Multilevel Binary Polar-Coded Modulation Achieving the Capacity of Asymmetric Channels

Constantin Runge ⬤, Thomas Wiegart ⬤, Diego Lentner ⬤, Tobias Prinz ⬤

Institute for Communications Engineering, Technical University of Munich, 80333 Munich, Germany

{constantin.runge, thomas.wiegart, diego.lentner, tobias.prinz}@tum.de

Abstract—A multilevel coded modulation scheme is studied that uses solely binary polar codes and Honda-Yamamoto probabilistic shaping. The scheme is shown to achieve the capacity of discrete memoryless channels with input alphabets of cardinality a power of two. The performance of finite-length implementations is compared to polar-coded probabilistic amplitude shaping and constant composition distribution matching.

Index Terms—coded modulation, polar codes, asymmetric channels, probabilistic shaping

I. INTRODUCTION

Reliable and power-efficient communication usually requires probabilistic shaping (PS) and/or geometric shaping .There are several ways to implement PS, e.g., many-to-one mappings [1, Sec. 6.2], trellis shaping [2], and others, see [3, Sec. II], [4]. More recent schemes are probabilistic amplitude shaping (PAS) [3], and Honda-Yamamoto (HY) PS [5] based on polar codes [6], [7].

PAS received significant attention from the optical fiber communications community and industry due to its performance and flexibility [8], [9]. PAS requires a target distribution P_X that factors as $P_X = P_A \cdot P_S$ so that P_S is a uniform binary distribution. Usually "A" and "S" refer to the amplitude and sign of X, respectively, but more general choices are permitted. We focus on $P_S(-1) = P_S(1) = 1/2$. An important component of PAS is a *distribution matching (DM)* device that maps uniformly distributed bits to real-alphabet symbols with distribution P_A, e.g., a constant composition distribution matching (CCDM) device [10]. These symbols are then protected with the parity bits of a systematic forward error control (FEC) code. Each parity bit is uniformly distributed and chooses one of two signs S so that $X = A \cdot S$ has $P_X(x) = P_X(-x)$. PAS in general does not allow for asymmetric P_X.

The HY scheme generates asymmetric P_X by performing joint DM and FEC. The scheme achieves the capacity of general binary-input discrete memoryless channels (biDMCs) [5] and has excellent performance for short block lengths. For instance, see [11] that compares the performance of different schemes for on-off keying (OOK) modulation over additive white Gaussian noise (AWGN) channels. An earlier scheme by Sutter et al. [12] also achieves the capacity of biDMCs. This scheme concatenates two separate polar codes for FEC and DM which reduces the error exponent by a factor of two as compared to HY coding [5].

Polar codes can be extended to higher-order modulation by using multilevel coding (MLC) [13]. In this paper, we study a multilevel Honda-Yamamoto (MLHY) coding scheme which is amenable to practical implementation. Our contributions are two-fold. First, we prove that MLHY coding achieves the capacity of general discrete memoryless channels (DMCs) with $M = 2^m$-ary channel inputs. Second, we compare the DM performance and shaping gains of PAS and MLHY coding for short block lengths. We evaluate the performance with unipolar ($X \geq 0$) and bipolar ($X \in \mathbb{R}$) modulation over AWGN channels. The proposed scheme performs on-par with polar-coded PAS [14] and does not need a DM device.

We remark that several polar coding architectures, including multilevel ones, were studied in [15]–[21] but these papers do not consider capacity proofs. Using polar lattice codes, a capacity proof for channels whose input alphabets have a lattice structure is given in [22]. Our proof and the proof in [22] are both based on the idea that each bitlevel polarizes, but we note that MLHY coding is not restricted to lattice inputs. This makes our proof simpler and more general. Note that the scheme of [22] is effectively a special case of the MLHY coding studied here for the case of amplitude-shift keying (ASK) modulation, Gaussian P_X and a set-partitioning bit-mapping [23].

Instead of using MLHY coding, the capacity of DMCs can also be achieved by combining HY coding with non-binary kernels [5], [24], [25]. However, binary polar codes are preferred in practice because non-binary polar codes and decoders are complex to implement and design [26], [27].

Polar-coded modulation has also been studied in the context of multiple-access channels [28, Sec. V] where each bitlevel of a channel input symbol corresponds to one user. Based on this approach, the authors of [28] describe a MLC scheme that achieves the symmetric capacity of DMCs with $M = 2^m$-ary channel inputs using independent binary polar codes for each bitlevel. For transmission over the AWGN channel, they combine the scheme with a many-to-one mapping, which is not desirable in practice.

This paper is organized as follows. Section II gives an overview of polar coding concepts. Section III shows that the MLHY scheme achieves capacity and develops error exponents. Finally, Section IV treats polar codes for short block lengths and provides numerical results.

II. PRELIMINARIES

A. Notation

Random variables are written with upper case letters such as X. Their alphabet, distribution, and realizations are written as $\mathcal{X}$, P_X, and x, respectively. Vectors are denoted by bold symbols such as $\boldsymbol{x}$. $\mathcal{X}^C$ and $|\mathcal{X}|$ are the complement and cardinality of $\mathcal{X}$, respectively. A set difference is denoted as $\mathcal{X} \setminus \mathcal{Y} = \mathcal{X} \cap \mathcal{Y}^C$. An index set from 1 to N is denoted as $[\![N]\!] \triangleq \{1, \ldots, N\}$. A set $\mathcal{S}$ may select entries of a vector, creating a substring $\boldsymbol{x}_\mathcal{S}$ with length $|\mathcal{S}|$, e.g., $\boldsymbol{x}_{[\![N]\!]}$. An event $\mathcal{E}$ has probability $\mathcal{P}(\mathcal{E})$.

