Technische Universität München
Lehrstuhl für Nachrichtentechnik
Prof. Dr. sc. techn. Gerhard Kramer

Master's Thesis

Higher-Order Polar-Coded Modulation for Asymmetric Channels

Vorgelegt von:
Constantin Marco Runge

München, November 2021

Betreut von:
Juan Diego Lentner Ibañez, M.Sc.
Thomas Wiegart, M.Sc.

Master's Thesis am
Lehrstuhl für Nachrichtentechnik (LNT)
der Technischen Universität München (TUM)
Titel : Higher-Order Polar-Coded Modulation for Asymmetric Channels
Autor : Constantin Marco Runge

Constantin Runge
Technische Universität München
Lehrstuhl für Nachrichtentechnik
Theresienstr. 90
80333 München
constantin.runge@tum.de

Ich versichere hiermit wahrheitsgemäß, die Arbeit bis auf die dem Aufgabensteller bereits bekannte Hilfe selbständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderung entnommen wurde.

München, November 15, 2021

Ort, Datum (Constantin Runge)

Contents

Acronyms

ASK amplitude-shift keying 11, 35, 36, 39, 40, 46, 51, 53–55, 57

AWGN additive white Gaussian noise 9–11, 27, 28, 39, 41, 44, 52–55, 57

BICM bit-interleaved coded modulation 22, 28

bpcu bits per channel use 11

BPSK binary phase-shift keying 46, 51, 53–55

CCDM constant composition distribution matching 12, 38–40

CRC cyclic redundancy check 15, 39, 40

DM distribution matching 3, 4, 12, 16, 17, 19–21, 25, 29, 34–40, 45, 48, 50, 53, 57

DMC discrete memoryless channel 3, 8, 13, 29, 30, 32, 33

DPC dirty paper coding 4, 49–51, 53–55, 57

dSNR design signal-to-noise ratio 39, 40, 57

FEC forward error correction 3, 4, 12, 13, 16, 19–21, 25, 29, 34, 35, 37–40, 48, 49, 53, 54, 57

FER frame error rate 35, 55

HY Honda-Yamamoto 4, 12, 16, 20, 21, 26–35, 37, 47–50, 53–55, 57

iff if and only if 6, 14

IM intensity modulation 3, 10, 41

LDPC low-density parity-check 15, 55

LLPS linear layered probabilistic shaping 54

LLR log-likelihood ratio 16, 27

MC Monte Carlo 15, 34, 35, 38, 51

MI mutual information 6, 7, 9, 10, 14, 15, 19, 21, 30

MLC multilevel coding 21, 22, 25, 30, 32, 33, 47, 50, 57

MLHY multilevel Honda-Yamamoto 4, 25, 28, 30–41, 47, 50, 51, 53–55, 57, 58

MLPC multilevel polar code 40, 41

MSD multistage decoding 21, 23, 30, 31

OOK on-off keying 27

PAM pulse-amplitude modulation 10, 11, 41

PAS probabilistic amplitude shaping 12, 39, 41, 45, 54

PC-PAS polar coded probabilistic amplitude shaping 39–41, 57

RCUB random coding union bound 39–41

SC successive cancellation 15, 17, 19–21, 27, 31–34, 57

SCL successive cancellation list 15, 19, 23, 33, 38, 39, 54

SE spectral efficiency 11, 34, 39–41, 53, 55, 57

SIR signal-to-interference ratio 46, 51

SNR signal-to-noise ratio 10, 11, 20, 35, 36, 38, 46, 51, 55

Abstract

This thesis proposes a coded modulation scheme to implement probabilistic shaping with arbitrary symbol distributions. The considered scheme is based on polar codes, a class of capacity-achieving, linear block codes with state-of-the-art finite-length performance and an explicit construction. The scheme is analysed from a theoretical standpoint as well as with numerical simulations. Furthermore, an extension to this scheme for state-dependent shaping is proposed, as well as an application to dirty paper coding. The effectiveness of this extension is shown in numerical simulations and its convergence properties are discussed.

1 Introduction

In 1948, Shannon [Sha48] proved maximally achievable information rates for a given noisy channel using codes with infinite lengths. Ever since then, communications engineers and information theorists are working on communication schemes which can approach these information rates using codes with finite lengths. Traditionally, linear forward error correction (FEC) codes are used which produce uniformly distributed channel input symbols for uniformly distributed data bits. While these codes can provide decoding rules with low probability of error without requiring too much overhead for redundancy, we know from Shannon's results that the optimal input distribution that achieves the channel capacity is generally not uniform. In order to further increase efficiency of communication, one can make use of distribution matching (DM) in addition to FEC. The goal of DM is to generate codewords with symbols that are distributed according to more beneficial distributions. As a FEC stage after DM spoils the symbol distribution and a DM stage the FEC destroys desirable properties of the FEC code, FEC and DM have to be performed in a coordinated manner. Such a joint coding for FEC and DM is called probabilistic shaping.

While there exist efficient shaping schemes for performing minimum-energy DM [For92], [KK93] or DM for symmetric target distributions, [BSS15], there is a lack of shaping schemes that are able to perform DM with arbitrary target distributions. However, arbitrary target distributions come up in certain cases, for example for communication over the intensity modulation (IM) channel or over the dirty paper channel [GP80], [Len18]. For these channels, the optimal input distribution does not contain any symmetries and thus, arbitrary symbol shaping is required.

We propose a shaping scheme based on polar codes. Polar codes [Sto02] are the first class of codes that were proven to achieve the capacity of discrete memoryless channels (DMCs) [Arı09]. Not only do they provide an explicit construction, but they can also be encoded and decoded with a complexity of $\Theta(N \log N)$ with block length N [Arı09]. Thanks to efficient list decoding [TV15], polar codes proved to be competitive for short and moderate block lengths, c.f. [CDJ+19]. Furthermore, there exist straightforward and powerful extensions to coded modulation [SSSH13], [PYB+17] as well as to shaping

[HY13], [WSSY19], making polar codes a good candidate for this still ongoing search for capacity-approaching communication systems.
In this thesis we build upon multilevel polar coded modulation [Ung82], [SSSH13] and Honda-Yamamoto (HY) codes for probabilistic shaping [HY13]. We combine these two concepts and propose multilevel Honda-Yamamoto (MLHY) coded modulation; a joint DM and FEC coding scheme. We show that MLHY coding is able to realize arbitrary target distributions on higher-order modulation alphabets. Furthermore, we extend existing polar coding proofs to MLHY coding and show that the proposed scheme achieves the constellation-constrained capacity of a memoryless channel and provide error exponents for the case of finite block lengths. To complement the theoretical results, we provide simulated performance results for DM and shaping scenarios. In addition to that, we discuss theoretical aspects to the applicability of HY coding and MLHY to state-dependent shaping, as is required for dirty paper coding (DPC). We give a short review on approaches to this problem and provide simulated results to show the practicability of the assumptions required for the state-dependent MLHY DPC scheme.
The remaining thesis is structured as follows. Chapter 2 and Chapter 3 provide some background on general aspects of information theory and a more focused review on polar coding, respectively. After this, Chapter 4 introduces the polar coding-based shaping scheme for higher-order coded modulation. For this scheme, we provide a theoretical analysis as well as simulated performance results. In Chapter 5, we then give a short introduction to channels with state and to DPC in particular. Chapter 6 extends the scheme proposed in Chapter 4 to state-dependent shaping and shows an exemplary application to DPC. The extension is again discussed from a theoretical standpoint and simulated performance results are presented. Finally, Chapter 7 concludes this thesis.

2 Preliminaries

2.1 Notation

Random variables X or Y with distribution $X, Y \sim P_{X,Y}$ are denoted by upper case symbols whereas their realizations are denoted by lower case symbols x and y, respectively. Vectors of realizations $\boldsymbol{x}$ or of random variables $\boldsymbol{X}$ are denoted by bold symbols. Elements x_i of vectors are denoted with lowered indices. Uppercase bold letters may also denote matrices $\boldsymbol{G}$. Sets $\mathcal{X}$ are denoted by calligraphic uppercase letters, their complement by $\mathcal{X}^C$ and their cardinality by $|\mathcal{X}|$. The set difference is denoted as $\mathcal{X} \setminus \mathcal{Y} = \mathcal{X} \cap \mathcal{Y}^C$. An index set from N to M is denoted as $N{:}M \triangleq \{N, \dots, M\}$. For index sets starting at index 1, we write $[\![N]\!] \triangleq 1{:}N = \{1, \dots, N\}$. An index set $\mathcal{S}$ may also index into a vector, creating a substring $\boldsymbol{x}_\mathcal{S}$ with length $|\mathcal{S}|$, e. g., $\boldsymbol{x}_{[\![N]\!]}$. Blackboard uppercase letters $\mathbb{F}$ denote fields.

All logarithms $\log x$ are with respect to base 2 if not stated otherwise. The Kronecker power $\boldsymbol{F}^{\otimes n}$ of a matrix is defined as $\underbrace{\boldsymbol{F} \otimes \cdots \otimes \boldsymbol{F}}_{n \text{ times}}$ with the tensor product $\boldsymbol{A} \otimes \boldsymbol{B}$.

2.2 Quantities of Information

The definitions and relations introduced in the following review can be found in [CT06]. Let P_X denote the probability distribution of a random variable X on a finite alphabet, p_X denote the probability mass function of a discrete random variable and f_X denote the probability density function of a continuous random variable. The probability of some general event $\mathcal{E}$ is denoted by $\mathcal{P}(\mathcal{E})$. The converse probability is denoted by $\mathcal{P}(\neg\mathcal{E}) = 1 - \mathcal{P}(\mathcal{E})$. Empirical stochastic quantities $\widehat{\mathcal{P}}(x)$ are denoted with a hat symbol. Conditioning a random variable X on Y is denoted as $X|Y$. The set of all probability mass functions over alphabet $\mathcal{X}$ is denoted as $\Pi(\mathcal{X})$.

A normally distributed random variable with mean μ and variance σ^2 is denoted by $X \sim \mathcal{N}(\mu, \sigma^2)$ and a binary random variable $X \in \mathbb{F}_2$ with $p_X(1) = q$ is denoted by $X \sim \mathrm{Ber}(q)$.

The entropy of some probability distribution p quantifying the accompanying uncertainty is defined as

$$\mathbb{H}(p) = -\sum_{x \in \mathrm{supp}(p)} p(x) \log p(x), \tag{2.1}$$

where $\mathrm{supp}(g) = \left\{x \in \mathcal{X} : g(x) \neq 0\right\}$ is the support of a function $g : \mathcal{X} \to \mathbb{R}$. For a discrete random variable $X \sim p_X$, we define the entropy $\mathbb{H}(X)$ of a random variable as a shorthand for

$$\mathbb{H}(X) \triangleq \mathbb{H}(p_X) = -\,\mathbb{E}[\log p_X(X)], \tag{2.2}$$

where $\mathbb{E}[\cdot]$ denotes the expectation value. In the same manner, the empirical entropy $\widehat{\mathbb{H}}(X)$ of a random variable is defined as the entropy of the respective empirical distribution

$$\widehat{\mathbb{H}}(X) \triangleq \mathbb{H}(\widehat{p}_X). \tag{2.3}$$

For a continuous random variable $X \sim f_X$, the differential entropy is defined likewise as

$$h(X) = -\,\mathbb{E}[\log f_X(X)]. \tag{2.4}$$

Similarly, the conditional entropy for two random variables $X, Y \sim p_{X,Y}$ is defined as

$$\mathbb{H}(X|Y) = -\,\mathbb{E}[\log p_{X|Y}(X|Y)] \tag{2.5}$$

and analogously for continuous random variables.

For discrete random variables,

$$0 \leq \mathbb{H}(X|Y) \leq \mathbb{H}(X) \leq \log|\mathcal{X}| \tag{2.6}$$

with equalities $\mathbb{H}(X) = 0$ for fully deterministic variables, $\mathbb{H}(X) = \log|\mathcal{X}|$ for $X \sim \mathcal{U}(\mathcal{X})$, where $\mathcal{U}$ denotes the uniform distribution over $\mathcal{X}$, and $\mathbb{H}(X) = \mathbb{H}(X|Y)$ if and only if (iff) X and Y are stochastically independent. For differential entropies, only the relation between $h(X|Y)$ and $h(X)$ holds.

Finally, the mutual information (MI) between two random variables X and Y is defined as

$$\mathbb{I}(X;Y) = \mathbb{H}(X) - \mathbb{H}(X|Y) = \mathbb{H}(Y) - \mathbb{H}(Y|X) \tag{2.7}$$

with straightforward extension to differential entropies. Furthermore, a conditioned definition exists with

$$\mathbb{I}(X;Y|S) = \mathbb{H}(X|S) - \mathbb{H}(X|Y,S) = \mathbb{H}(Y|S) - \mathbb{H}(Y|X,S). \tag{2.8}$$

Both, the MI as well as the entropy, obey the so-called chain rule so that

$$\mathbb{H}(X_1 X_2 \dots X_N) = \sum_{i \in [\![N]\!]} \mathbb{H}(X_i | X_1, \dots, X_{i-1}) \tag{2.9}$$

as well as

$$\mathbb{I}(X_1 X_2 \dots X_N; Y) = \sum_{i \in [\![N]\!]} \mathbb{I}(X_i; Y | X_1, \dots, X_{i-1}) \tag{2.10}$$

holds.

