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) [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 $X$, $P_X$, and $x$, respectively. Vectors are denoted by bold symbols such as $x$. $X^C$ and $|X|$ are the complement and cardinality of $X$, respectively. A set difference is denoted as $X \setminus Y = X \cap Y^C$. An index set from 1 to $N$ is denoted as $[N] = \{1, \ldots, N\}$. A set $S$ may select entries of a vector, creating a substring $x_S$ with length $|S|$, e.g., $x_{[N]}$. An event $E$ has probability $P(E)$.

We denote by $H(X)$, $H(X|Y)$, and $I(X;Y)$ 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 \cite{9} is defined as $Z(X|Y) = 2E[\sqrt{P_{X|Y}(0)^{2}P_{X|Y}(1)^2}]$ and satisfies \cite{9} \cite{22} Lem. 6
\[ Z(X|Y)^2 \leq H(X|Y) \leq Z(X|Y) \quad \text{(1)} \]
\[ Z(X|Y,S) \leq Z(X|Y) \quad \text{(2)} \]
where $X,Y,S \sim P_{X,Y,S}$.

B. Polarization and Polar Coding

Polar codes \cite{6}, \cite{7} are linear block codes of length $N = 2^n$ for $n \in N$. They are defined through the polar transform that maps a vector $u \in \mathbb{F}_2^N$ to a codeword $x \in \mathbb{F}_2^N$
\[ x = uG_N \text{ with } G_N = B_N \begin{bmatrix} 1 & 0 \ 1 & 1 \end{bmatrix}^n \quad \text{(3)} \]
where $B_N$ is the bit-reversal matrix as in \cite{7}, and where $F^{\otimes n}$ is the $n$-fold Kronecker product of $F$. The polar transform satisfies $G_N^{-1} = G_N$. For encoding, we will consider the codeword $x$ to have independent and identically distributed (i.i.d.) entries. The codeword $x$ is transmitted over $N$ uses of a biDMC $W: X \rightarrow Y$ resulting in a vector of channel observations $y \in Y^N$. Consider the sets
\[ \mathcal{L}_{U|Y} = \{i \in [N] : Z(U_i|U_{[i-1]},Y) < \delta_N \} \quad \text{(4)} \]
\[ \mathcal{H}_U = \{i \in [N] : Z(U_i|U_{[i-1]}) > 1 - \delta_N \} \quad \text{(5)} \]
with $\delta_N \triangleq 2^{-N^\beta}$ for any $\beta < \frac{1}{2}$. It is known \cite{5} Eqs. (38), (39)] that these index sets polarize, i.e., we have
\[ \lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{L}_{U|Y}| = 1 - H(X|Y) \quad \text{(6)} \]
\[ \lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{H}_U| = H(X|Y) \quad \text{(7)} \]
The encoder places the data bits on the reliable bit positions of $u$, i.e., those with $Z(U_i|U_{[i-1]},Y) \approx 0$. The remaining positions in $u$ are frozen, i.e., set to fixed values. The receiver uses successive cancellation (SC) decoding of the non-frozen bits via $\hat{u}_i = \text{arg max}_{x} P_{U|U_{[i-1]},Y}(u|U_{[i-1]},y)$. Honda and Yamamoto \cite{5} consider two more sets:
\[ \mathcal{L}_U = \{i \in [N] : Z(U_i|U_{[i-1]}) < \delta_N \} \quad \text{(8)} \]
\[ \mathcal{H}_U = \{i \in [N] : Z(U_i|U_{[i-1]}) > 1 - \delta_N \}. \quad \text{(9)} \]
Note that \cite{7} and \cite{5} yield $|H_U|/N \xrightarrow{N \rightarrow \infty} \mathbb{H}(X)$. With \cite{5}, 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 $\mathcal{I} = \mathcal{H}_U \cup \mathcal{L}_{U|Y}$, we thus have \cite{5} Thm. 1
\[ \lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{I}| = I(X;Y). \quad \text{(10)} \]