We denote by $\mathbb{H}(X)$, $\mathbb{H}(X|Y)$, and $\mathbb{I}(X;Y)$ the entropy of X, the entropy of X conditioned on Y and the mutual information (MI) of X and Y, respectively. The conditional Bhattacharyya parameter [29] is defined as $Z(X|Y) = 2\,\mathbb{E}\big[\sqrt{P_{X|Y}(0|Y)P_{X|Y}(1|Y)}\big]$ and satisfies [29] [22, Lem. 6]

$$Z(X|Y)^2 \leq \mathbb{H}(X|Y) \leq Z(X|Y) \tag{1}$$
$$Z(X|Y,S) \leq Z(X|Y) \tag{2}$$

where $X, Y, S \sim P_{X,Y,S}$.

B. Polarization and Polar Coding

Polar codes [6], [7] are linear block codes of length $N = 2^n$ for $n \in \mathbb{N}$. They are defined via the polar transform that maps a vector $\boldsymbol{u} \in \mathbb{F}_2^N$ to a codeword $\boldsymbol{x} \in \mathbb{F}_2^N$

$$\boldsymbol{x} = \boldsymbol{u}\boldsymbol{G}_N \text{ with } \boldsymbol{G}_N = \boldsymbol{B}_N \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes n} \tag{3}$$

where $\boldsymbol{B}_N$ is the bit-reversal matrix as in [7], and where $\boldsymbol{F}^{\otimes n}$ is the n-fold Kronecker product of $\boldsymbol{F}$. The polar transform satisfies $\boldsymbol{G}_N^{-1} = \boldsymbol{G}_N$. For encoding, we will consider the codeword $\boldsymbol{x}$ to have independent and identically distributed (i.i.d.) entries. The codeword $\boldsymbol{x}$ is transmitted over N uses of a biDMC $W: X \to Y$ resulting in a vector of channel observations $\boldsymbol{y} \in \mathcal{Y}^N$. Consider the sets

$$\mathcal{L}_{U|Y} = \{i \in [\![N]\!] : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y}) < \delta_N\}, \tag{4}$$
$$\mathcal{H}_{U|Y} = \{i \in [\![N]\!] : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y}) > 1 - \delta_N\} \tag{5}$$

with $\delta_N \triangleq 2^{-N^\beta}$ for any $\beta < \frac{1}{2}$. It is known [5, Eqs. (38), (39)] that these index sets polarize, i.e., we have

$$\lim_{N\to\infty} \frac{1}{N}|\mathcal{L}_{U|Y}| = 1 - \mathbb{H}(X|Y) \tag{6}$$
$$\lim_{N\to\infty} \frac{1}{N}|\mathcal{H}_{U|Y}| = \mathbb{H}(X|Y). \tag{7}$$

The encoder places the data bits on the reliable bit positions of $\boldsymbol{u}$, i.e., those with $Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y}) \approx 0$. The remaining positions in $\boldsymbol{u}$ are frozen, i.e., set to fixed values. The receiver uses successive cancellation (SC) decoding of the non-frozen bits via $\hat{u}_i = \arg\max_u P_{U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y}}(u|\hat{\boldsymbol{u}}_{[\![i-1]\!]}, \boldsymbol{y})$.

Honda and Yamamoto [5] consider two more sets:

$$\mathcal{L}_U = \{i \in [\![N]\!] : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}) < \delta_N\} \tag{8}$$
$$\mathcal{H}_U = \{i \in [\![N]\!] : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}) > 1 - \delta_N\}. \tag{9}$$

Note that (7) and (9) yield $|\mathcal{H}_U|/N \overset{N\to\infty}{\to} \mathbb{H}(X)$. With (2), we have $\mathcal{L}_U \subseteq \mathcal{L}_{U|Y}$ and thus $\mathcal{L}_U \cap \mathcal{H}_{U|Y} = \emptyset$. For the "data" set $\mathcal{I} = \mathcal{H}_U \cap \mathcal{L}_{U|Y}$, we thus have [5, Thm. 1]

$$\lim_{N\to\infty} \frac{1}{N}|\mathcal{I}| = \mathbb{I}(X;Y). \tag{10}$$

To achieve capacity, Honda and Yamamoto chose the $\mathcal{I}$ bits as data bits and the remaining bits in $\boldsymbol{u}$ randomly with probability $P_{U_i|\boldsymbol{U}_{[\![i-1]\!]}}(\cdot|\boldsymbol{u}_{[\![i-1]\!]})$. To calculate these probabilities, the same SC structure as for decoding is employed. The random bits must be known to the receiver. We describe the method in more detail in Section III-A.

C. Conditional Polarization

We next consider conditional polarization which helps to prove our main results.

Lemma 1. Let $X \in \mathbb{F}_2$ and S be the input to a biDMC $W: X \to Y$ with side information S with joint distribution $(X,Y,S) \sim P_{Y|X,S}P_{X|S}P_S$. Let $(\boldsymbol{X},\boldsymbol{S},\boldsymbol{Y})$ be N i.i.d. realizations of (X,S,Y) and let $\boldsymbol{U} = \boldsymbol{X}\boldsymbol{G}_N$. Then all the index sets

$$\mathcal{L}_{U|S} = \{i \in [\![N]\!] : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S}) < \delta_N\} \tag{11}$$
$$\mathcal{H}_{U|S} = \{i \in [\![N]\!] : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S}) > 1 - \delta_N\} \tag{12}$$
$$\mathcal{L}_{U|S,Y} = \{i \in [\![N]\!] : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S}, \boldsymbol{Y}) < \delta_N\} \tag{13}$$
$$\mathcal{H}_{U|S,Y} = \{i \in [\![N]\!] : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S}, \boldsymbol{Y}) > 1 - \delta_N\} \tag{14}$$

polarize with $\delta_N \triangleq 2^{-N^\beta}$ for any $\beta < \frac{1}{2}$, yielding

$$\lim_{N\to\infty} \frac{1}{N}|\mathcal{L}_{U|S} \cup \mathcal{H}_{U|S,Y}| = 1 - \mathbb{I}(X;Y|S) \tag{15}$$
$$\lim_{N\to\infty} \frac{1}{N}|\mathcal{H}_{U|S} \cap \mathcal{L}_{U|S,Y}| = \mathbb{I}(X;Y|S). \tag{16}$$