2.3 Channel Models

2.3.1 General Channel Properties

Let $W : X \to Y$ denote a channel with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and transition probabilities $W(y|x)$.
Such a channel W is discrete if both input and output are discrete random variables.
A channel $W(y|x)$ is called strongly symmetric if it satisfies

$$\big\{W(y|x) : y \in \mathcal{Y}\big\} = \big\{W(y|x') : y \in \mathcal{Y}\big\} \, \forall x, x' \in \mathcal{X}, \tag{2.11}$$

$$\big\{W(y|x) : x \in \mathcal{X}\big\} = \big\{W(y'|x) : x \in \mathcal{X}\big\} \, \forall y, y' \in \mathcal{Y}. \tag{2.12}$$

Such a channel has equal sets of transition probabilities for every input symbol and equal sets of input likelihoods for each output symbol. When the output alphabet $\mathcal{Y}$ of a channel W can be partitioned into sets $\mathcal{Y}_i$, so that each channel $\mathcal{X} \to \mathcal{Y}_i$ is strongly symmetric, the channel W is said to be symmetric.
Furthermore, a channel W is memoryless, if $W(y_i|x_i)$ does not depend on the previous symbols $\boldsymbol{x}_{[\![i-1]\!]}$ and $\boldsymbol{y}_{[\![i-1]\!]}$. For a memoryless channel, the joint probability of N consecutive accesses is

$$W^N(\boldsymbol{y}_{[\![N]\!]} | \boldsymbol{x}_{[\![N]\!]}) = \prod_{i \in [\![N]\!]} W(y_i|x_i). \tag{2.13}$$

Shannon [Sha48] proved that for coding over a memoryless channel with block length $N \to \infty$ and rate $R < \mathbb{I}(X;Y)$, there exists a code whose decoding error probability $\mathcal{P}(\mathcal{E}) \to 0$, where $\mathcal{E}$ is the event of a block decoding error. Furthermore, for $R > \mathbb{I}(X;Y)$ no such code exists and $\mathcal{P}(\mathcal{E}) \to 1$ for $N \to \infty$. Maximizing the MI with respect to the channel input distribution provides this channel capacity

$$C = \max_{P_X} \mathbb{I}(X;Y). \tag{2.14}$$

Also, one can define constrained capacities s.t. $P_X \in \pi \subseteq \Pi(\mathcal{X})$ such as the maximally achievable rate under an average power constraint

$$C = \max_{P_X} \mathbb{I}(X;Y) \text{ s.t. } \mathbb{E}[\,\|X\|^2] \leq P_{\text{tx}}. \tag{2.15}$$

Shannon used random codes to prove these results. However, these codes are infeasible for practical use, which motivates the search for practically usable channel codes with good finite block length performances.

For a symmetric DMC W, the unconstrained channel capacity is achieved with a uniform input distribution $P_X = \mathcal{U}(\mathcal{X})$. The symmetric capacity of a not necessarily symmetric channel is defined as $\mathbb{I}(W) = \mathbb{I}(X;Y)\big|_{X\sim\mathcal{U}(\mathcal{X})} = \log|\mathcal{X}| - \mathbb{H}(X|Y)\big|_{X\sim\mathcal{U}(\mathcal{X})}$, which coincides with the channel capacity for symmetric channels.

In a coded modulation setting with continuous channel input $\mathcal{X}$, the effective input alphabet $\mathcal{X}_{\text{CM}}$ of W as well as its distribution f_X is constrained by the signal constellation $\mathcal{X}_{\text{CM}} \subset \mathcal{X}$. We call the achievable rate under this constraint the constellation-constrained channel capacity. This is also known as the coded modulation capacity in the literature.

Channels $W(y|x,s)$ can also be state dependent. When the channel state is known to the transmitter as well as to the receiver acausally [EK11, Section 7.4], the capacity of the system is

$$C = \max_{P_{X|S}} \mathbb{I}(X;Y|S) = \max_{P_{X|S}} \mathbb{H}(X|S) - \mathbb{H}(X|Y,S)\ . \tag{2.16}$$

A special case where the state is unknown to the receiver is considered in Chapter 5.

2.3.2 Degraded Channels

A so-called broadcast channel $W(y_1, y_2|x)$ with outputs y_1 and y_2 is said to be physically degraded if X — Y_1 — Y_2 forms a Markov chain.

We say that a conditional probability distribution $P_{Y_2|X}$ is stochastically degraded with respect to a distribution $P_{Y_1|X}$, denoted as $P_{Y_2|X} \preceq P_{Y_1|X}$, if there exists a distribution $\widetilde{P}_{Y_2|Y_1}$ such that

$$P_{Y_2|X}(y_2|x) = \sum_{y_1 \in \mathcal{Y}_1} \widetilde{P}_{Y_2|Y_1}(y_2|y_1) P_{Y_1|X}(y_1|x). \tag{2.17}$$

This is the case when $P_{Y_2|X}$ and $P_{Y_1|X}$ are the marginals of some physically degraded broadcast channel and X — Y_1 — Y_2 forms a Markov chain. One interpretation of this is that everything that can be decoded from $\boldsymbol{Y}_2$ can also be decoded from $\boldsymbol{Y}_1$. The channel $P_{Y_1|X}$ is in some sense more reliable than the channel $P_{Y_2|X}$.

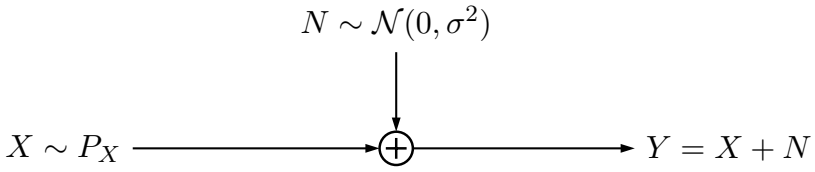


Figure 2.1: AWGN channel.

For two channels $P_{Y_2|X} \preceq P_{Y_1|X}$,

$$\mathbb{I}(X;Y_2) \leq \mathbb{I}(X;Y_1) \tag{2.18}$$

because X — Y_1 — Y_2 forms a Markov chain. By subtracting $\mathbb{H}(X)$ and multiplying with -1 at both sides, we also get

$$\mathbb{H}(X|Y_2) \geq \mathbb{H}(X|Y_1). \tag{2.19}$$

We remark that the statement cannot be made in the opposite direction. In general, one cannot infer $P_{Y_2|X} \preceq P_{Y_1|X}$ from $\mathbb{I}(X;Y_2) \leq \mathbb{I}(X;Y_1)$ for two channels with the same input X [EK11, Section 5.6].

2.3.3 The AWGN Channel

The additive white Gaussian noise (AWGN) channel with noise variance σ^2 is a continuous, symmetric channel $Y = X + N$ with additive noise $N \sim \mathcal{N}(0, \sigma^2)$ that is stochastically independent of X. The channel can also be modeled as $W(\cdot|x) = \mathcal{N}(x, \sigma^2)$. A block diagram is shown in Figure 2.1.

The MI of an AWGN channel is

$$\mathbb{I}(X;Y) = h(Y) - h(Y|X) = h(Y) - \frac{1}{2}\log(2\pi e\sigma^2) \tag{2.20}$$

for $\mathcal{X} \subseteq \mathbb{R}$. Observing that the only term dependent on P_X is $h(Y) = h(X+N)$, one can use the independence of X and N to conclude that

$$P_X^* = \arg\max_{P_X} \mathbb{H}(X) \tag{2.21}$$

is required to achieve capacity. This particular P_X^* is called maximum-entropy distribution.

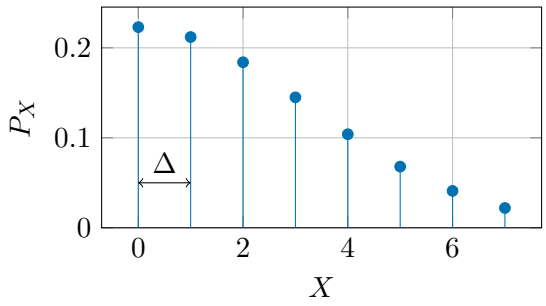


Figure 2.2: PAM Constellation and distribution for IM with $M = 8$, $\Delta = 1$, and $\sqrt{\mathbb{E}[\,|X|^2]} \approx 3$.

For $\mathcal{X} = \mathbb{R}$ and an average power constraint $\mathbb{E}[\,\|X\|^2] \leq P_{\text{tx}}$, this maximum-entropy distribution is a Gaussian distribution. For discrete $\mathcal{X}$ and an average power constraint, P_X resembles a discretized Gaussian distribution where

$$P_X(x) \propto \exp(-c\|x\|^2) \tag{2.22}$$

with factor c depending on the specific $\mathcal{X}$ as well as on the constraint. This is also referred to as Maxwell-Boltzmann distribution in some literature [KP93]. The continuous AWGN channel has a channel capacity of $C = \frac{1}{2}\log(1+\gamma)$ with the signal-to-noise ratio (SNR) $\gamma = \frac{\mathbb{E}[\,\|X\|^2]}{\sigma^2}$. Complex signalling achieves a channel capacity of $C = \log(1+\gamma)$.

2.3.4 The Intensity Modulation Channel

In many optical communication systems, transmission schemes only modulate the power of the transmitted signal. As the power is a non-negative scalar, the possible constellations are $\mathcal{X} \subset \mathbb{R}_0^+$. Other than that, this so-called IM channel can be modelled as an AWGN channel [WDY+20].

We assume a finite pulse-amplitude modulation (PAM) constellation with M equidistant symbols as shown in Figure 2.2. The parameter Δ specifies the distance between two adjacent symbols. As the channel is an AWGN channel, the constellation-constrained capacity can again be achieved with a maximum entropy distribution. Using an average power constraint, this distribution is again of the form $P_X(x) \propto \exp(-c\|x\|^2)$. For given SNR and M, the constellation-constrained MI $\mathbb{I}(X;Y)$ is a function of Δ, which makes the optimal $\Delta^*(\gamma)$ achieving the capacity a function of the SNR [BSS15].

As the constellation is one-sided, each symbol is associated with a probability that is distinct from the probabilities of the other symbols. Furthermore, when each symbol x is represented by $m = \log M$ bits $x^{\text{B},1}, \ldots, x^{\text{B},m}$, then there is no decomposition $P_X(x) =$

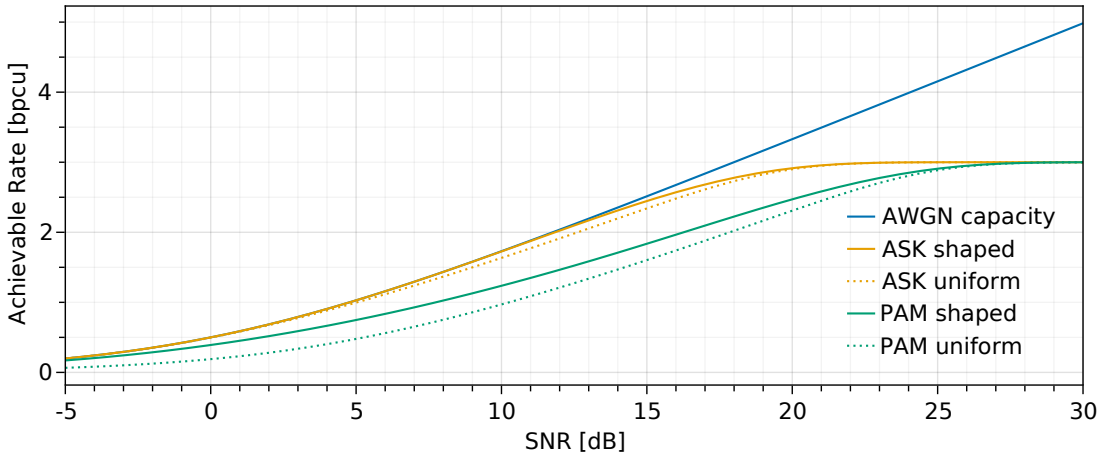


Figure 2.3: Achievable rates for ASK and IM with $M = 8$.

$\mathcal{P}(x^{\mathrm{B},[\![m-1]\!]}) \cdot \mathcal{P}(x^{\mathrm{B},m})$, i. e., it is not sufficient to distribute a bit $x^{\mathrm{B},\ell}$ according to its marginal, because $\mathcal{P}(x^{\mathrm{B},\ell}) \neq \mathcal{P}(x^{\mathrm{B},\ell}|x^{\mathrm{B},[\![\ell-1]\!]})$. As we will see in the next Section, these aspects impose challenges on a practical communication scheme.

2.4 Probabilistic Shaping

In order to close the gap between channel capacity and practically implementable systems, a transmission scheme aims at maximizing spectral efficiency (SE) and at minimizing the probability of decoding error. The SE is defined as the rate in bits per channel use (bpcu), at which information can be transmitted. By using non-uniform symbol distributions P_X, e. g., the maximum-entropy distributions discussed in the last two sections, higher SEs becomes possible without requiring larger SNR. The shaping gain is defined as the reduction in required SNR by using shaped instead of uniform transmission for a given achievable rate. Figure 2.3 compares the achievable rates for amplitude-shift keying (ASK) and PAM over the AWGN channel for uniform as well as optimal signalling. For PAM with $M = 8$, shaping gains of over 2 dB are possible.

Generally, a communication system may assume uniformly distributed messages. For most coded transmission schemes, also the parity bits are distributed uniformly [BSS15]. The generation of non-uniformly distributed codewords requires probabilistic shaping.

In accordance to the terminology from [CO90], we focus on direct probabilistic shaping approaches. Direct probabilistic shaping tries to generate energy-optimal transmission sequences by generating codewords, so that the empirical distribution of the transmitted symbols follows a target distribution $\widetilde{P}_X$. This target distribution over the constellation

is designed to approximate the distribution for which the channel capacity is achieved. For this reason, these approaches are also often referred to as DM [GFAW20]. In case the DM is implemented via a code, it is often called a shaping code.
To motivate the need for practical shaping schemes, we give a quick overview over existing ideas and their limitations. We do not aim at covering all existing concepts and their derived schemes exhaustively. A summary of existing schemes can be found in [BSS15] and [GFAW20] with further references. We characterize a few conceptually important schemes in the following list.