To achieve capacity, Honda and Yamamoto chose the $I$ bits as data bits and the remaining bits in $u$ randomly with probability $P_{U|U_{[i-1]}},(u|U_{[i-1]})$. 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 $W: X \rightarrow Y$ with side information $S$ with joint distribution $(X,Y,S) \sim P_{Y|X,S}P_XP_S$. Let $(X,S,Y)$ be $N$ i.i.d. realizations of $(X,Y,S)$ and let $U = XG_N$. Then all the index sets
\[ \mathcal{L}_{U|S} = \{i \in [N] : Z(U_i|U_{[i-1]},S) < \delta_N \} \quad \text{(11)} \]
\[ \mathcal{H}_U|S = \{i \in [N] : Z(U_i|U_{[i-1]},S) > 1 - \delta_N \} \quad \text{(12)} \]
\[ \mathcal{L}_{U|S,Y} = \{i \in [N] : Z(U_i|U_{[i-1]},S,Y) < \delta_N \} \quad \text{(13)} \]
\[ \mathcal{H}_U|S,Y = \{i \in [N] : Z(U_i|U_{[i-1]},S,Y) > 1 - \delta_N \} \quad \text{(14)} \]
polarize with $\delta_N \triangleq 2^{-N^\beta}$ for any $\beta < \frac{1}{2}$, yielding
\[ \lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{L}_{U|S}| = 1 - I(X;Y|S) \quad \text{(15)} \]
\[ \lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{H}_U|S \cup \mathcal{H}_U|S,Y| = I(X;Y|S). \quad \text{(16)} \]
Proof. Using \cite{5} Eqs. (38), (39), the sets \cite{11} to \cite{14} polarize analogously to \cite{9}, \cite{7}, i.e.,
\[ 1 - \lim_{N \rightarrow \infty} \frac{|\mathcal{L}_{U|S}|}{N} = \lim_{N \rightarrow \infty} \frac{|\mathcal{H}_U|S}{N} = \mathbb{H}(X|S) \quad \text{(17)} \]
\[ 1 - \lim_{N \rightarrow \infty} \frac{|\mathcal{L}_{U|S,Y}|}{N} = \lim_{N \rightarrow \infty} \frac{|\mathcal{H}_U|S,Y}{N} = \mathbb{H}(X|S,Y). \quad \text{(18)} \]
We next show \cite{15}. Basic set theory gives
\[ |\mathcal{L}_{U|S}| + |\mathcal{H}_U|S,Y| = |\mathcal{L}_{U|S,Y}| - |\mathcal{L}_{U|S}| \cap \mathcal{H}_U|S,Y|. \quad \text{(19)} \]
By \cite{7}, 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 \cite{17} and the second term of \cite{18} into \cite{19} gives \cite{15}. To prove \cite{16}, observe that
\[ \lim_{N \rightarrow \infty} \frac{|\mathcal{H}_U|S \cap \mathcal{L}_{U|S,Y}|}{N} = 1 - \lim_{N \rightarrow \infty} \frac{|(\mathcal{H}_U|S \cup \mathcal{L}_{U|S,Y})^C|}{N} \quad \text{(20)} \]
\[ = 1 - \lim_{N \rightarrow \infty} \frac{|\mathcal{L}_{U|S}|}{N} \quad \text{(21)} \]
\[ = 1 - \lim_{N \rightarrow \infty} \frac{|\mathcal{L}_{U|S}|}{N} \quad \text{(22)} \]
where the last equality follows from \cite{17} and \cite{18}. Combining \cite{15} and \cite{22} yields \cite{16}. □
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., $x = f(x_{B,1}B,2 \ldots x_{B,m})$ where $f(\cdot)$ is invertible. A codeword has a length of $N$ symbols or $mN$ bits. Each bitlevel $x_{B,\ell}$, $\ell \in [m]$, is encoded with a separate polar transform as $u_\ell = x_{B,\ell}G_N$. Using Lem. [1], we can prove the polarization of such a multilevel polar code.

\textbf{Theorem 1.} Let $W: X \rightarrow Y$ be a DMC with joint distribution $X,Y \sim P_{X|Y}P_X$ and $|\mathcal{X}| = 2^m$. Let $X_{B,\ell}$, $\ell \in [m]$, be the $\ell$-th bit of the binary representation $X_{B}$ of symbol $X = f(X_{B})$, and let $U_{\ell} = X_{B,\ell}G_N$. Then, the sets

\begin{align}
L_U' &= \{ (\ell, i) : Z(U_{\ell i}^{|Y}) \leq \delta_N \} \\
H_U' &= \{ (\ell, i) : Z(U_{\ell i}^{|Y}) > \delta_N \} \\
L_{U|Y}' &= \{ (\ell, i) : Z(U_{\ell i}^{|Y}, Y) \leq \delta_N \} \\
H_{U|Y}' &= \{ (\ell, i) : Z(U_{\ell i}^{|Y}, Y) > \delta_N \}
\end{align}