Proof. Using [5, Eqs. (38), (39)], the sets (11) to (14) polarize analogously to (6), (7), i.e.,

$$1 - \lim_{N\to\infty}\frac{|\mathcal{L}_{U|S}|}{N} = \lim_{N\to\infty}\frac{|\mathcal{H}_{U|S}|}{N} = \mathbb{H}(X|S) \tag{17}$$
$$1 - \lim_{N\to\infty}\frac{|\mathcal{L}_{U|S,Y}|}{N} = \lim_{N\to\infty}\frac{|\mathcal{H}_{U|S,Y}|}{N} = \mathbb{H}(X|S,Y). \tag{18}$$

We next show (15). Basic set theory gives

$$|\mathcal{L}_{U|S} \cup \mathcal{H}_{U|S,Y}| = |\mathcal{L}_{U|S}| + |\mathcal{H}_{U|S,Y}| - |\mathcal{L}_{U|S} \cap \mathcal{H}_{U|S,Y}|. \tag{19}$$

By (2), we have $\mathcal{L}_{U|S} \subseteq \mathcal{L}_{U|S,Y}$ and thus $\mathcal{L}_{U|S} \cap \mathcal{H}_{U|S,Y} = \emptyset$. Inserting the first term of (17) and the second term of (18) into (19) gives (15). To prove (16), observe that

$$\lim_{N\to\infty}\frac{|\mathcal{H}_{U|S} \cap \mathcal{L}_{U|S,Y}|}{N} = 1 - \lim_{N\to\infty}\frac{|(\mathcal{H}_{U|S} \cap \mathcal{L}_{U|S,Y})^C|}{N} \tag{20}$$
$$= 1 - \lim_{N\to\infty}\frac{|\mathcal{H}_{U|S}^C \cup \mathcal{L}_{U|S,Y}^C|}{N} \tag{21}$$
$$= 1 - \lim_{N\to\infty}\frac{|\mathcal{L}_{U|S} \cup \mathcal{H}_{U|S,Y}|}{N} \tag{22}$$

where the last equality follows from (17) and (18). Combining (15) and (22) yields (16). $\qquad\square$
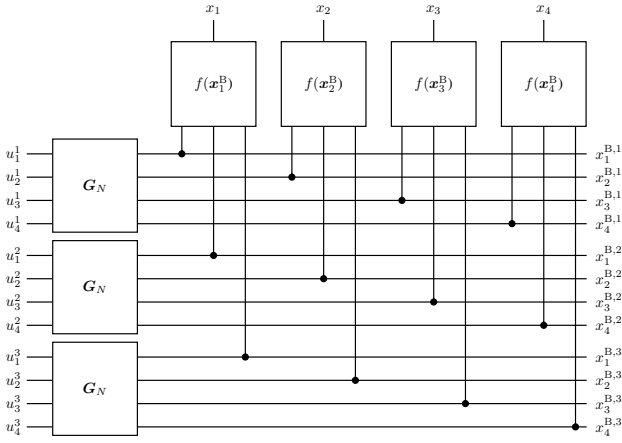
Fig. 1. Multilevel polar-coded modulation with $M = 2^m = 8$ and $N = 4$.

III. MULTILEVEL POLAR CODES FOR DMCs

Consider the multilevel code construction in [13] depicted in Fig. 1. For a channel with input alphabet $\mathcal{X}$ of cardinality $M = |\mathcal{X}| = 2^m$, each symbol x is labelled with m bits, i.e., $x = f(x^{\mathrm{B},1} x^{\mathrm{B},2} \cdots x^{\mathrm{B},m})$ where $f(\cdot)$ is invertible. A codeword has a length of N symbols or mN bits. Each bitlevel $\boldsymbol{x}^{\mathrm{B},\ell}$, $\ell \in [\![m]\!]$, is encoded with a separate polar transform as $\boldsymbol{u}^\ell = \boldsymbol{x}^{\mathrm{B},\ell} \boldsymbol{G}_N$. Using Lem. 1, we can prove the polarization of such a multilevel polar code.