Gallager's scheme proposed in [Gal68] and also known as alphabet extension, maps many code symbols to one transmission symbol. With this, uniform code symbols can realize arbitrary distributions $P_X(x) = \frac{a}{b}$ represented by rational numbers. As the size of the extended alphabet increases with the denominator of the desired probabilities, this scheme becomes infeasible for many scenarios.

Trellis shaping proposed in [For92] uses the trellis decoder of a convolutional shaping code to generate the transmission sequence from an encoded data sequence. This effectively finds a minimum energy shaping codeword for a given FEC codeword. This shaping codeword then represents the transmission sequence sent to the receiver. The main problem of using trellis shaping in practical communication schemes is the design and rate matching of the shaping code.

Probabilistic amplitude shaping (PAS) proposed in [BSS15] combines DM with linear FEC. PAS uses the uniformly distributed parity bits of the FEC as sign bits, thus requiring a symmetric target distribution. The non-uniform amplitude shaping is realized via DM such as the constant composition distribution matching (CCDM) scheme proposed by [SB15] which maps a uniform, fixed-length input sequence to a fixed-length output sequence with the desired distribution P_X by using sequence look-ups of sufficient size. PAS itself can only synthesize certain distributions. In particular, the uniform parity bit shaping requires a distribution for symbols x represented by m bits $\boldsymbol{x}^{\mathrm{B},[\![m]\!]}$ to be decomposable as $P_X(x) = \frac{1}{2}\,\mathcal{P}(\boldsymbol{x}^{\mathrm{B},[\![m-1]\!]})$. Also, the incorporated DM may suffer from large look-ups or suboptimal rates. Extensions to non-uniform probabilistic shaping of the parity bits exist [BLCS19]. The syndrome DM proposed there is non-trivial to implement efficiently.

HY coding is explained in more detail in Section 3.2.

3 Polar Coding

This chapter presents polar coding proposed by Arıkan [Arı09] and its extensions to coded modulation and probabilistic shaping proposed by Seidl et al. [SSSH13], and Honda and Yamamoto [HY13], respectively.

3.1 Polar Codes for Error Correction

Polar codes, originally proposed for FEC in [Sto02], [Arı09], are an instance of linear block codes where the general idea is to transform $N = 2^n$, $n \in \mathbb{N}$, consecutive accesses to a symmetric binary-input DMC $W : \mathbb{F}_2 \to \mathcal{Y}$ to a combined channel $W_N : \mathbb{F}_2^N \to \mathcal{Y}^N$. An individual channel $W_N^{(i)}(\boldsymbol{y}, \boldsymbol{u}_{[\![i-1]\!]}|u_i)$ with data bits $\boldsymbol{u}$, observations $\boldsymbol{y}$ and index $i \in [\![N]\!]$ is called bitchannel. The transform is constructed in a way, so that with $N \to \infty$, the individual bitchannels tend to polarize to be either deterministic or completely useless with transition probability $p = \frac{1}{2}$. Bitchannels that are useless are then set to frozen values known at the receiver whereas deterministic bitchannels are used for data transmission. Following Arıkan's original construction, this polarizing transform is achieved by recursively applying a 2×2 base transform also known as polarization kernel. The originally proposed base transform encodes a vector $\boldsymbol{u} = [u_1, u_2] \in \mathbb{F}_2^2$ using a kernel matrix

$$\boldsymbol{F} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \tag{3.1}$$

to a codeword $\boldsymbol{x} = \boldsymbol{u} \cdot \boldsymbol{F} = [u_1 + u_2, u_2]$. The transform is visualized in Figure 3.1.

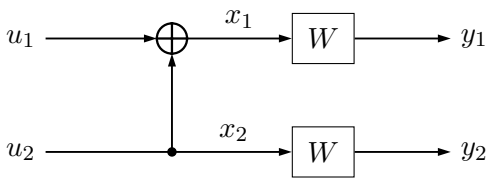


Figure 3.1: Combined channel W_2.

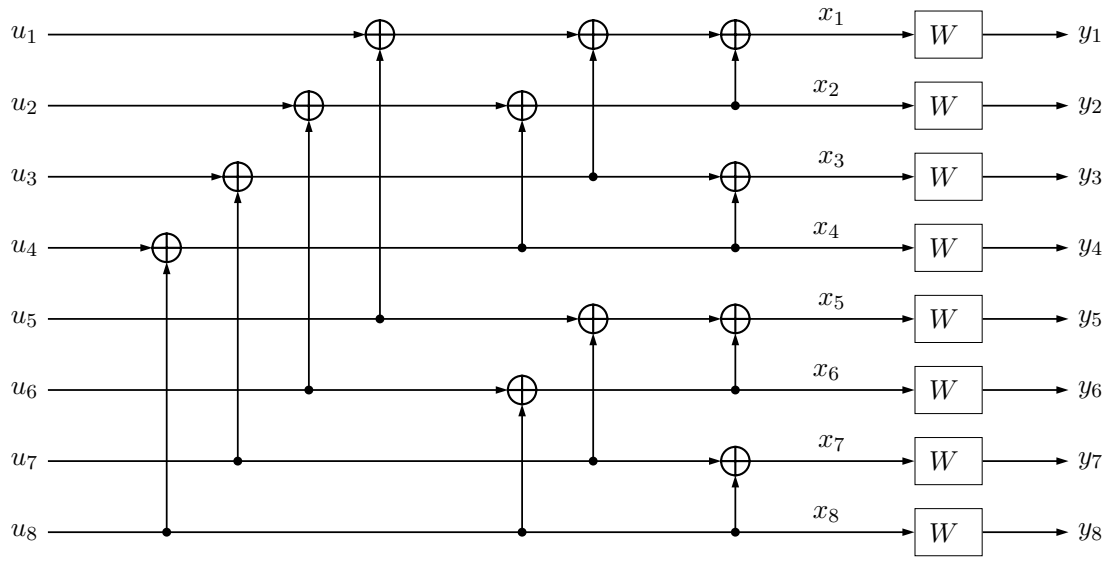


Figure 3.2: Polar transform for $N = 8$.

Recursively applying this transform leads to a transform $\boldsymbol{F}^{\otimes n}$ with

$$\boldsymbol{x} = \boldsymbol{u} \cdot \boldsymbol{F}^{\otimes n}. \tag{3.2}$$

A transform for a block length of $N = 8$ is shown in Figure 3.2. Each group of XOR-nodes stemming from the recursive application is called a polarization stage.

As the linear kernel $\boldsymbol{F}$ as well as its recursive application $\boldsymbol{F}^{\otimes n}$ is invertible, the total entropies of the random vectors $\boldsymbol{U}$ and $\boldsymbol{X}$, $\mathbb{H}(\boldsymbol{U})$ and $\mathbb{H}(\boldsymbol{X})$, are equal. From this, also the MI stays invariant under the base transform and $\mathbb{I}(\boldsymbol{U};\boldsymbol{Y}) = \mathbb{I}(\boldsymbol{X};\boldsymbol{Y}) = N\,\mathbb{I}(W)$. Furthermore, it was shown by [Arı09], [AT09] that $\sum_{i\in[\![N]\!]} \mathbb{I}(W_N^{(i)}) = \mathbb{I}(\boldsymbol{X};\boldsymbol{Y})$ and

$$\lim_{N\to\infty} \frac{1}{N} \left| \left\{ i : Z(W_N^{(i)}) \leq 2^{-N^\beta} \right\} \right| = \mathbb{I}(W), \tag{3.3}$$

$$\lim_{N\to\infty} \frac{1}{N} \left| \left\{ i : Z(W_N^{(i)}) \geq 1 - 2^{-N^\beta} \right\} \right| = 1 - \mathbb{I}(W) \tag{3.4}$$

with some $\beta < \frac{1}{2}$ and the Bhattacharyya parameter $Z(W) = \sum_y \sqrt{W(y|0)W(y|1)}$ where $\mathbb{I}(W) \approx 1$ iff $Z(W) \approx 0$ and $\mathbb{I}(W) \approx 0$ iff $Z(W) \approx 1$. Therefore, for any $0 < \delta < 1$,

$$\lim_{N\to\infty} \frac{1}{N} \left| \left\{ i : \mathbb{I}(W_N^{(i)}) > 1 - \delta \right\} \right| = \mathbb{I}(W), \tag{3.5}$$

$$\lim_{N\to\infty} \frac{1}{N} \left| \left\{ i : \mathbb{I}(W_N^{(i)}) < \delta \right\} \right| = 1 - \mathbb{I}(W). \tag{3.6}$$

Applying the chain rule of MI results in

$$\mathbb{I}(\boldsymbol{U};\boldsymbol{Y}) = \sum_{i=1}^{N} \mathbb{I}(U_i;\boldsymbol{Y}|\boldsymbol{U}_{[\![i-1]\!]}). \tag{3.7}$$

This decomposition gives rise to the so-called successive cancellation (SC) decoding, a sequential soft-in hard-out decoding rule that first decodes u_1 and then u_2 given u_1 recursively. For this, the SC decoder calculates the probabilities $\mathcal{P}(u_i|\boldsymbol{u}_{[\![i-1]\!]}, \boldsymbol{y})$ from the provided $\mathcal{P}(x_i|y_i)$ for $\boldsymbol{u} = \boldsymbol{x}(\boldsymbol{F}^{\otimes n})^{-1}$ so that $\boldsymbol{x}$ is a codeword, i. e., so that $\boldsymbol{u}$ adheres to the known frozen bit positions and values.

With multiple polarization stages, this algorithm has a complexity of $\mathcal{O}(N \log N)$ and executes a specific decoding order that is different from the natural bit order. As the definition of a bitchannel and the decomposition using the chain rule for MI rely on the previous bits and thus on the decoding order, often $\boldsymbol{B}_N\boldsymbol{F}^{\otimes n}$ with the bit-reversal matrix $\boldsymbol{B}_N$ is used as polar transform instead. The bit-reversal matrix acts as a permutation that assigns each index its bit-reversed index.

Example 1. For a polar code with $n = 4$ stages and thus a block length of $N = 2^4 = 16$, each bitchannel index is represented by four bits. In that case, e. g., the index 12 (1100) is assigned to index 3 (0011) and is therefore decoded fourth after bits 0, 8 and 4.

As the SC principle is based on the chain rule for MI, the decoder can be shown to be capacity achieving in the case of $N \to \infty$ [AT09]. In the finite-block length case, the SC decoder was shown to have decoding error probabilities of $\mathcal{P}(\mathcal{E}) = \mathcal{O}(2^{-N^\beta})$ [KŞU10]. In order to increase performance at smaller block lengths, Tal and Vardy [TV15] proposed a successive cancellation list (SCL) decoder with decoding complexity of $\mathcal{O}(LN \log N)$ with list size L. This decoding rule allows to achieve error rates comparable with the more complex and often larger low-density parity-check (LDPC) codes at short and moderate block lengths around and also below 2^{13} c.f. [CDJ+19]. To increase the performance of the SCL decoder further, an outer cyclic redundancy check (CRC) code can be used to prune invalid list items [TV15].

For the construction of a polar code and the allocation of data bits and frozen bits, the symmetric capacity of each bitchannel needs to be determined. Those capacities depend on the channel and need to be computed for the specific channel. This can always be done in a Monte Carlo (MC) simulation, but more efficient approximations are available [MT09], [TV13], [Tah17]. We remark that the theoretical polarization results as well as the computed reliabilities use SC decoding. For SCL decoding, one may need to alter the bitchannel allocation [RV19], [YPB+19].

Extensions to systematic codes [LZH15] and log-likelihood ratio (LLR)-based decoding [BBB15] exist.

3.2 Polar Codes for Probabilistic Shaping

Besides their capacity-achieving property for symmetric channels, polar codes also proved to be asymptotically optimal for lossy and lossless source coding [KU10a], [Arı10]. This property allows them to be used for channel input shaping as well, by employing a shaping decoder at the transmitter similar to [For92].

Furthermore, polar codes decouple the error-correcting or information compressing redundancy from the code in terms of its codewords. While the redundancy is defined by the polar transform itself, the set of valid codewords is defined by choosing the positions and values for the frozen bits. This can be used to construct structurally similar codes with different codewords and also codes with partly overlapping codewords.

Based on this, Honda and Yamamoto [HY13] proposed a transmission scheme that allows for joint FEC and DM using a single polar code. This polar code is defined by the positions and values for the frozen bits as well as an additional set of bitchannel indices called the shaping set. The bitchannels with the highest $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y})$ are again allocated for frozen bits. For DM, the bitchannels with the lowest $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]})$ are chosen. We call this transmission scheme HY coding, to differentiate it from other polar code-based shaping schemes such as [MKM+19], [ZLJ+21].

This construction can be interpreted as follows. The bits chosen for FEC have $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y}) \approx 1$. This entropy quantifies the uncertainty about U_i, or the "missing information" about U_i which still remains after knowing the channel observations $\boldsymbol{y}$ as well as the previous bits $\boldsymbol{u}_{[\![i-1]\!]}$. The larger this uncertainty, the higher the chance that the decoder guesses the bit u_i wrongly. If the uncertainty regarding bit U_i is 1 bit, the bit is equivalent to a coin flip. This bit cannot be decoded successfully and is added to the set of frozen bits, as knowledge about this bit is required for the decoding of later bits. If the uncertainty regarding this bit is low, it can be decoded successfully with high probability. The same idea applies to $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]})$, which quantifies the available entropy at the transmitter. For non-uniform channel input distributions, one has $\mathbb{H}(\boldsymbol{X}) < N \log|\mathcal{X}|$. Therefore, not all bitchannels will have $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}) \approx 1$. For data bits, $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}) \approx 1$ and the transmitter can fit 1 bit of information into bit U_i. For DM bits on the other hand, $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}) \approx 0$ and the bit U_i does not exhibit sufficient entropy, for the transmitter to be able to fill it with data. These bits are decided based on the previous bits. The construction is visualized in Figure 3.3.

