to generate bit levels that are as independent as possible [14].

Fig. 2 depicts the variance of the bit levels for ASK modulation using both SP and (binary-reflected) Gray labeling.

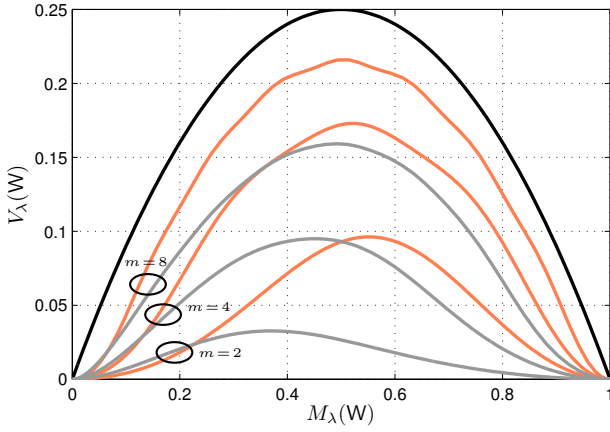


Fig. 2. Bit level variance for  $2^m$ -ary ASK signalling with multi-stage decoding over the AWGN channel ( $m = 2, 4, 8$ ). Red: SP labeling, gray: Gray labeling.

It can be observed that – except for small capacities  $M_\lambda(W)$  – the SP labeling approach leads to significantly larger bit level variances compared to the Gray labeling, as expected. Therefore, for multilevel polar codes, SP labeling should be preferably applied. Furthermore, when compared to the corresponding variance curves of polar codes over a single B-DMC for  $N = 2, 4, 8$  as shown in Fig. 1, especially in case of SP labeling the achieved bit level variance is significantly higher, emphasizing the importance of the careful choice of the labeling  $\mathcal{L}_\lambda$  in this first step of polarization for multilevel polar codes.

## V. SIMULATION RESULTS

We finally present some numerical results in terms of rate-vs.-power-efficiency plots in order to illustrate the error performance of polar-coded modulation with SC decoding over the AWGN channel.

Besides common Monte-Carlo simulations, we also present approximate results obtained by density evolution (DE) [10]. Here, for multilevel polar codes, we numerically determine the bit channel capacities  $I(B_\lambda^{(i)})$  of the transform  $\lambda$  as defined in (33). Then, we calculate the bit channel capacities – and the corresponding error probabilities  $p_e(B_{\lambda \circ \pi^n}^{(i)})$  – of the component polar codes by performing density evolution with the well-known Gaussian approximation [15], i.e., we simply assume the output bit channels of each SBP in the chain  $\lambda \otimes \pi \otimes \dots \otimes \pi$  to be Gaussian. The word error rate under SC decoding  $\text{WER}_{\text{SC}}$  is obtained from (28).

Fig. 3 depicts the performance of multilevel polar codes with 16-ASK modulation under SC decoding for different labelings  $\mathcal{L}_\lambda$  and various block lengths. The large performance loss of Gray labeling w.r.t. SP labeling can clearly be observed. Furthermore, by comparison of the results obtained by DE and the simulation points, the inaccuracy induced by the Gaussian assumption can be obviously neglected.

## VI. CONCLUSIONS

In this paper, we have combined polar coding and multilevel coding by representing both as sequential binary partitions

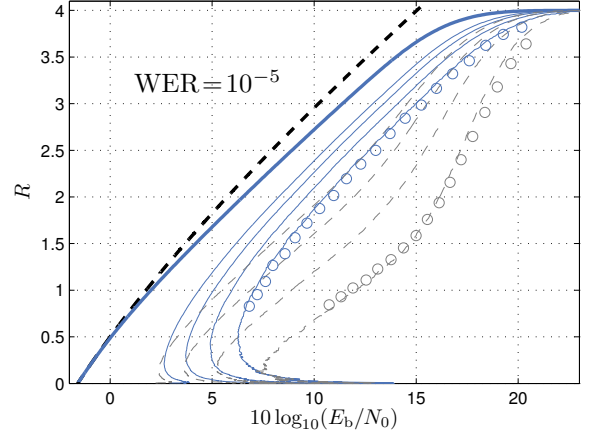


Fig. 3. 16-ASK / AWGN: Rate vs. power efficiency of multilevel polar codes using SP (blue) and Gray labeling (dashed gray). Markers: Simulation points for  $mN = 512$ , lines: DE results with overall block length (from right to left)  $mN = 2^k$ ,  $k = 9, 11, 13, 15$ . Bold blue: coded-modulation capacity, dashed black: Shannon bound for real constellations.

(SBPs). Based on this representation, we have derived rules for optimization of multilevel polar codes.

Future work will extend this novel framework of channel partitions to bit-interleaved coded modulation (BICM) and incorporate fading scenarios.

## REFERENCES

- [1] E. Arkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3051–3073, Jul. 2009.
- [2] E. Sasoglu, E. Telatar, and E. Arkan, "Polarization for arbitrary discrete memoryless channels," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Oct. 2009, pp. 144–148.
- [3] D.-M. Shin, S.-C. Lim, and K. Yang, "Mapping selection and code construction for  $2^m$ -ary polar-coded modulation," *IEEE Commun. Lett.*, vol. 16, pp. 905–908, Jun. 2012.
- [4] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, pp. 927–946, May 1998.
- [5] H. Imai and S. Hiraoka, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 23, pp. 371–377, 1977.
- [6] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1361–1391, Jul. 1999.
- [7] E. Arkan, "Polar coding: Status and prospects," Plenary Talk at *IEEE Int. Symp. Inf. Theory (ISIT)*, Aug. 2011, available: <http://polaran.com/images/arikan2011isit.pdf>.
- [8] S. Korada, E. Sasoglu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *IEEE Trans. Inf. Theory*, vol. 56, pp. 6253–6264, Dec. 2010.
- [9] I. Tal and A. Vardy, "List decoding of polar codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aug. 2011, pp. 1–5.
- [10] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Commun. Lett.*, vol. 13, pp. 519–521, Jul. 2009.
- [11] I. Land and J. B. Huber, "Information combining," *Found. Trends Commun. Inf. Theory*, vol. 3, no. 3, pp. 227–330, Nov. 2006.
- [12] M. Seidl, A. Schenk, C. Stierstorfer, and J. B. Huber, "Aspects of polar-coded modulation," in *Proc. 9th Int. ITG Conf. Systems, Commun. and Coding (SCC)*, Munich, Germany, Jan. 2013.
- [13] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets Part I: Introduction," *IEEE Commun. Mag.*, vol. 25, no. 2, pp. 5–11, Feb. 1987.
- [14] C. Stierstorfer and R. F. H. Fischer, "(Gray) mappings for bit-interleaved coded modulation," in *Proc. IEEE Veh. Technol. Conf. Spring (VTC Spring)*, Dublin, Ireland, Apr. 2007, pp. 1703–1707.
- [15] S.-Y. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 47, pp. 657–670, Feb. 2001.