Theorem 1. *Let* $W \colon X \to Y$ *be a DMC with joint distribution* $X, Y \sim P_{Y|X} P_X$ *and* $|\mathcal{X}| = 2^m$. *Let* $X^{\mathrm{B},\ell}$, $\ell \in [\![m]\!]$, *be the ℓ-th bit of the binary representation* $\boldsymbol{X}^{\mathrm{B}}$ *of symbol* $X = f(\boldsymbol{X}^{\mathrm{B}})$, *and let* $\boldsymbol{U}^\ell = \boldsymbol{X}^{\mathrm{B},\ell} \boldsymbol{G}_N$. *Then, the sets*

$$\mathcal{L}_U' = \{(\ell, i) : Z(U_i^\ell | \boldsymbol{V}_i^\ell) < \delta_N\} \tag{23}$$

$$\mathcal{H}_U' = \{(\ell, i) : Z(U_i^\ell | \boldsymbol{V}_i^\ell) > 1 - \delta_N\} \tag{24}$$

$$\mathcal{L}_{U|Y}' = \{(\ell, i) : Z(U_i^\ell | \boldsymbol{V}_i^\ell, \boldsymbol{Y}) < \delta_N\} \tag{25}$$

$$\mathcal{H}_{U|Y}' = \{(\ell, i) : Z(U_i^\ell | \boldsymbol{V}_i^\ell, \boldsymbol{Y}) > 1 - \delta_N\} \tag{26}$$

with $\ell \in [\![m]\!]$, $i \in [\![N]\!]$, *and* $\boldsymbol{V}_i^\ell = (\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{X}^{\mathrm{B},[\![\ell-1]\!]})$ *polarize, i.e., we have*

$$\lim_{N \to \infty} \frac{1}{N} |\mathcal{L}_U' \cup \mathcal{H}_{U|Y}'| = 1 - \mathbb{I}(X; Y) \tag{27}$$

$$\lim_{N \to \infty} \frac{1}{N} |\mathcal{I}| = \mathbb{I}(X; Y) \tag{28}$$

where $\mathcal{I} = \mathcal{H}_U' \cap \mathcal{L}_{U|Y}'$.

Proof. We begin by showing (28). For each bitlevel $\ell \in [\![m]\!]$, consider the sets

$$\mathcal{H}_{U^\ell}' = \{(\ell, i) : Z(U_i^\ell | \boldsymbol{V}_i^\ell) > 1 - \delta_N\} \tag{29}$$

$$\mathcal{L}_{U^\ell|Y}' = \{(\ell, i) : Z(U_i^\ell | \boldsymbol{V}_i^\ell, \boldsymbol{Y}) < \delta_N\}. \tag{30}$$

By Lem. 1, we have

$$\lim_{N \to \infty} \frac{1}{N} |\mathcal{H}_{U^\ell}' \cap \mathcal{L}_{U^\ell|Y}'| = \mathbb{I}(X^{\mathrm{B},\ell}; Y | \boldsymbol{X}^{\mathrm{B},[\![\ell-1]\!]}). \tag{31}$$

Note that $\mathcal{H}_{U^\ell}' \cap \mathcal{H}_{U^k}' = \mathcal{L}_{U^\ell|Y}' \cap \mathcal{L}_{U^k|Y}' = \mathcal{H}_{U^\ell}' \cap \mathcal{L}_{U^k|Y}' = \emptyset$ by definition for $\ell \neq k$ and

$$\mathcal{H}_U' = \bigcup_{\ell \in [\![m]\!]} \mathcal{H}_{U^\ell}', \quad \mathcal{L}_{U|Y}' = \bigcup_{\ell \in [\![m]\!]} \mathcal{L}_{U^\ell|Y}'. \tag{32}$$

Additionally, since $f(\cdot)$ is invertible we have

$$\lim_{N \to \infty} \frac{1}{N} |\mathcal{I}| = \lim_{N \to \infty} \frac{1}{N} \sum_{\ell \in [\![m]\!]} |\mathcal{H}_{U^\ell}' \cap \mathcal{L}_{U^\ell|Y}'| \tag{33}$$

$$= \sum_{\ell \in [\![m]\!]} \mathbb{I}(X^{\mathrm{B},\ell}; Y | \boldsymbol{X}^{\mathrm{B},[\![\ell-1]\!]}) \tag{34}$$

$$= \mathbb{I}(X; Y) \tag{35}$$

which proves (28). The proof of (27) follows analogously to the proof of (16). $\square$

A. Encoding and Decoding

The encoding is similar to [5]. Define the message set $\mathcal{M} \subseteq \mathcal{I}$ as the set of bit positions (ℓ, i) populated by data bits. The remaining bits u_i^ℓ, $(\ell, i) \in \mathcal{M}^C$, are chosen successively and randomly with probability $P_{U_i^\ell | \boldsymbol{V}_i^\ell}(\cdot | \boldsymbol{v}_i^\ell)$, where $\boldsymbol{v}_i^\ell$ again includes the bits decided before (ℓ, i). To compute $P_{U_i^\ell | \boldsymbol{V}_i^\ell}(\cdot | \boldsymbol{v}_i^\ell)$, we factor $P_X(x_i)$ as

$$P_X(x_i) = \prod_{\ell \in [\![m]\!]} P_{X^{\mathrm{B},\ell} | \boldsymbol{X}^{\mathrm{B},[\![\ell-1]\!]}}(x_i^{\mathrm{B},\ell} | \boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]}). \tag{36}$$

For each bitlevel ℓ, multistage decoding (MSD) computes $P_{X^{\mathrm{B},\ell} | \boldsymbol{X}^{\mathrm{B},[\![\ell-1]\!]}}(\cdot | \boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]})$, $i \in [\![N]\!]$ [30], and provides these values to a SC decoder that computes $P_{U_i^\ell | \boldsymbol{V}_i^\ell}(\cdot | \boldsymbol{v}_i^\ell)$ and decides on u_i^ℓ.

The decoder uses the same MSD structure with SC decoding. The bits $\hat{u}_i^\ell$, $(\ell, i) \in \mathcal{M}$, are estimated as $\hat{u}_i^\ell = \arg\max_u P_{U_i^\ell | \boldsymbol{V}_i^\ell, \boldsymbol{Y}}(u | \boldsymbol{v}_i^\ell, \boldsymbol{y})$ assuming perfect knowledge of the previous bits $\boldsymbol{v}_i^\ell$. The non-message bits are decided from $P_{U_i^\ell | \boldsymbol{V}_i^\ell}(\cdot | \boldsymbol{v}_i^\ell)$ requiring randomness that is shared by the transmitter and receiver. The decoding error probability $\mathcal{P}(\{\hat{\boldsymbol{U}} \neq \boldsymbol{U}\})$ is averaged over this randomness.