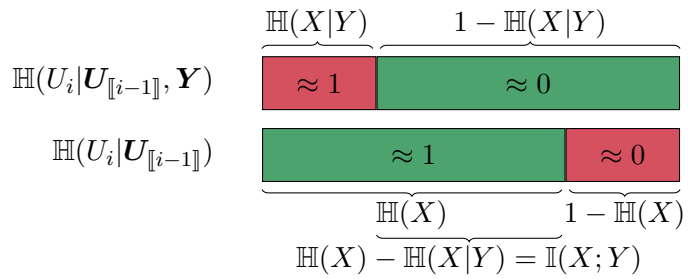


Figure 3.3: Bitchannel allocation for Honda-Yamamoto code [WSSY19].

The scheme in [HY13] modifies the encoder as well as the decoder. The encoder at the transmitter is replaced with a probabilistic decoder that finds a random data word $\boldsymbol{u}$ with the given data bits and frozen bits so that the empirical distribution of the respective codeword $\boldsymbol{x} = \boldsymbol{u}\boldsymbol{F}^{\otimes n}$ obey the target bit probabilities. This shaping decoder can use the original SC architecture that calculates the probabilities $\mathcal{P}(u_i|\boldsymbol{u}_{[\![i-1]\!]})$. Instead of deciding for the most likely bit $\arg\max_{u_i} \mathcal{P}(u_i|\boldsymbol{u}_{[\![i-1]\!]})$, the probabilistic decoder decides u_i randomly with probability $\mathcal{P}(u_i|\boldsymbol{u}_{[\![i-1]\!]})$. The receiver uses two-fold SC decoding. Data bits are decoded from channel observations using $\mathcal{P}(u_i|\boldsymbol{u}_{[\![i-1]\!]}, \boldsymbol{y})$ whereas DM bits are decoded probabilistically from the a priori probabilities using $\mathcal{P}(u_i|\boldsymbol{u}_{[\![i-1]\!]})$, identically to the encoder. Decoding the DM bits from $\mathcal{P}(u_i|\boldsymbol{u}_{[\![i-1]\!]})$ requires a common source of randomness that is shared between encoder and decoder.

3.2.1 Achieving Capacity

To show the capacity achieving and diminishing error probability properties of this scheme, the (source) Bhattacharyya parameter $Z(X|Y)$ for some random pair X, Y is required.

Definition 1 (Bhattacharyya parameter [Arı10])**.** Let $X \sim \mathrm{Ber}(p)$ and Y be two random variables with some joint distribution $(X, Y) \sim P_X P_{Y|X}$. Then the Bhattacharyya parameter $Z(X|Y)$ is defined as

$$Z(X|Y) = 2\,\mathbb{E}\left[\sqrt{P_{X|Y}(0|Y)P_{X|Y}(1|Y)}\right]. \tag{3.8}$$

This parameter $Z(X|Y)$ is used for polarization results and corresponds to the conditional entropy as stated by the following proposition.

Proposition 1 ([Arı10, Proposition 2])**.**

$$(Z(X|Y))^2 \leq \mathbb{H}(X|Y) \leq Z(X|Y). \tag{3.9}$$

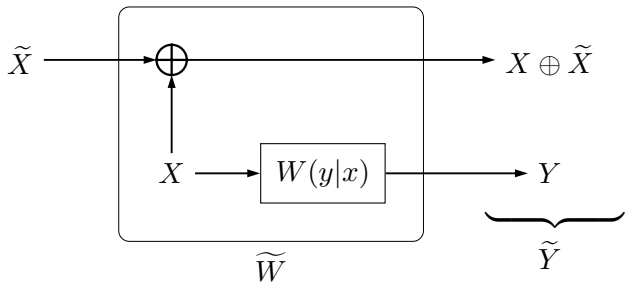


Figure 3.4: Symmetrized channel $\widetilde{W}$.

Most importantly, this means that a conditional entropy $\mathbb{H}(X|Y)$ polarizes together with the respective Bhattacharyya parameter $Z(X|Y)$. Using a channel symmetrization argument by [Kor09], one can show that polar codes do not only polarize the transmission over a symmetric channel but in fact polarize over any pair X, Y of random variables. This is proven by introducing a symmetric channel $\widetilde{W} : \widetilde{X} \to \widetilde{Y}$ as shown in Figure 3.4 for which $\mathbb{I}(\widetilde{W}) = 1 - \mathbb{H}(X|Y)$. Therefore, Arıkan's original polarization theorem is applicable to this symmetrized channel, which captures the important properties of the original asymmetric channel. The respective result is covered in the next proposition.

Proposition 2. *Let $X \sim \mathrm{Ber}(p)$ and Y be two random variables with some joint distribution $(X, Y) \sim P_X P_{Y|X}$. Let $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{F}^{\otimes n}$ be the polar transform of the random vector $\boldsymbol{U}$ and $N = 2^n$. Then for any $\beta < \frac{1}{2}$,*

$$\lim_{N\to\infty} \frac{1}{N} \left| \left\{ i : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y}) \leq 2^{-N^\beta} \right\} \right| = 1 - \mathbb{H}(X|Y), \tag{3.10}$$

$$\lim_{N\to\infty} \frac{1}{N} \left| \left\{ i : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y}) \geq 1 - 2^{-N^\beta} \right\} \right| = \mathbb{H}(X|Y). \tag{3.11}$$

Proof. The proof to this proposition can be found in the proof to Theorem 1 in [HY13, Eqs. (38), (39)]. □

Based on this, we provide a sketch of the capacity proof from [HY13] hereinafter. For this, we also need the following Lemma.

Lemma 1 ([Liu16, Lemma 4.3.2])**.** *Let $(X, Y, S) \sim P_{X,Y,S}$ be three random variables with $\mathcal{X} = \mathbb{F}_2$. Then,*

$$Z(X|Y, S) \leq Z(X|Y). \tag{3.12}$$

Using the polarization for any two variables suggested by Proposition 2, one can consider the bitchannel entropies $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y})$ and $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]})$ (or their respective Bhat-

tacharyya parameters), where the random variable Y is set to a constant for the second expression. This leads to the four index sets denoted by

$$\mathcal{L}_U = \left\{ i : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}) \leq 2^{-N^\beta} \right\}, \tag{3.13}$$

$$\mathcal{H}_U = \left\{ i : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}) \geq 1 - 2^{-N^\beta} \right\}, \tag{3.14}$$

$$\mathcal{L}_{U|Y} = \left\{ i : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y}) \leq 2^{-N^\beta} \right\}, \tag{3.15}$$

$$\mathcal{H}_{U|Y} = \left\{ i : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y}) \geq 1 - 2^{-N^\beta} \right\}. \tag{3.16}$$

By Lemma 1, $\mathcal{L}_U \subseteq \mathcal{L}_{U|Y}$ and thus $\mathcal{L}_U \cap \mathcal{H}_{U|Y} = \emptyset$. Therefore, we can assign $\mathcal{L}_U$ for DM, $\mathcal{H}_{U|Y}$ for FEC redundancy and $\mathcal{H}_U \cap \mathcal{L}_{U|Y}$ for data transmission. The proportion of bitchannels for data transmission thus converges to the MI and $\lim_{N\to\infty} \frac{\left|\mathcal{H}_U \cap \mathcal{L}_{U|Y}\right|}{N} = \mathbb{I}(X;Y)$. By choosing $P_X = \arg\max_{P_X} \mathbb{I}(X;Y)$, the proportion of data transmission bits approaches the channel capacity. The authors of [HY13] show furthermore that such a construction has a decoding error probability of $\mathcal{P}(\mathcal{E}) = \mathcal{O}(2^{-N^{\beta'}})$ with $\beta' < \beta < \frac{1}{2}$ for uniformly chosen messages.

3.2.2 Practical Implementation

We remark that transmission of the all-zero codeword to speed up calculation of the bitchannel reliabilities cannot be assumed in the asymmetric case anymore [WKP05].

While the theoretical proof requires probabilistic encoding as well as two-fold decoding, a practical transmission scheme is also possible by using an unmodified polar decoder and a deterministic encoder [CB15], [MHU18]. The unmodified receiver performs well in practice, because by Equation (2.6) DM bitchannels with low $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]})$ also have low $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y}) \leq \mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]})$. Therefore, those bits can as well be decoded successfully from the channel outputs without the need for a shared source of randomness. The deterministic encoder that finds the most likely data word instead of a random one also works well in practice, as the bitchannels used for DM ideally have $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}) \approx 0$. Thus, even with a probabilistic decision rule, these bits are almost deterministic. To find this deterministic most likely data word, an unmodified SC decoder is used. The frozen bits for this decoder at the transmitter are set to the frozen bits as well as to the data bits of the channel code. This decoder then finds the appropriate DM bits by applying SC decoding. The data word found by the shaping decoder is then encoded with a normal polar encoder, resulting in the most-likely codeword for the given data and frozen bits and the target distribution. Decoding at transmitter and receiver can again be improved using modified decoding rules such as SCL decoding.

When list decoding is employed at the transmitter, then instead of the most likely codeword one can also choose the codeword with minimum energy c.f. [For92]. As all options in the list are valid codewords for the given data bits, each codeword in this list is a valid choice for transmission. For priors P_X that are Gaussian-type distributions, i. e., $P_X(x) \propto \exp(-c\|x\|^2)$, the negative log-likelihood of a codeword is an affine function of the energy, i. e., $-\log(\mathcal{P}(\boldsymbol{x})) = m\|\boldsymbol{x}\|^2 + b$ with $m > 0$. Thus, the most likely codeword is also the minimum-energy codeword for these distributions irrespectively of the code.

It is important to note that the deterministic encoder can only realize the target distribution accurately if all bitchannels used for DM have $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}) \approx 0$ and the remaining bitchannels have $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}) \approx 1$. In practical applications, code construction may result in DM, FEC and data index sets which are slightly different from $\mathcal{L}_U$, $\mathcal{H}_{U|Y}$ and $\mathcal{H}_U \cap \mathcal{L}_{U|Y}$, respectively. Especially when polar code constructions are reused at different SNRs or when rate matching is implemented using the bitchannel selection, it may happen that bitchannels with $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}) \gg 0$ are used for DM. In this case, these bits need to be encoded with a probabilistic decision rule. Otherwise, a deterministic decision rule would result in an empirical entropy of $\widehat{\mathbb{H}}(U_i|\boldsymbol{U}_{[\![i-1]\!]}) \approx 0$. Then, due to $\sum_{i\in[\![N]\!]} \mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}) = \mathbb{H}(\boldsymbol{X})$, the actual empirical entropy of the channel input distribution would be $\widehat{\mathbb{H}}(\boldsymbol{X}) < \mathbb{H}(\boldsymbol{X})$. In this case, the generated distribution will differ from the target distribution. With a different empirical distribution and different entropy, also the energy of a codeword changes.

3.2.3 Discussion

The main advantage of this shaping scheme is two-fold. In comparison to Gallager's scheme, arbitrary target distributions can be achieved efficiently, even if there are no small natural numbers a and b so that $P_X(x) \approx \frac{a}{b}$. On the other hand, this scheme is able to shape bits with an arbitrary distribution whereas existing practical shaping schemes require some symmetry constraints to be fulfilled. The scheme is able to perform DM on data bits as well as parity bits, without an increase in decoding complexity.

Similar to trellis shaping, HY coding is an instance of coset coding [For88]. Coset coding for shaping employs a shaping code so that multiple different codewords of the shaping code may correspond to the same codeword of the channel code. Then, an encoder for the channel code is combined with a decoder for the shaping code, where the shaping decoder selects either a random or the most likely transmission sequence. Where trellis shaping uses convolutional coding and trellis decoders, HY coding depends on the SC decoding rule to find an approximate most likely transmission sequence. By this, the favourable properties of polar codes can also be used for the shaping code. In particular,

HY coding provides flexible rate adaption, provable capacity achievability and provable decoding error convergence [HY13]. As both, the shaping code and the FEC code, are defined via the polar transform and the DM frozen index set contains the FEC frozen index set, the shaping code is trivially a subcode of the FEC code. Both codes can be constructed as a single, joint code [HY13].

3.3 Polar-Coded Modulation

The second important extension of polar codes discussed here is polar-coded modulation and in particular multilevel coding (MLC).

MLC [Ung82], [WFH99] protects each of the m bitlevels for a constellation with 2^m symbols with a separate code. For each bitlevel ℓ, the component code produces a length-N binary codeword $\boldsymbol{x}_{[\![N]\!]}^{\mathrm{B},\ell}$. These codewords are then encoded to a symbol codeword $\boldsymbol{x}_{[\![N]\!]}$ by applying a labelling rule $x_i = f(\boldsymbol{x}_i^{\mathrm{B},[\![m]\!]})$. The overall multilevel code has a rate $R = \sum_{\ell \in [\![m]\!]} R^\ell$, where R^ℓ is the rate of the ℓ-th component code. For decoding, the multistage decoding (MSD) rule is used. MSD iteratively demaps and decodes a bitlevel ℓ conditioned on the bits decoded at the previous bitlevels $\ell' < \ell$. The binary-input channel $W : X^{\mathrm{B},\ell} \to Y$ with side information $\boldsymbol{X}^{\mathrm{B},[\![\ell-1]\!]}$ known at transmitter and receiver is called equivalent channel [WFH99] and has an achievable rate

$$\bar{R}^\ell = \mathbb{I}(X^{\mathrm{B},\ell}; Y | \boldsymbol{X}^{\mathrm{B},[\![\ell-1]\!]}). \tag{3.17}$$

By the chain rule of MI, the achievable rate of the whole multilevel code is equal to

$$\bar{R} = \sum_{\ell \in [\![m]\!]} \bar{R}^\ell = \mathbb{I}(X;Y). \tag{3.18}$$

Choosing $P_X^* = \arg\max_{P_X} \mathbb{I}(X;Y)$ and $P_{X^{\mathrm{B},\ell}|\boldsymbol{X}^{\mathrm{B},[\![\ell-1]\!]}}$ so that $\prod_{\ell \in [\![m]\!]} P_{X^{\mathrm{B},\ell}|\boldsymbol{X}^{\mathrm{B},[\![\ell-1]\!]}} = P_X^*$, the achievable rate for MLC with MSD is the channel capacity.