with $\ell \in [m]$, $i \in [N]$, and $V_{\ell i}^{|Y} = (U_{\ell i}^{|Y}, X_{B,|\ell - 1})$ polarize, i.e., we have

\begin{align}
\lim_{N \rightarrow \infty} \frac{1}{N}|L_U' \cup H_U'| &= 1 - I(X;Y) \\
\lim_{N \rightarrow \infty} \frac{1}{N}|I| &= I(X;Y)
\end{align}

where $I = H_U' \cap L_{U|Y}'$.

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

\begin{align}
H_{U|Y}' &= \{ (\ell, i) : Z(U_{\ell i}^{|Y}) \leq \delta_N \} \\
L_{U|Y}' &= \{ (\ell, i) : Z(U_{\ell i}^{|Y}, Y) > \delta_N \}.
\end{align}

By Lem. [1] we have

\begin{align}
\lim_{N \rightarrow \infty} \frac{1}{N}|H_{U|Y}' \cup L_{U|Y}'| &= I(X_{B,\ell}^{|Y}, Y|X_{B,|\ell - 1}^{|})
\end{align}

Note that $H_{U|Y}' \cap L_{U|Y}' = L_{U|Y}'$, and $H_{U|Y}' \cap L_{U|Y}' = \emptyset$ by definition for $\ell \neq k$ and

\begin{align}
H_{U} &= \bigcup_{\ell \in [m]} H_{U|Y}' \\
L_{U|Y}' &= \bigcup_{\ell \in [m]} L_{U|Y}'
\end{align}

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

\begin{align}
\lim_{N \rightarrow \infty} \frac{1}{N}|I| &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\ell \in [m]} |H_{U|Y}' \cup L_{U|Y}'| \\
&= \sum_{\ell \in [m]} I(X_{B,\ell}; Y|X_{B,|\ell - 1}^{|}) \\
&= I(X;Y)
\end{align}

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

\section{Encoding and Decoding}

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

\begin{equation}
P_{X(x_i)} = \prod_{\ell \in [m]} P_{X_{B,\ell}|X_{B,|\ell - 1}}(x_{B,\ell}^{|X_{B,|\ell - 1}^{|}})
\end{equation}

For each bitlevel $\ell$, multistage decoding (MSD) computes $P_{X_{B,\ell}|X_{B,|\ell - 1}}(x_{B,\ell}^{|X_{B,|\ell - 1}^{|}})$, $i \in [N]$ [30], and provides these values to a SC decoder that computes $P_{U_{\ell i}^{|Y} | \mathcal{M}^c}$ and decides on $u_{\ell i}$.

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

\textbf{Theorem 2.} Let $W: X \rightarrow Y$ and define $\mathcal{I}$ as in Thm. 7. Let $\mathcal{M} \subseteq \mathcal{I}$, and consider encoding and decoding as described above. Then the average decoding error probability is $P(\{\hat{U} \neq U\}) = \mathcal{O}(2^{-N^2} \beta^2)$ for any $0 < \beta' < \beta < 1/2$ by choosing the polarization sets with $\delta_N = 2^{-N^2}$.

\textbf{Proof.} Consider

\begin{equation}
P(\{\hat{U} \neq U\}) = 1 - \prod_{\ell \in [m]} \left( 1 - P(\mathcal{E}^{|Y} | \mathcal{E}^{|Y}|_{\ell - 1}) \right)
\end{equation}

where $\mathcal{E}^{|Y} = \{U^{|Y} \neq \hat{U}^{|Y}\}$ and $\mathcal{E}^{|Y}|_{\ell - 1} = \{U^{|Y}|_{\ell - 1} = \hat{U}^{|Y}|_{\ell - 1}\}$. Let the equivalent channel for the $\ell$-th bitlevel be the channel that has bit $\ell$ as input and bits $[\ell - 1]$ as side-information available at transmitter and receiver. By [31] Thm. 4.3.9, [5] Thm. 3, and [1] Thm. 3, the $U_{\ell}$ code over this equivalent channel for bitlevel $\ell$ has an average decoding error probability $P(\mathcal{E}^{|Y} | \mathcal{E}^{|Y}|_{\ell - 1}) = \mathcal{O}(2^{-N^2})$ with $\beta' < \beta < 1/2$ and uniformly chosen messages.
Thus, for each bitlevel $\ell$ there is a positive constant $c_{\ell}$ and a block length $N_{\ell}$ so that $\Pr(\mathcal{E}_{\ell}^{c[\ell-1]}) \leq e^{2^{-N_{\ell}^{2}}} \Pr$ 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

$$\Pr(\mathcal{E}_{\ell}^{c[\ell-1]}) \leq e^{2^{-N^{*}}}$$

(38)