Theorem 2. *Let* $W \colon X \to Y$ *and define* $\mathcal{I}$ *as in Thm. 1. Let* $\mathcal{M} \subseteq \mathcal{I}$, *and consider encoding and decoding as described above. Then the average decoding error probability is* $\mathcal{P}(\{\hat{\boldsymbol{U}} \neq \boldsymbol{U}\}) = \mathcal{O}(2^{-N^{\beta'}})$ *for any* $0 < \beta' < \beta < 1/2$ *by choosing the polarization sets with* $\delta_N = 2^{-N^\beta}$.

Proof. Consider

$$\mathcal{P}(\{\hat{\boldsymbol{U}} \neq \boldsymbol{U}\}) = 1 - \prod_{\ell \in [\![m]\!]} \left(1 - \mathcal{P}\left(\mathcal{E}^\ell \big| \mathcal{C}^{[\![\ell-1]\!]}\right)\right) \tag{37}$$

where $\mathcal{E}^\ell = \{\hat{\boldsymbol{U}}^\ell \neq \boldsymbol{U}^\ell\}$ and $\mathcal{C}^{[\![\ell-1]\!]} = \{\hat{\boldsymbol{U}}^{[\![\ell-1]\!]} = \boldsymbol{U}^{[\![\ell-1]\!]}\}$. Let the equivalent channel for the ℓ-th bitlevel be the channel that has bit ℓ as input and bits $[\![\ell-1]\!]$ as side-information available at transmitter and receiver. By [31, Thm. 4.3.9], [5, Thm. 3], the HY code over this equivalent channel for bitlevel ℓ has an average decoding error probability $\mathcal{P}(\mathcal{E}^\ell | \mathcal{C}^{[\![\ell-1]\!]}) = \mathcal{O}(2^{-N^{\beta'}})$ with $\beta' < \beta < 1/2$ and uniformly chosen messages.

Thus, for each bitlevel ℓ there is a positive constant c_ℓ and a block length N_ℓ so that $\mathcal{P}\big(\mathcal{E}^\ell\big|\mathcal{C}^{[\![\ell-1]\!]}\big) \leq c_\ell 2^{-N^{\beta'}}$ for all $N > N_\ell$. By choosing $c = \max_{\ell \in [\![m]\!]} c_\ell$ and $N^* = \max_{\ell \in [\![m]\!]} N_\ell$, we can bound the error probability for any $\ell \in [\![m]\!]$ by

$$\mathcal{P}\Big(\mathcal{E}^\ell\Big|\mathcal{C}^{[\![\ell-1]\!]}\Big) \leq c 2^{-N^{\beta'}} \tag{38}$$

for all $N > N^*$. The average decoding error probability under MSD can thus be bounded as

$$\mathcal{P}(\{\hat{\boldsymbol{U}} \neq \boldsymbol{U}\}) = 1 - \prod_{\ell \in [\![m]\!]} \left(1 - \mathcal{P}\Big(\mathcal{E}^\ell\Big|\mathcal{C}^{[\![\ell-1]\!]}\Big)\right) \tag{39}$$

$$\leq 1 - \left(1 - c\,2^{-N^{\beta'}}\right)^m \tag{40}$$

$$\leq mc\,2^{-N^{\beta'}} \tag{41}$$

where the final step follows by Bernoulli's inequality. □

We remark that one can extend Thms. 1 and 2 to discrete-input, continuous-output channels along the lines of [32, Part IV, Appendix 7].

IV. SHORT BLOCKLENGTH CODES

A pragmatic approach is to choose the non-data bits with a deterministic rule [33], [34] where the bits with large $\mathbb{H}(U_i^\ell|\boldsymbol{V}_i^\ell,\boldsymbol{Y})$ are fixed to 0 and the bits with small $\mathbb{H}(U_i^\ell|\boldsymbol{V}_i^\ell)$ are decided as $\arg\max_u P_{U_i|\boldsymbol{V}_i^\ell}(u|\hat{\boldsymbol{v}}_i^\ell)$ ("DM bits"). The remaining bits are data bits. The decoder estimates the non-frozen bits via $\arg\max_u P_{U_i|\boldsymbol{V}_i^\ell,\boldsymbol{Y}}(u|\hat{\boldsymbol{v}}_i^\ell,\boldsymbol{y})$. The DM bits with $\mathbb{H}(U_i^\ell|\boldsymbol{V}_i^\ell) \approx 0$ also have $\mathbb{H}(U_i^\ell|\boldsymbol{V}_i^\ell,\boldsymbol{Y}) \approx 0$ and are thus reliably estimated. We call the resulting scheme MLHY coding. The entropies used for code construction can be computed with, e.g., Monte Carlo (MC) integration or density evolution with Gaussian approximation [15]. Similar to [13], we jointly compute the bitchannel entropies over all bitlevels.

A. Distribution Matching

Consider first a code that performs only DM, i.e., there are no frozen bits. To evaluate the performance, we consider the rate loss [3, Sec. V-B], [11], [20]

$$\Delta_{\mathrm{R}} = \mathbb{H}(\widehat{P}_X) - R = \mathbb{H}(\widehat{P}_X) - \frac{|\mathcal{U}|}{N} \tag{42}$$

where $\widehat{P}_X$ is the empirical distribution of X, and $\mathcal{U}$ indexes the bitchannels with uniformly-distributed data bits. Typically, $\mathcal{U}$ consists of the bitchannels with $\mathbb{H}(U_i^\ell|\boldsymbol{V}_i^\ell) \approx 1$.