Seidl et al. [SSSH13] proposed a straightforward way to combine MLC with polar codes. They notice that both the polar transform as well as MLC rely on splitting and combining channels and both decoding rules, SC decoding and MSD, implement the chain rule for MI. This correspondence allows a direct combination of both, where the overall multilevel polar code itself is treated as a polar-like code.

Their transmission scheme is based on MLC, where each of the m bitlevels for a constellation with 2^m symbols is protected by a separate length-N polar component code and decoded iteratively. A block diagram depicting the multilevel polar code can be seen in Figure 3.5. Instead of requiring an explicit rate allocation during code design

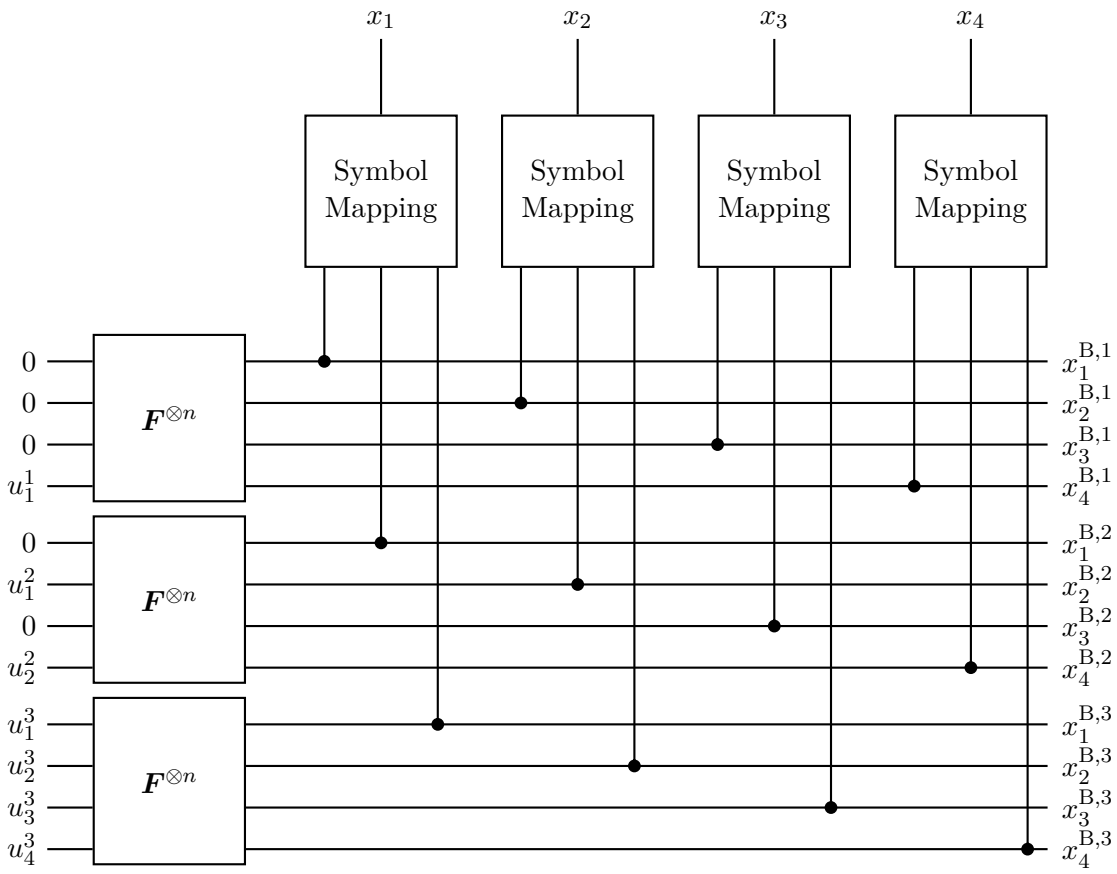


Figure 3.5: Multilevel Polar-Coded Modulation Scheme with $m = 3$ and $N = 4$.

as for standard MLC, considering the overall coding scheme as a single polar-like code translates into an implicit rate allocation. This implicit rate allocation is done during the polar code construction step. There, the reliabilities of all bitchannels from all bitlevels are computed and the most reliable bitchannels are chosen for transmission. The distribution of information bitchannels and frozen bitchannels over the bitlevels results in the implicit and optimal rate allocation. Also, since the overall code can be considered a polar-like code with block length mN, the component codes can decrease in block length by a factor of m without sacrificing code performance relative to a scalar polar code with block length N [Sei15].

Furthermore, it can be shown that multilevel polar-coded modulation is capacity achieving with provably decreasing error probabilities [SSSH13], [Sei15].

The authors of the original work note that for MLC the labels should be chosen by set-partitioning [WFH99] whereas for bit-interleaved coded modulation (BICM) the labels should be chosen by Gray labelling.

The efficient SCL-based decoder can be used for MSD as well, by making the decoding lists of one component code available to the component code of the next bitlevel [PY18]. Also, there exist schemes to approximate bitchannel reliabilities in a computationally more efficient manner using surrogate channels [BPYS17].

4 Symbol Shaping using Polar-Coded Modulation

In this chapter, we propose a coded modulation scheme which is able to provide arbitrary symbol-level probabilistic shaping and which achieves the coded modulation capacity. We first introduce the proposed MLHY coded modulation scheme and discuss its theoretical properties, before we consider aspects of construction. Finally, we show simulated performance results.

The scheme aims at scenarios where the target distribution P_X does not exhibit any exploitable symmetries. Existing practical shaping schemes either require that the distribution is symmetric, i. e., that it is decomposable as $P_X(x) = \frac{1}{2}\,\mathcal{P}(\boldsymbol{x}^{\mathrm{B},[\![m-1]\!]})$ [BSS15], or they have to use very large alphabets in order to obtain arbitrary distributions [Gal68]. Also, MLC-based shaping schemes often only allow independent shaping of the bitlevels [FHW98], which imposes a product distribution constraint onto P_X. In particular, this is the case when each bitlevel is shaped separately without the possibility of adapting the shaping to the previous bitlevels.

To approach capacity with a practical scheme, a transmitter needs to generate channel input sequences that follow a target distribution and that can be decoded with low error probability and low computational effort at the receiver. We use an approach that integrates a source decoder as well as a channel encoder into the transmitter, where the former executes the DM and the latter enables FEC. Following [HY13], we use a single polar code for both. Similar to [BİX20], the encoding structure facilitates MLC to implement

$$P_X(x_i) = \prod_{\ell \in [\![m]\!]} \mathcal{P}(x_i^{\mathrm{B},\ell} | \boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]}). \tag{4.1}$$

4.1 Multilevel Honda-Yamamoto Coded Modulation

We combine the idea of using a polar decoder to find the most likely transmission sequence with the extension of polar codes to MLC.

Figure 4.1 depicts the proposed scheme. To find the most likely transmission sequence in a coded modulation case, the polar decoder at the transmitter is replaced with a

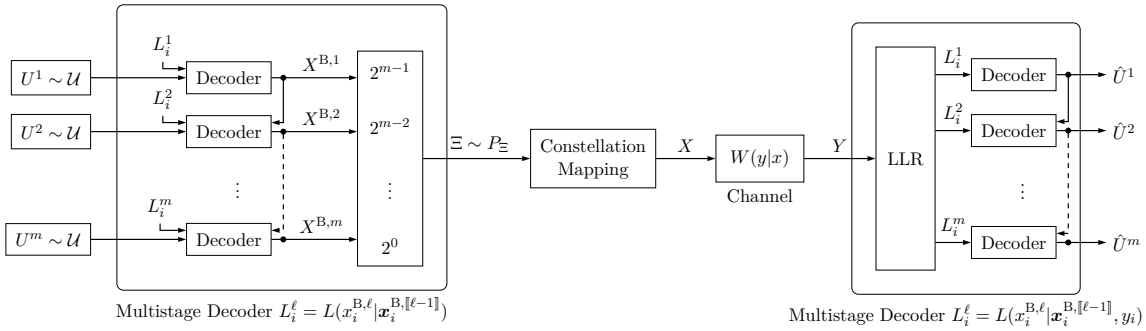


Figure 4.1: Multilevel Honda-Yamamoto Coded Modulation.

multistage decoder with polar decoders for its component codes. This multistage decoder can be the same decoder as used on the receiver side which calculates the likelihoods for each component code based on some input symbols and on the previously decided bits. Instead of channel observations, the constant symbol prior P_X is used for calculation of the likelihoods at the transmitter. The resulting multilevel code is again treated as a single code with regards to construction. We assume the same polarization stage count n for each component code, an alphabet size of $|\mathcal{X}| = M = 2^m$ and a one-to-one mapping $f : \Xi \to \mathcal{X}$ between label ξ and symbol x. Thus, the total block length is equal to N channel uses and the binary block length is equal to $Nm = 2^n m$ bits.

Symbol Mapping Each transmission symbol $x_i \in \mathcal{X}$ with $i \in [\![N]\!]$ is assigned with a label $\xi_i \in \{0, \dots, M-1\}$. As the labelling f is bijective, the distributions P_X and P_Ξ describe the symbol distribution equivalently. Each label ξ_i can be represented by a group of m bits $\boldsymbol{x}_i^{\mathrm{B},[\![m]\!]}$. With slight abuse of notation, we define $P_\Xi(\boldsymbol{x}^{\mathrm{B},[\![m]\!]})$ as the probability $P_\Xi(\xi)$ of the respective label ξ.

Encoding The component HY encoder for bitlevel ℓ encodes a sequence of uniform bits $u_j^\ell \sim \mathcal{U}(\mathbb{F}_2)$ with $j \in [\![N^{\mathrm{I}}(\ell)]\!]$, where $N^{\mathrm{I}}(\ell)$ denotes the number of information bits of the component code ℓ.

Each block of bits $\boldsymbol{x}^{\mathrm{B},\ell}$ for a bitlevel ℓ is a codeword of the component HY code. The HY code is used to produce codewords where the bits are distributed according to $\mathcal{P}(x_i^{\mathrm{B},\ell}|\boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]})$, so that the distribution for each label ξ_i is

$$\widehat{\mathcal{P}}(\xi_i) = \widehat{\mathcal{P}}(\boldsymbol{x}_i^{\mathrm{B},[\![m]\!]}) = \prod_{\ell \in [\![m]\!]} \mathcal{P}(x_i^{\mathrm{B},\ell}|\boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]}) = P_\Xi(\boldsymbol{x}_i^{\mathrm{B},[\![m]\!]}). \tag{4.2}$$

The probabilities $\mathcal{P}(x_i^{\mathrm{B},\ell}|\boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]})$ are calculated by the multistage decoder. For this, the multistage decoder conditions P_Ξ on all previous bitlevels $\ell' < \ell$ and marginalizes P_Ξ over all bitlevels $\ell' > \ell$. We write

$$\mathcal{P}(x_i^{\mathrm{B},\ell}|\boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]}) = \frac{\mathcal{P}(\boldsymbol{x}_i^{\mathrm{B},[\![\ell]\!]})}{\mathcal{P}(\boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]})} = \frac{\sum_{\boldsymbol{x}_i^{\mathrm{B},\ell+1:m}} P_\Xi(\boldsymbol{x}_i^{\mathrm{B},[\![m]\!]})}{\sum_{\boldsymbol{x}_i^{\mathrm{B},\ell:m}} P_\Xi(\boldsymbol{x}_i^{\mathrm{B},[\![m]\!]})}, \tag{4.3}$$

where $\sum_{\boldsymbol{x}_i^{\mathrm{B},\ell:m}} P_\Xi(\boldsymbol{x}_i^{\mathrm{B},[\![m]\!]})$ denotes the sum over all bitlevels $\ell < \ell' < m$,

$$\sum_{\boldsymbol{x}_i^{\mathrm{B},\ell:m}} P_\Xi(\boldsymbol{x}_i^{\mathrm{B},[\![m]\!]}) = \sum_{x_i^{\mathrm{B},\ell}\in\{0,1\}} \cdots \sum_{x_i^{\mathrm{B},m}\in\{0,1\}} P_\Xi(\boldsymbol{x}_i^{\mathrm{B},[\![m]\!]}). \tag{4.4}$$

The likelihoods $\mathcal{P}(x_i^{\mathrm{B},\ell}|\boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]})$ are then supplied to the HY encoder in order to encode the input sequence $\boldsymbol{u}^\ell_{[\![N^{\mathrm{I}}(\ell)]\!]}$ into a HY codeword $\boldsymbol{x}^{\mathrm{B},\ell}$ with the desired distribution.
This process happens successively. First the multistage decoder calculates $\mathcal{P}(x_i^{\mathrm{B},1})$ for $i \in [\![N]\!]$. With this, the component encoder generates a HY codeword $\boldsymbol{x}^{\mathrm{B},1}$. This codeword is then used by the multistage decoder to calculate $\mathcal{P}(x_i^{\mathrm{B},2}|x_i^{\mathrm{B},1})$ and supply those to the next HY encoder. The process continues until each bitlevel generated a codeword $\boldsymbol{x}^{\mathrm{B},\ell}$ with $\ell \in [\![m]\!]$. These HY codewords are then used for symbol mapping.
Depending on the used decision rule for the SC decoding, the encoder either finds the most likely symbol codeword $\boldsymbol{x}$ given $\boldsymbol{u}$ and P_X or generates symbol codewords $\boldsymbol{x}$ for $\boldsymbol{u}$ randomly distributed according to P_X.