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

$$\Pr(\{\hat{U} \neq U\}) = 1 - \prod_{\ell \in [m]} \left(1 - \Pr(\mathcal{E}_{\ell}^{c[\ell-1]})\right)$$

(39)

$$\leq 1 - \left(1 - e^{2^{-N^{*}}}\right)^{m}$$

(40)

$$\leq mc^{2^{-N^{*}}}$$

(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 $\mathbb{H}(U_{i}^{\ell}|V_{i}^{\ell}, Y)$ are fixed to 0 and the bits with small $\mathbb{H}(U_{i}^{\ell}|V_{i}^{\ell})$ are decided as $\arg \max_{u} P_{U_{i}^{\ell}|V_{i}^{\ell}}(u|\hat{b}_{i}^{\ell})$ (DM bits). The non-frozen bits are data bits. The decoder estimates the non-frozen bits via $\arg \max_{u} P_{U_{i}^{\ell}|V_{i}^{\ell}, Y}(u|\hat{b}_{i}^{\ell}, y)$. The DM bits with $\mathbb{H}(U_{i}^{\ell}|V_{i}^{\ell}) \approx 0$ also have $\mathbb{H}(U_{i}^{\ell}|V_{i}^{\ell}, Y) \approx 0$ 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], [17], [20]

$$\Delta_{R} = \mathbb{H}(\hat{P}_{X}) - R = \mathbb{H}(\hat{P}_{X}) - \frac{|U|}{N}$$

(42)

where $\hat{P}_{X}$ is the empirical distribution of $X$, and $U$ indexes the bitchannels with uniformly-distributed data bits. Typically, $U$ consists of the bitchannels with $\mathbb{H}(U_{i}^{\ell}|V_{i}^{\ell}) \approx 1$.

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)

- $R = 1.625 \text{ bpcu}$ for $M = 4$;
- $R = 2.375 \text{ bpcu}$ for $M = 8$;
- $R = 3.250 \text{ bpcu}$ for $M = 16$.

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

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 MLHY increases with $M$ for short and moderate block lengths whereas the rate loss of 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

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

For both cases, we choose $P_{X}(x) \propto \exp(-\nu|x|^{2})$ so that $\nu$ 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 $\mathbb{H}(U_{i}^{\ell}|V_{i}^{\ell}, Y)$. Second, we introduce a design parameter $\kappa$ for code optimization and choose the rate-optimal $P_{X}$ at $\kappa \cdot \text{dSNR}$ as our target distribution based on which we also compute $\mathbb{H}(U_{i}^{\ell}|V_{i}^{\ell})$. 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 $N_{\text{DM}}$ of DM positions. The code is constructed by choosing the $N_{\text{DM}}$ positions with lowest $\mathbb{H}(U_{i}^{\ell}|V_{i}^{\ell})$ for DM and the $N(1-R)-N_{\text{DM}}$ positions with highest $\mathbb{H}(U_{i}^{\ell}|V_{i}^{\ell}, Y)$ 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
not fulfill this requirement. Instead, we approximate a one-
sided sampled Gaussian distribution $P_X$ for $M$-PAM by
assigning $M/2$ different probability masses to pairs of points
as described in [39], [40], i.e., $P_X(2i) = P_X(2i + 1) \forall i = 0, 1, \ldots, M/2 - 1$. The input distribution is thus suboptimal.
Second, polar-coded [28] 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 parameters.

We describe potential improvements. Because CCDM codewords 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 channel 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 (3-CRC) 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
$P_X$ is approximately 0.34 dB. Without CRC, the MLHY
curve exhibits the predicted shaping gain and outperforms
PC-PAS (3-CRC) 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.

Acknowledge

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
PC-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

The authors wish to thank Prof. Gerhard Kramer for sugges-
tions. This work was supported in part by the German Federal
Ministry of Education and Research (BMBF) under the Grant
6G-life, and by the German Research Foundation (DFG) under
Grant KR 3517/9-1.