Fig. 2 shows the rate loss for CCDM [10] and for MLHY DM with successive cancellation list (SCL) encoding [35] with list size $L = 32$ instead of randomized encoding. The target distributions are the maximum entropy distributions for the rates in bits per channel use (bpcu):

- $R = 1.625$ bpcu for $M = 4$;
- $R = 2.375$ bpcu for $M = 8$;
- $R = 3.250$ bpcu for $M = 16$.

The MLHY code is constructed by using the RN bitchannels with the largest $\mathbb{H}(U_i^\ell|\boldsymbol{V}_i^\ell)$ for data. The quantized distribution and the rate for CCDM are determined by [36, Algorithm 2] and [3, Eq. (37)], respectively.
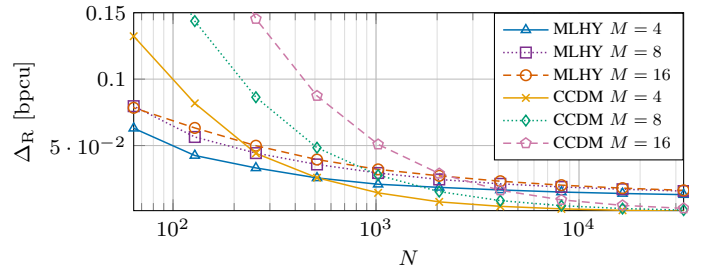

Fig. 2. Rate loss for MLHY with $L = 32$ and CCDM with different M.

Observe that CCDM is better than MLHY DM for large block lengths [37], [11]. This is expected since the polar code has a rigid structure. However, MLHY DM has a smaller rate loss than CCDM for practically-relevant block lengths up to $N = 1024$. We observe that the rate loss of CCDM increases with M for short and moderate block lengths whereas the rate loss of MLHY degrades only slightly. MLHY DM thus has superior performance for short block lengths and offers the flexibility to design joint DM and FEC schemes.

B. End-to-end Frame Error Rates

We compare MLHY coding with the polar-coded probabilistic amplitude shaping (PC-PAS) scheme proposed in [14]. PC-PAS uses the PAS architecture [3] with a systematic multilevel polar code as FEC and CCDM [10] for DM.

Consider bipolar ASK and unipolar pulse-amplitude modulation (PAM). The input alphabets of cardinality $M = 8$ and $M = 4$, respectively, are

$$\mathcal{X}_{\mathrm{ASK}} = \{\pm 7, \pm 5, \pm 3, \pm 1\}, \quad \mathcal{X}_{\mathrm{PAM}} = \{0, 1, 2, 3\}.$$

For both cases, we choose $P_X(x) \propto \exp(-\nu|x|^2)$ so that ν minimizes the frame error rate (FER).

The transmitter and receiver use SCL decoding with list size $L = 32$ and an optional outer cyclic redundancy check (CRC) code. The code is designed for a specific rate R and block length N. There are three relevant design parameters. The first is the design signal-to-noise ratio (dSNR) that determines the noise variance for computing $\mathbb{H}(U_i^\ell|\boldsymbol{V}_i^\ell,\boldsymbol{Y})$. Second, we introduce a design parameter κ for code optimization and choose the rate-optimal P_X at $\kappa \cdot$ dSNR as our target distribution based on which we also compute $\mathbb{H}(U_i^\ell|\boldsymbol{V}_i^\ell)$. This parameter can improve the finite length performance because the optimal channel input distribution might deviate from the asymptotically optimal one. Finally, we optimize over the number N_{DM} of DM positions. The code is constructed by choosing the N_{DM} positions with lowest $\mathbb{H}(U_i^\ell|\boldsymbol{V}_i^\ell)$ for DM and the $N(1-R) - N_{\mathrm{DM}}$ positions with highest $\mathbb{H}(U_i^\ell|\boldsymbol{V}_i^\ell,\boldsymbol{Y})$ for FEC. The remaining positions are used for data. We use set-partitioning labelling [38] for the channel input symbols.

The scheme from [14] must be modified to transmit PAM symbols with polar-coded PAS. First, PAS requires symmetric distributions, as described in the introduction. The one-sided sampled Gaussian distribution that we use for PAM does
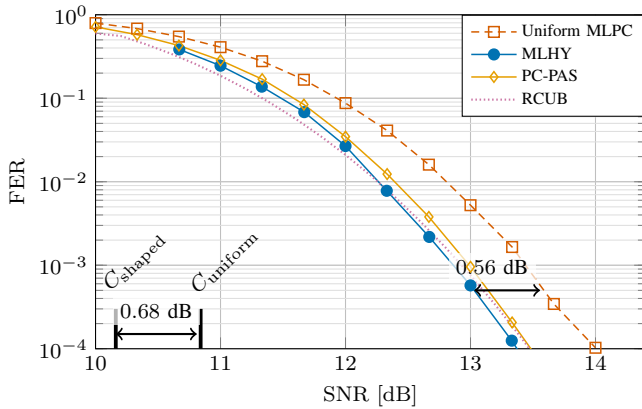
Fig. 3. Performance of MLHY coding compared to PC-PAS [14], uniform MLPC, and the RCUB, with an 8-ASK constellation, $N = 64$, $L = 32$, at a $R = 1.75$ bpcu, shaped dSNR $= 13$ dB, $\kappa = -1$ dB, $N_{\mathrm{DM}} = 23$, and uniform dSNR $= 13$ dB. The MLHY and MLPC codes use an outer CRC-7 and PC-PAS a CRC-4 together with a type check. For PC-PAS, dSNR $= 8$ dB and $\kappa = -0.6$ dB.
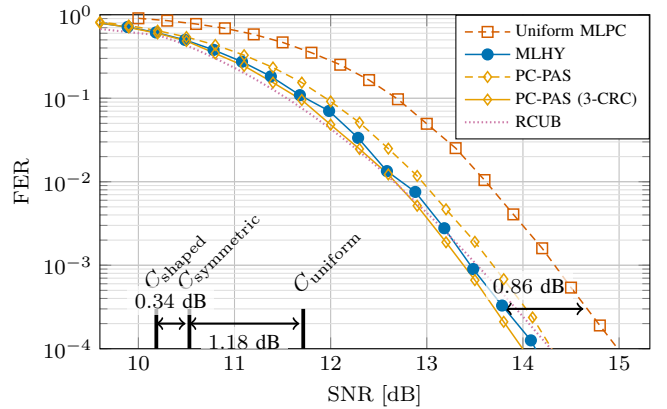