Demapping and Decoding For demapping, the multistage decoder at the receiver calculates the probabilities $\mathcal{P}(x_i^{\mathrm{B},\ell}|\boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]}, y_i)$. Again, these condition on already decoded bits and marginalize over future bits so that

$$\mathcal{P}(x_i^{\mathrm{B},\ell}|\boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]}, y_i) \propto \sum_{\boldsymbol{x}_i^{\mathrm{B},\ell+1:m}} \mathcal{P}(\boldsymbol{x}_i^{\mathrm{B},[\![m]\!]}|y_i). \tag{4.5}$$

Example 2. For on-off keying (OOK) with amplitude $f(1) = \Delta$ and distribution $P_X(\Delta) = p$, $P_X(0) = 1-p$ over an AWGN channel $Y = X + N$ with noise $N \sim \mathcal{N}(0, \sigma^2)$, the LLR-based demapper calculates

$$L(x_i^{\mathrm{B}}|y_i) = \ln \frac{\mathcal{P}(x_i^{\mathrm{B}} = 0|y_i)}{\mathcal{P}(x_i^{\mathrm{B}} = 1|y_i)} \tag{4.6}$$

$$= \ln \frac{\mathcal{P}(y_i|x_i^{\mathrm{B}} = 0)P_X(0)}{\mathcal{P}(y_i|x_i^{\mathrm{B}} = 1)P_X(\Delta)} \tag{4.7}$$

$$= \ln \frac{1-p}{p} + \ln \frac{\exp(-\frac{1}{2}\left|\frac{y}{\sigma}\right|^2)}{\exp(-\frac{1}{2}\left|\frac{y-\Delta}{\sigma}\right|^2)} \tag{4.8}$$

$$= \ln \frac{1-p}{p} + \frac{\Delta\left(\Delta - 2y_i\right)}{2\sigma^2}, \tag{4.9}$$

where the logarithms are natural logarithms.

Finally, provided with $\mathcal{P}(x_i^{\mathrm{B},\ell}|\boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]}, y_i)$ for $i \in [\![N]\!]$, the polar decoders calculate $\mathcal{P}(\widehat{u}_i^\ell|\widehat{\boldsymbol{u}}_{[\![i-1]\!]}^\ell, \boldsymbol{x}^{\mathrm{B},[\![\ell-1]\!]}, \boldsymbol{y})$ as well as estimates for the HY codewords $\boldsymbol{x}^{\mathrm{B},\ell}$. As with the encoding process, this happens iteratively.

By changing the computation of the likelihoods, other decoding schemes can be realized. In particular, by replacing the conditioning with marginalization, BICM-based schemes similar to the one proposed by [İBX17], [İBX19] can be implemented as well. In that case, the likelihoods can be computed and decoded in parallel but only product distributions $\mathcal{P}(\Xi = x^{\mathrm{B},1}\cdots x^{\mathrm{B},m}) = \prod_{\ell\in[\![m]\!]} \mathcal{P}(x^{\mathrm{B},\ell})$ can be realized exactly.

The proposed MLHY coded modulation scheme is similar to the independently developed concept of polar lattices proposed in [Liu16], [LYLW18]. There, the authors use a multilevel lattice scheme as proposed by [FTC00] with HY component codes. The main difference between both schemes is that the scheme proposed in this work is more flexible by not requiring lattice coding and can work for any constellation. This means that in order to achieve the capacity of the general, continuous AWGN channel, a sufficiently well-performing constellation has to be chosen. Also, in Section 4.3 we propose a more straightforward construction procedure than the one described in [Liu16].

4.2 Theoretical Results

In this section we show that the proposed MLHY coded modulation scheme is capacity achieving and can be decoded with diminishing decoding error probabilities for increasing block lengths. As the achievable rates of any modulated transmission scheme are constrained by the constellation, we will show achievability with respect to the constellation-constrained channel capacity.

We can combine the insight that bitchannels of multilevel polar codes still polarize with the insight from Proposition 2 that $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y})$ polarizes to $\mathbb{H}(X|Y)$ for any Y. Then, following the construction used by Honda and Yamamoto considering the two entropies $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]})$ and $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y})$ instead of just $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y})$, we expect a capacity-achieving scheme.

We start by proving results for the binary-input equivalent channel for each component code which we will then use to characterize the overall scheme.

Lemma 2. *Let $X \in \mathbb{F}_2$ and S be the input to a binary-input DMC $W : X \to Y$ with side information $\boldsymbol{S}$ known at both the transmitter and the receiver acausally with joint distribution $(X, Y, S) \sim P_{Y|X,S} P_{X|S} P_S$. Then, there exists a HY transmission scheme that for $N \to \infty$ achieves the capacity of W, i. e., $\max_{P_{X|S}} \mathbb{I}(X;Y|S)$ s.t. $P_{X|S}P_S \in \pi \subseteq \Pi(\mathcal{X} \times \mathcal{S})$ under a constraint π.*

Proof. Let $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{F}^{\otimes n}$ be the polar transform of $\boldsymbol{U}$ and $N = 2^n$. Similar to the proof of original HY coding [HY13, Proof of Theorem 1], we define

$$\mathcal{L}_{U|S} = \left\{ i : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S}) \leq 2^{-N^\beta} \right\}, \tag{4.10}$$

$$\mathcal{H}_{U|S} = \left\{ i : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S}) \geq 1 - 2^{-N^\beta} \right\}, \tag{4.11}$$

$$\mathcal{L}_{U|S,Y} = \left\{ i : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S}, \boldsymbol{Y}) \leq 2^{-N^\beta} \right\}, \tag{4.12}$$

$$\mathcal{H}_{U|S,Y} = \left\{ i : Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S}, \boldsymbol{Y}) \geq 1 - 2^{-N^\beta} \right\} \tag{4.13}$$

with some $\beta < \frac{1}{2}$.

By Lemma 1, $\mathcal{L}_{U|S} \subseteq \mathcal{L}_{U|S,Y}$ and thus $\mathcal{L}_{U|S} \cap \mathcal{H}_{U|S,Y} = \emptyset$. We again identify $\mathcal{L}_{U|S}$ as DM bits at the transmitter, $\mathcal{H}_{U|S,Y}$ as FEC bits and $\mathcal{H}_{U|S} \cap \mathcal{L}_{U|S,Y}$ as data transmission bits with large degree of freedom at the encoder and low uncertainty at the decoder. With Proposition 2, we see that

$$\lim_{N\to\infty} \frac{1}{N} \left| \mathcal{L}_{U|S} \right| = 1 - \mathbb{H}(X|S) \tag{4.14}$$

and

$$\lim_{N\to\infty} \frac{1}{N} \left| \mathcal{H}_{U|S,Y} \right| = \mathbb{H}(X|S,Y). \tag{4.15}$$

Furthermore, we note that

$$\lim_{N\to\infty} \frac{1}{N} \left| \mathcal{L}_{U|S}^C \setminus \mathcal{H}_{U|S} \right| = 0 \tag{4.16}$$

and equivalently for $\mathcal{L}_{U|S,Y}$ and $\mathcal{H}_{U|S,Y}$.

We can see that

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \left| \mathcal{L}_{U|S} \cup \mathcal{H}_{U|S,Y} \right| &= \lim_{N\to\infty} \frac{\left|\mathcal{L}_{U|S}\right| + \left|\mathcal{H}_{U|S,Y}\right| - \left|\mathcal{L}_{U|S} \cap \mathcal{H}_{U|S,Y}\right|}{N} \\ &= 1 - \mathbb{H}(X|S) + \mathbb{H}(X|S,Y) \\ &= 1 - \mathbb{I}(X;Y|S) \end{aligned} \tag{4.17}$$

and thus $\lim_{N\to\infty} \frac{1}{N}\left|\mathcal{L}^C_{U|S} \cap \mathcal{H}^C_{U|S,Y}\right| = \mathbb{I}(X;Y|S)$. From Equation (4.16), we can conclude that $\lim_{N\to\infty} \frac{1}{N}\left|\mathcal{H}_{U|S} \cap \mathcal{L}_{U|S,Y}\right| = \mathbb{I}(X;Y|S)$.
By choosing $P_{X|S} = \arg\max_{P_{X|S}} \mathbb{I}(X;Y|S)$ s.t. $P_{X|S} \in \pi$, the proportion of data transmission bitchannels $\frac{1}{N}\left|\mathcal{H}_{U|S} \cap \mathcal{L}_{U|S,Y}\right|$ approaches the channel capacity. □

We can also extend the error exponent results proven for HY coding [HY13, Theorem 3] to this case with side information.

Proposition 3 ([Liu16, Theorem 4.3.9])**.** *Let W be the binary-input DMC with side information as defined for Lemma 2. Then, the respective capacity-achieving HY scheme for W has a decoding error probability of $\mathcal{P}(\mathcal{E}) = \mathcal{O}(2^{-N^{\beta'}})$ for $\beta' < \beta < \frac{1}{2}$ with increasing block length N for a uniformly chosen message.*

We can now proceed to the main theoretical results of this chapter.

Theorem 1. *Let $W : X \to Y$ be a DMC with joint distribution $X, Y \sim P_{Y|X}P_X$ and $|\mathcal{X}| = 2^m$. Then, there exists a MLHY transmission scheme that for $N \to \infty$ achieves the channel capacity of W, i. e., $\max_{P_X} \mathbb{I}(X;Y)$ s.t. $P_X \in \pi \subseteq \Pi(\mathcal{X})$ under an arbitrary constraint π.*

Proof. For an m-bit coded modulation scheme, each of the 2^m symbols $x \in \mathcal{X}$ is labelled with m bits $x^{\mathrm{B},1} \ldots, x^{\mathrm{B},m}$. The equivalent channel under MSD for each bitlevel ℓ has binary input $X^{\mathrm{B},\ell}$, output Y and state $X^{\mathrm{B},1}, \ldots, X^{\mathrm{B},\ell-1}$ and has a MI of $\mathbb{I}(X^{\mathrm{B},\ell}; Y|\boldsymbol{X}^{\mathrm{B},[\![\ell-1]\!]})$. For each bitlevel, by Lemma 2, there exists a HY transmission scheme that achieves the equivalent channel's MI.
For component codes with equal block length, MLC has a code rate exactly equal to the sum of the component code rates [WFH99]. Furthermore, under MSD the rates of the component codes are summed and we have

$$\sum_{\ell\in[\![m]\!]} \mathbb{I}(X^{\mathrm{B},\ell}; Y|\boldsymbol{X}^{\mathrm{B},[\![\ell-1]\!]}) = \mathbb{I}(\boldsymbol{X}^{\mathrm{B},[\![m]\!]}; Y). \tag{4.18}$$

As each component HY code has a rate approaching $\mathbb{I}(X^{\mathrm{B},\ell}; Y|\boldsymbol{X}^{\mathrm{B},[\![\ell-1]\!]})$, the code rate of the MLHY scheme approaches $\mathbb{I}(\boldsymbol{X}^{\mathrm{B},[\![m]\!]}; Y)$, which is identical to $\mathbb{I}(X;Y)$ because the mapping $f : \Xi \to \mathcal{X}$ is bijective.
Finally, by choosing $P_X = \arg\max_{P_X} \mathbb{I}(X;Y)$ s.t. $P_X \in \pi$ and supplying each component encoder at bitlevel ℓ with the prior given by Equation (4.3) calculated by the multistage decoder, the empirical distribution of X approaches P_X and the code rate of the scheme approaches the channel capacity. □

Theorem 2. *MLHY coded modulation as in Theorem 1 has a decoding error probability of* $\mathcal{P}(\mathcal{E}) = \mathcal{O}(2^{-N^{\beta'}})$ *for* $\beta' < \beta < \frac{1}{2}$ *with increasing symbol block length* N *for a uniformly chosen message.*