Fig. 4. Performance of MLHY coding compared to PC-PAS, uniform MLPC, and the RCUB, with an 4-PAM constellation, $N = 64$, $L = 32$, at a $R = 1.25$ bpcu, shaped dSNR $= 18.1$ dB, $\kappa = -0.9$ dB, $N_{\mathrm{DM}} = 24$, and uniform dSNR $= 19.25$ dB. The MLHY and MLPC codes do not use an outer CRC. For PC-PAS we depict curves with and without outer CRC. Further, for PC-PAS, dSNR $= 14.5$ dB and $\kappa = -3.9$ dB.

not fulfill this requirement. Instead, we approximate a one-sided sampled Gaussian distribution $\tilde{P}_X$ for M-PAM by assigning $M/2$ different probability masses to pairs of points as described in [39], [40], i.e., $\tilde{P}_X(2i) = \tilde{P}_X(2i + 1)$ $\forall i = 0, 1, \ldots, M/2 - 1$. The input distribution is thus suboptimal. Second, polar-coded PAS uses a set-partitioning labelling. For ASK modulation as in [14], the last bitlevel carries the sign of the constellation. For PAM modulation, the first bitlevel refers to the bit that maps the transmitted signal to either the one or the other point of a pair. This facilitates systematic encoding and we can omit the labelling transformation described in [14].

Fig. 3 shows the FER for an 8-ASK constellation and $N = 64$. We also show the random coding union bound (RCUB) [41] computed for the distribution realized by the MLHY encoder, and the FER for uniform multilevel polar coding (MLPC) [13]. The codes and bounds are designed for $R = 1.75$ bpcu. The bold black lines at 10.16 dB and 10.84 dB show the constellation-constrained capacities for shaped and uniform transmission, respectively.

The error curve slopes for MLHY coding and uniform MLPC are similar, resulting in an almost constant shaping gain in the waterfall region. Both MLHY coding and PC-PAS perform close to the theoretical shaping gain of 0.68 dB and to the RCUB. The MLHY scheme thus performs on par with PC-PAS, even without a dedicated code optimization beyond a random search over the design paramters.

We describe potential improvements. Because CCDM code-words are all of the same type, PC-PAS permits an additional list pruning step [14] so that the length of the outer CRC code can be reduced. The performance of MLHY coding may be improved by further adjusting the design parameters, optimizing the bitchannel selection process, optimizing the CRC polynomial and length, and checking candidate codewords against DM constraints at the decoder.

Fig. 4 depicts the FERs for a 4-PAM constellation with

$N = 64$ and $R = 1.25$ bpcu. We show shaped MLHY coding and PC-PAS, uniform MLPC, and the RCUB for PAM over the AWGN channel. The additional shaping gain of using the rate-optimal, asymmetric P_X over a symmetric distribution $\tilde{P}_X$ is approximately 0.34 dB. Without CRC, the MLHY curve exhibits the predicted shaping gain and outperforms PC-PAS without CRC. It further lies on top of the RCUB. With list pruning by CRC and type checking, PC-PAS gains approximately 0.2 dB. First results using an additional outer CRC code in the MLHY scheme did not provide a noticeable coding gain. We therefore did not include the CRC curves in this case. We expect to recover the full shaping gain by further optimizing the polar and CRC codes.

Recall that MLHY coding uses the same binary polar multistage decoder at the transmitter and the receiver. The implementation complexity is thus reduced as compared to PAS. Furthermore, the use of CCDM as an outer code causes the end-to-end bit error rate (BER) of PAS to typically be much higher than for MLHY coding for the same FER.

V. Conclusion

We showed that multilevel polar coded modulation with binary polar codes and Honda-Yamamoto probabilistic shaping can achieve the capacity of DMCs with input alphabets of cardinality a power of two. The performance is on-par with state-of-the-art PC-PAS for short and moderate block lengths. Future research may further optimize the code for these block lengths, and investigate how the constraints induced by the deterministic DM process can be used to aid decoding.

Acknowledgment

REFERENCES

[1] R. G. Gallager, *Information theory and reliable communication*. New York, NY: John Wiley & Sons, Inc., 1968.

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

[3] 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, Oct. 2015.

[4] Y. C. Gültekin, T. Fehenberger, A. Alvarado, and F. M. J. Willems, "Probabilistic shaping for finite blocklengths: Distribution matching and sphere shaping," *Entropy*, vol. 22, no. 5, Apr. 2020.

[5] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7829–7838, Sep. 2013.