Proof. To account for the fact that each HY code for the bitlevel ℓ is decoded after the previous bitlevels were decoded, we analyse the decoding error probability $\mathcal{P}(\mathcal{E}^\ell|\neg\mathcal{E}^1,\dots,\neg\mathcal{E}^{\ell-1})$ for each component code, for which we assume that the previous bitlevels decoded correctly similarly to the analogous assumption made for SC decoding. By Proposition 3, each bitlevel transmission scheme has a decoding error probability of $\mathcal{P}(\mathcal{E}^\ell|\neg\mathcal{E}^1,\dots,\neg\mathcal{E}^{\ell-1}) = \mathcal{O}(2^{-N^{\beta'}})$ with $\beta' < \beta < \frac{1}{2}$ for uniformly chosen messages.
Thus, for each bitlevel ℓ there exists a positive constant c_ℓ and a block length N_ℓ so that $\mathcal{P}(\mathcal{E}^\ell|\neg\mathcal{E}^1,\dots,\neg\mathcal{E}^{\ell-1}) \leq c_\ell 2^{-N^{\beta'}}$ for all $N > N_\ell$. By choosing $c = \max_{\ell\in[\![m]\!]} c_\ell$ and $N_0 = \max_{\ell\in[\![m]\!]} N_\ell$, we can bound the error probability for any $\ell \in [\![m]\!]$ by

$$\mathcal{P}(\mathcal{E}^\ell|\neg\mathcal{E}^1,\dots,\neg\mathcal{E}^{\ell-1}) \leq c2^{-N^{\beta'}} \tag{4.19}$$

for all $N > N_0$. The total decoding error probability under MSD for a sufficiently large number of stages n is consequently

$$\begin{aligned}
\mathcal{P}(\mathcal{E}) &= 1 - \prod_{\ell\in[\![m]\!]} \mathcal{P}(\neg\mathcal{E}^\ell|\neg\mathcal{E}^{[\![\ell-1]\!]}) = 1 - \prod_{\ell\in[\![m]\!]} \underbrace{\left(1 - \mathcal{P}(\mathcal{E}^\ell|\neg\mathcal{E}^{[\![\ell-1]\!]})\right)}_{\geq 1-c2^{-N^{\beta'}}} \\
&\leq 1 - (1 - c2^{-N^{\beta'}})^m \\
&= 1 - \sum_{k=0}^{m} \binom{m}{k} (-c2^{-N^{\beta'}})^k \\
&= 1 - \sum_{k=0}^{m} \frac{m!}{k!(m-k)!} (-1)^k c^k 2^{-kN^{\beta'}} \\
&\overset{(a)}{=} mc2^{-N^{\beta'}} - \sum_{k=2}^{m} \underbrace{(-1)^k \frac{m!}{k!(m-k)!}}_{\geq -m!} c^k 2^{-kN^{\beta'}} \\
&\leq mc2^{-N^{\beta'}} + \sum_{k=2}^{m} m! c^k \underbrace{2^{-kN^{\beta'}}}_{\leq 2^{-2N^{\beta'}}} \overset{(b)}{} \\
&\overset{(c)}{\leq} mc2^{-N^{\beta'}} + m!C2^{-2N^{\beta'}} \\
&= 2^{-N^{\beta'}} (mc + m!C \underbrace{2^{-N^{\beta'}}}_{\leq 2}) \\
&\overset{(d)}{=} \mathcal{O}(2^{-N^{\beta'}})
\end{aligned} \tag{4.20}$$

with

(a) from assuming $m \geq 2$ and moving the terms for $k = 0$ and $k = 1$ out of the sum with $\frac{m!}{k!(m-k)!}(-c)^k 2^{-kN^{\beta'}}\Big|_{k=0} = 1$ and $\frac{m!}{k!(m-k)!}(-c)^k 2^{-kN^{\beta'}}\Big|_{k=1} = -mc2^{-N^{\beta'}}$,

(b) since $k \geq 2$ and $N^{\beta'} > 0$,

(c) by defining the constant $C = \sum_{k=2}^{m} c^k$ and

(d) because $mc + 2m!C$ does not depend on N.

We further remark that with the assumption of $m \geq 2$ for (a) we do not loose generality. For $m = 1$, this scheme is equal to binary HY coding and all results from [HY13] are applicable without modification. □

Similarly to an extension of HY codes to the non-binary case [ŞTA09], [HY13], MLHY coding achieves the channel capacity for arbitrary DMCs. In case the size of the channels input alphabet is not a power of two, π can be chosen so that some symbols have probability zero. While construction as well as decoding of polar codes for non-binary inputs impose additional complexity, c.f. [YS18] and references therein, construction, encoding and decoding of MLHY codes can directly facilitate existing implementations for binary MLC and polar coding.

We remark that the theorems can be easily extended from the DMC case to discrete-input, continuous-output memoryless channels. This construction follows along the lines of Shannon's extension to continuous channels [Sha48, Part IV, Appendix 7] by dividing $\mathcal{Y}$ into suitable subsets $\mathcal{Y}_k \subset \mathcal{Y}$, $k \in [\![K]\!]$ and letting $K \to \infty$. From this extension, it then also follows that MLHY coding achieves the constellation-constrained capacity of continuous channels. Furthermore, one can easily extend the proofs to channels with state known ahead of time at both the receiver and the transmitter so that the achievable rate is given by $\mathbb{I}(X;Y|S)$.

Comparing the decoding error probability of the MLHY transmission scheme with the decoding error probability of a classical HY transmission scheme supports the claim that a multilevel polar code can be treated as a polar-like code. Also in the case of an asymmetric channel, the multilevel construction does not worsen the convergence properties of the polar codes. We remark, though, that $\mathcal{P}(\mathcal{E})$ also increases with the number of bitlevels m, as each bitlevel introduces a separate chance to decoding failure. This is not automatically counteracted by the increased binary block length mN of the multilevel code, as the order of $\mathcal{P}(\mathcal{E})$ is given in terms of the symbol block length N instead.

As with the scalar HY coding, the proof depends on probabilistic SC encoding and the two-fold SC decoding. For practical use, those can be replaced by a deterministic encoder

and a normal polar decoder. Also, existing SCL-based decoding rules for multilevel polar codes can be used at the receiver as well as the transmitter to improve finite block length performance.

4.3 Selection of Bitchannels

4.3.1 Conceptual Considerations

It is important to consider the proposed MLHY code as a single code. Applying the original MLC formalism to HY coding naively results in conceptual problems. Whilst for a symmetric channel the equivalent channels for each bitlevel are independent of each other, this is not the case for asymmetric channels. Special care has to be taken with the HY component codes.

When using a multilevel communication scheme over an asymmetric channel, the channels of each bitlevel depend on each other. In that case, each equivalent channel ℓ depends on the codewords transmitted over the previous bitlevels $k < \ell$. As the polar code and thereby also the HY code is constructed for an individual channel, this poses the question on how to handle these state-dependent equivalent channels. This is a difference to the multilevel polar-coded modulation proposed for symmetric channels, where a component code always encodes for the same equivalent channel.

Conceptually, a similar problem also already exists for standard HY coding. Consider a HY code with n stages over an asymmetric binary-input DMC. Adding a polarization stage effectively couples two subsequent codewords and flips one of the codewords dependent on the other one, so that two length-2^n polar codes are chained together. Due to the asymmetric channel, this bit-flip also changes the channel seen by one of those two stage-n polar codes. Instead of constructing multiple distinct stage-n polar codes, Honda and Yamamoto only consider the overall stage-$(n+1)$ polar code. As the entropies $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]})$ and $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y})$ are expectation values and due to the law of total expectation, i. e., $\mathbb{E}[X] = \mathbb{E}_Y\big[\mathbb{E}[X|Y]\big]$, this effectively averages over the respective channels for the stage-n polar code experiencing the bit-flips.

We can apply the same intuition to a MLC-based HY coded modulation scheme. Using the insights from [SSSH13] that a multilevel polar code can be considered as one single polar-like code, we also consider the MLHY code as a single code. By replacing the SC decoder with a multistage decoder and by replacing the encoder at the transmitter with a decoder, we arrive at the desired scheme. Instead of constructing component codes conditional to the previous bits, the entropies $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]})$ and $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y})$ for all bitchannels i and all bitlevels ℓ are computed jointly. This effectively makes the

component code consider an averaged equivalent channel. The additional conditioning on $\boldsymbol{U}^{[\![\ell-1]\!]}$ is adequate. Even though the individual component SC decoder does not consider bits $\boldsymbol{u}^{[\![\ell-1]\!]}$ from the previous bitlevels, the likelihoods provided to the component decoder by the multistage decoder do depend on these previous bits.

4.3.2 Code Construction

To select the bitchannels for DM, FEC and data transmission, we calculate $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y})$ as well as $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]})$. This can be done using a MC simulation that calculates the probabilities $\mathcal{P}(U_i|\boldsymbol{U}_{[\![i-1]\!]})$ for the complete MLHY code in every iteration.

As the reliabilities for the full MLHY code are computed in one MC simulation, also the implicit rate allocation can be facilitated. For this, we choose the bitchannels with the highest $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y})$ from all bitlevels for FEC and the bitchannels with the lowest $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]})$ from all bitlevels for DM. Heuristic rate allocation schemes as proposed in [WFH99], [DPN21] are not required. Because of Equation (2.6), $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y}) \leq \mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]})$ and the bitchannel sets for FEC and DM are disjoint.

The expectations $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y})$ and $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]})$ are with respect to U as well as Y. Therefore, the MC simulation needs to sample U so that $X \sim P_X$.

To construct a MLHY code with SE of r in bpcu, we propose to first choose the rN bitchannels for transmission that have the largest $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}) - \mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y}) = \mathbb{I}(U_i^\ell; \boldsymbol{Y}|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]})$. The remaining bitchannels are bitchannels with either low $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]})$ or high $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y})$. The former ones are chosen for DM and the latter ones for FEC.

In case off-the-shelf polar codes with known reliability orderings are available, these can be used for HY coding and MLHY coding without modification as well [İBX17]. For this, one can choose the least reliable bitchannels for FEC and the most reliable bitchannels for DM. Although this is not necessarily optimal, bitchannels with the lowest receiver uncertainty $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y})$ are often also amongst the bitchannels with the lowest transmitter freedom $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]})$.

Compared to the construction described by [WSSY19], where first the DM bits and then the remaining bits are chosen, the proposed method does not require any simulations to determine the number of DM bits. Compared to the construction described by [İBX17], where the HY code is constructed using only bitchannel reliabilities, i. e., $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y})$, the proposed method does not rely in the assumption that bitchannels with lowest $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y})$ also have low $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]})$. The main differ-

ence here in comparison to the construction detailed in [Liu16] is the used channel. Liu proposes to construct each bitlevel separately, whereas we propose to optimize the entire scheme using the coded modulation channel model directly. Their channel models used to construct each bitlevel are the symmetrized channels Honda and Yamamoto used to prove the polarization theorems. This results in a more complex construction procedure, without providing any apparent benefits.

4.4 Simulation Results

In this chapter we present exemplary performance curves for the proposed MLHY scheme. All bitchannel entropies as well as frame error rate (FER) performances were obtained using MC methods.

4.4.1 Selection of Bitchannels

Figure 4.2 shows the bitchannel entropies $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]})$ for the two symbol distributions shown in Figure 4.3. The symbols are labelled by set-partitioning [WFH99]. The $N = 128$ bitchannels of each of the $m = 3$ bitlevels are shown concatenated together in decoding order. The symbol entropy in X for the case optimized for an SNR of 16.5 dB is $\mathbb{H}(X) \approx 2.5$ bit whereas for the case optimized for an SNR of 22.5 dB it is $\mathbb{H}(X) \approx 2.92$ bit. For the higher-SNR case, the distribution is more uniform and thus less strong DM is required. Therefore, also fewer bitchannels have $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}) \approx 0$ and more bitchannels can be used for FEC and data. As with polar coding for FEC, the HY construction optimizes the code to the specific channel and target distribution. We can see this effect in Figure 4.2 and Figure 4.3. The steeper the target distribution and thus the smaller its entropy, the higher the number of bitchannels with $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}) \approx 0$ and thus the lower the resulting DM rate.
For both cases in Figure 4.2, the DM bits in each bitlevel lie more towards the end, i. e., are bits that are decided after most of the other bits were decided. Furthermore, there are more DM bits allocated to the bitlevel that is decoded last. This can be understood as a generalization of sign-bit shaping introduced by [For92], as becomes more evident from Figure 4.4 and Figure 4.5. These show the same bitchannel allocations, but for ASK constellations. With sign-bit shaping, the last bit $x^{\mathrm{B},m}$ of a m-bit symbol is used for DM. For this, $\boldsymbol{x}^{\mathrm{B},m}$ is decided after the previous bitlevels. When the probabilities of the two symbols which differ only in their last bit add up to $\frac{2}{M}$, then sign-bit shaping can realize the target distribution exactly. As an intuition behind this, the last bit decides between two possible eventual symbols. Therefore, by deciding the last bit conditional to the previous bits and by using proper labelling, the last bit can determine if a symbol

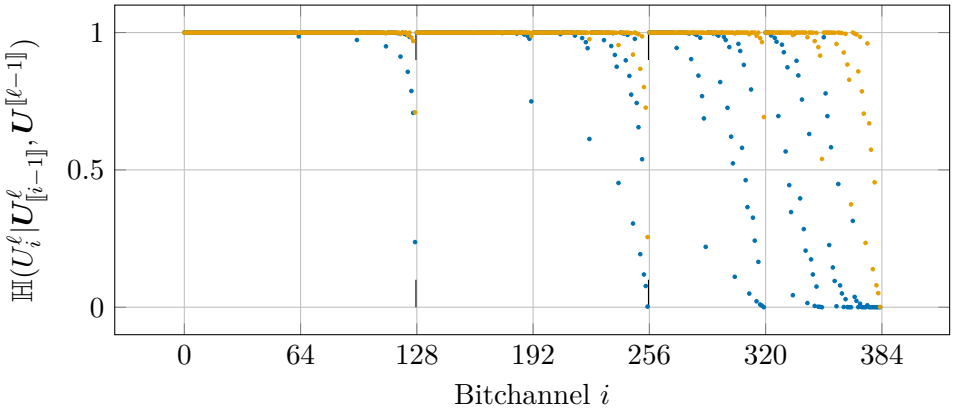


Figure 4.2: Bitchannel entropies for PAM with $M = 8$ and $N = 128$ for an SNR of 16.5 dB (●) and an SNR of 22.5 dB (●).