[6] N. Stolte, "Rekursive Codes mit der Plotkin-Konstruktion und ihre Decodierung," Ph.D. Thesis, Technische Universität Darmstadt, Jan. 2002. [Online]. Available: http://elib.tu-darmstadt.de/diss/000183

[7] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jun. 2009.

[8] 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," *J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, Apr 2016.

[9] G. Böcherer, P. Schulte, and F. Steiner, "Probabilistic shaping and forward error correction for fiber-optic communication systems," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 230–244, 2019.

[10] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Nov. 2015.

[11] T. Wiegart, F. Steiner, P. Schulte, and P. Yuan, "Shaped on–off keying using polar codes," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 1922–1926, Jul. 2019.

[12] D. Sutter, J. M. Renes, F. Dupuis, and R. Renner, "Achieving the capacity of any DMC using only polar codes," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Lausanne, Switzerland, Sep. 2012, pp. 114–118.

[13] M. Seidl, A. Schenk, C. Stierstorfer, and J. B. Huber, "Polar-coded modulation," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4108–4119, Sep. 2013.

[14] T. Prinz, P. Yuan, G. Böcherer, F. Steiner, O. İşcan, R. Böhnke, and W. Xu, "Polar coded probabilistic amplitude shaping for short packets," in *Proc. IEEE Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, Sapporo, Japan, Jul. 2017, pp. 1–5.

[15] G. Böcherer, T. Prinz, P. Yuan, and F. Steiner, "Efficient polar code construction for higher-order modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*. San Francisco, CA: IEEE, Mar. 2017, pp. 1–6.

[16] O. İşcan, R. Böhnke, and W. Xu, "Shaped polar codes for higher order modulation," *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 252–255, 2018.

[17] ——, "Probabilistic shaping using 5G new radio polar codes," *IEEE Access*, vol. 7, pp. 22 579–22 587, 2019.

[18] O. İşcan, R. Böhnke, and W. Xu, "Sign-bit shaping using polar codes," *Trans. Emerging Telecommun. Technol.*, vol. 31, no. 10, p. e4058, 2020.

[19] T. Matsumine, T. Koike-Akino, D. S. Millar, K. Kojima, and K. Parsons, "Polar-coded modulation for joint channel coding and probabilistic shaping," in *Optical Fiber Communication Conference (OFC)*. San Diego, CA: OSA, Apr. 2019, p. M4B.2.

[20] R. Böhnke, O. İşcan, and W. Xu, "Multi-level distribution matching," *IEEE Commun. Lett.*, vol. 24, no. 9, pp. 2015–2019, May 2020.

[21] M. Y. Şener, R. Böhnke, W. Xu, and G. Kramer, "Dirty paper coding based on polar codes and probabilistic shaping," *IEEE Commun. Lett.*, pp. 3810–3813, Sep. 2021.

[22] L. Liu, Y. Yan, C. Ling, and X. Wu, "Construction of capacity-achieving lattice codes: Polar lattices," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 915–928, Oct. 2018.

[23] G. Ungerböck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.

[24] E. Şaşoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Taormina, Italy, Oct. 2009, pp. 144–148.

[25] W. Park and A. Barg, "Polar codes for q-ary channels, $q = 2^r$," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 955–969, Sep. 2013.

[26] T. C. Gulcu, M. Ye, and A. Barg, "Construction of polar codes for arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 309–321, 2018.

[27] P. Yuan and F. Steiner, "Construction and decoding algorithms for polar codes based on 2×2 non-binary kernels," in *Proc. IEEE Int. Symp. Turbo Codes & Iter. Inf. Process. (ISTC)*, Hong Kong, China, Dec. 2018, pp. 1–5.

[28] E. Abbe and E. Telatar, "Polar codes for the m-user multiple access channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5437–5448, May 2012.

[29] E. Arıkan, "Source polarization," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, Jun. 2010, pp. 899–903.

[30] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 23, no. 3, pp. 371–377, 1977.

[31] L. Liu, "Polar codes and polar lattices for efficient communication and source quantization," Doctoral Dissertation, Imperial College London, Sep. 2016. [Online]. Available: https://spiral.imperial.ac.uk/handle/10044/1/48001

[32] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[33] R. A. Chou and M. R. Bloch, "Using deterministic decisions for low-entropy bits in the encoding and decoding of polar codes," in *Proc. Allerton Conf. Commun., Contr., Comput.* Monticello, IL: IEEE, Oct. 2015, pp. 1380–1385.

[34] M. Mondelli, S. H. Hassani, and R. L. Urbanke, "How to achieve the capacity of asymmetric channels," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3371–3393, Jan. 2018.

[35] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, Mar. 2015.

[36] G. Böcherer and B. C. Geiger, "Optimal quantization for distribution synthesis," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6162–6172, Sep. 2016.

[37] P. Schulte and B. C. Geiger, "Divergence scaling of fixed-length, binary-output, one-to-one distribution matching," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 3075–3079.

[38] U. Wachsmann, R. F. Fischer, and J. B. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[39] Z. He, T. Bo, and H. Kim, "Probabilistically shaped coded modulation for IM/DD system," *Opt. Express*, vol. 27, no. 9, pp. 12 126–12 136, Apr. 2019.

[40] D. Kim, Z. He, T. Bo, Y. Yu, and H. Kim, "Transmission of 36-Gbaud PAM-8 signal in IM/DD system using pairwise-distributed probabilistic amplitude shaping," in *Proc. Optical Fiber Commun. Conf. (OFC)*. San Diego, CA: OSA, Mar. 2020, p. M3J.3.

[41] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, Apr. 2010.