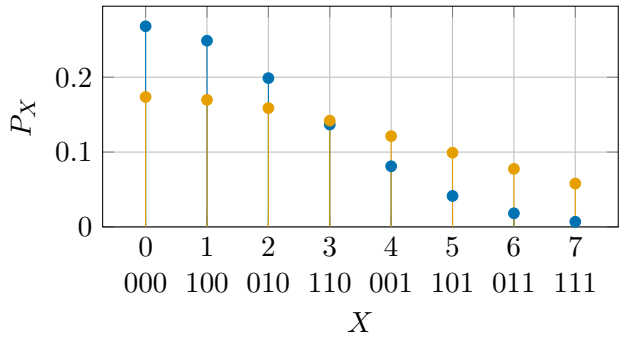


Figure 4.3: Optimal constellations for PAM with $M = 8$ for an SNR of 16.5 dB (●) and an SNR of 22.5 dB (●).

has large or small amplitude and then decide for the smaller amplitude more often. This concept is observable in one bitlevel as well as over the whole block. Within one bitlevel, the later bits are mainly responsible for DM. Over the whole block, the last bitlevel will be used for DM to a large extent whereas the first bitlevel requires almost no DM. In fact, depending on the target distribution and the used labelling, MLHY can be an instance of sign bit shaping when all DM bits are allocated to the final bitlevel only. This is for example the case for the MLHY coding with the 8-ASK constellation for an SNR of 14.6 dB as shown above.

As with the polarization of $X|Y$, the polarization of X introduces weakly polarized bitchannels in the finite-length case. This means that there are bitchannels with either $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]})$ or $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y})$ neither close to 0 nor close to 1. As the weakly polarized bitchannels for the two entropies are usually different ones, a

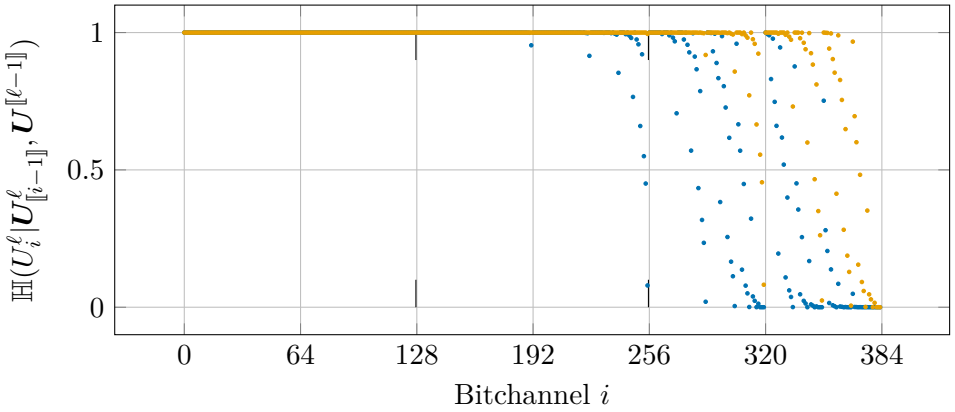


Figure 4.4: Bitchannel entropies for ASK with $M = 8$ and $N = 128$ for an SNR of 8.6 dB (●) and an SNR of 14.6 dB (●).

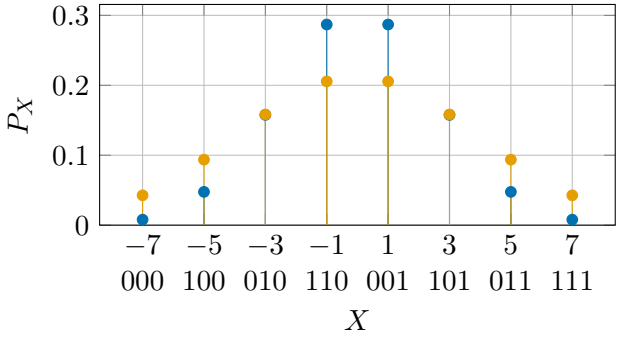


Figure 4.5: Optimal constellations for ASK with $M = 8$ for an SNR of 8.6 dB (●) and an SNR of 14.6 dB (●).

HY construction increases the number of weakly polarized bitchannels compared to a FEC-only polar construction.

4.4.2 Finite Block Length Distribution Matching

One commonly used metric to evaluate the finite length performance of a DM scheme is the rate loss [BSS15, Sec. V-B], [WSSY19], [BİX20]

$$\Delta_{\mathrm{R}} = \mathbb{H}(\widehat{P}_X) - \frac{1}{N}\,\mathbb{H}(\widehat{P}_{\boldsymbol{X}}). \tag{4.21}$$

Assuming a MLHY based scheme which performs only DM, i. e., one that has no frozen bits, the rate loss becomes

$$\Delta_{\mathrm{R}} = \mathbb{H}(\widehat{P}_X) - \frac{|\mathcal{U}|}{N}, \tag{4.22}$$

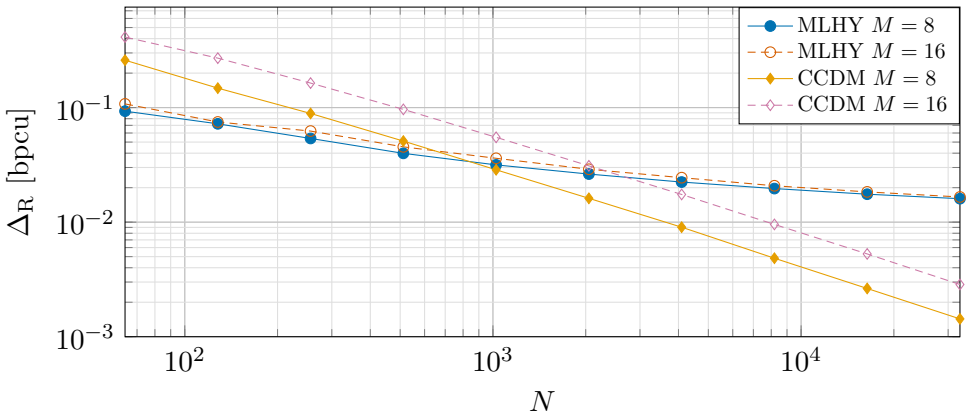


Figure 4.6: Rate loss for MLHY with $L = 32$ and CCDM with and $M = 8$ and $M = 16$.

where $\mathcal{U}$ denotes the bitchannels that are filled with uniform data. Typically, $\mathcal{U}$ consists of the bitchannels with $\mathbb{H}(U_i^\ell | \boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}) \approx 1$.

Equation (4.22) provides an additional argument to deterministic encoding over probabilistic encoding. For probabilistic encoding, the bitchannels $\mathcal{U}^C$ used for DM may contain some randomness whereas for deterministic encoding they do not. This additional randomness increases $\mathbb{H}(\widehat{P}_X)$ when the respective $\mathbb{H}(U_i^\ell | \boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}) > 0$. As the number of data bitchannels $|\mathcal{U}|$ stays the same, this increases the rate loss. In many cases, optimizing the rate loss is preferable [BİX20] which supports the choice towards a fully deterministic encoder.

Figure 4.6 shows the rate loss for MLHY DM with fully deterministic SCL encoding as well as CCDM [SB15] for different alphabet sizes. The target distribution with alphabet size $M = 8$ is the distribution for an SNR of 16.5 dB from Figure 4.3. Figure 4.7 depicts the target distribution with alphabet size $M = 16$ which is optimized for an SNR of 22.5 dB. The MLHY code is constructed by using all bitchannels with $\mathbb{H}(U_i^\ell | \boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}) \leq 0.95$ for DM and its generated empirical distribution $\widehat{P}_X$ is determined in a MC simulation. Distribution and rate for the CCDM are determined by [BG16, Algorithm 2] and [BSS15, Eq. (37)], respectively. The simulations only consider DM without any FEC.

Consistently with previous results [SG17], [WSSY19], CCDM shows lower rate losses for large block lengths. For short block lengths up to 1024, MLHY dominates CCDM in terms of rate loss. Furthermore, the increase of the number of symbols from $M = 8$ to $M = 16$ worsens the rate loss of CCDM by a factor of 2. MLHY shows only negligible penalties in rate loss from the increase in alphabet size. While MLHY may perform DM with a higher rate loss at larger block lengths, it provides superior performance for short block lengths and offers more flexibility for the design of joint DM and FEC schemes in general.

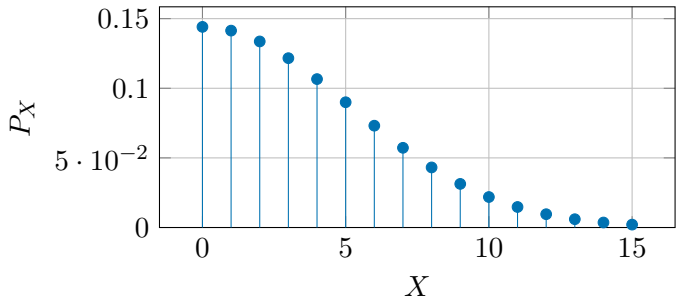


Figure 4.7: 16-PAM target distribution.

4.4.3 Coding for Channels with Symmetric Input Distributions

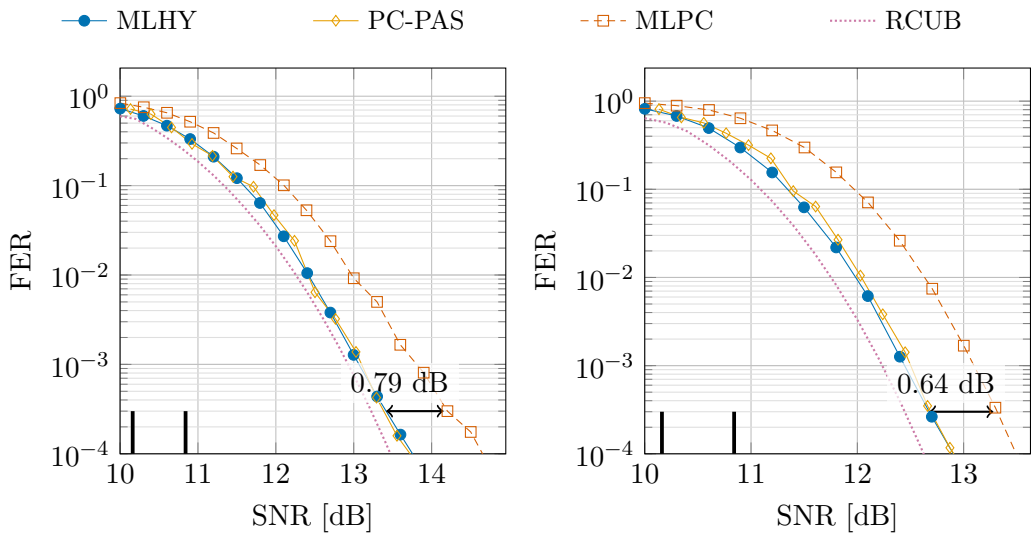
In this chapter we provide simulation results for transmission with ASK over the AWGN channel. For such a scenario, efficient methods for joint DM and FEC, e. g., the already mentioned PAS exist.

We compare MLHY with polar coded probabilistic amplitude shaping (PC-PAS) proposed by [PYB+17]. PC-PAS uses the PAS architecture with a systematic multilevel polar code as FEC and CCDM [SB15] as DM.

The MLHY simulations hereinafter employ SCL decoding at the transmitter as well as at the receiver with $L = 32$. The fully deterministic receiver selects the maximum-likely codeword and, as the channel input distribution is of sampled Gaussian type, thus also the minimum-energy codeword. At the receiver, list decoding is aided by an outer CRC code. Code design is performed for a chosen design signal-to-noise ratio (dSNR) as well as a chosen SE by using the method described in Section 4.3. The dSNR also determines the target distribution for the DM. Set-partitioning labelling [WFH99] is used to label the channel input symbols.

Figure 4.8 shows performance curves for 8-ASK MLHY coding as well as PC-PAS. We also compare our results to the random coding union bound (RCUB) [PPV10], which we compute for the same channel input distribution as the distribution realized by the MLHY encoder. The codes and bounds are designed for a SE of 1.75 bpcu. The two bold black lines at 10.162 dB and 10.841 dB denote the constellation-constrained channel capacities at this SE for shaped and uniform transmission, respectively. The theoretical shaping gain is 0.679 dB.

Contrary to the MLHY code, the empirical distributions of the codewords of CCDM are all of the same type. This allows an additional list pruning step for PC-PAS, where the decoder candidates are checked to be of the correct type [PYB+17]. This type check acts as an additional outer code, so that the length of the outer CRC for PC-PAS can be reduced. A similar check might be possible with MLHY as well, although there the



(a) $N = 64$
shaped dSNR = 16.5 dB
uniform dSNR = 18.0 dB

(b) $N = 128$
shaped dSNR = 14.6 dB
uniform dSNR = 15.1 dB

Figure 4.8: Performance of MLHY coding over the AWGN channel compared to PC-PAS [PYB+17], uniform MLPC, and the RCUB, with an 8-ASK constellation, $L = 32$, and at a SE = 1.75 bpcu. The MLHY codes use an outer CRC-8, PC-PAS a CRC-4 together with a type check.

codewords are not necessarily of any type. Also, as the PC-PAS codes are constructed using [BPYS17], they achieve optimal performance for different dSNRs than the MLHY codes presented.

We observe comparable slopes in the shaped MLHY and uniform multilevel polar code (MLPC) case, resulting in an almost constant shaping gain in the waterfall region. For both block lengths $N = 64$ and $N = 128$ the codes perform close to the theoretical shaping gain of 0.679 dB as well as close to the RCUB. Without any dedicated code optimization beyond a coarse grid search over the dSNR, the MLHY scheme performs at least on par with PC-PAS. This is in accordance to the rate loss discussion in the previous chapter. Whilst the MLHY scheme has lower rate loss than CCDM and is therefore expected to perform better at identical SEs, CCDM facilitates its rate loss at the decoder to prune the list of candidate codewords and can thereby recover some performance. We remark that MLHY performance may be further optimized by adjusting the dSNR, using separate dSNR for DM and FEC construction, optimizing the bitchannel selection process and optimizing the CRC code.

Figure 4.9 compares MLHY coding with PC-PAS for a 16-ASK transmission with SE of 3 bpcu and a block length of $N = 1024$. Again, MLHY performs slightly better than

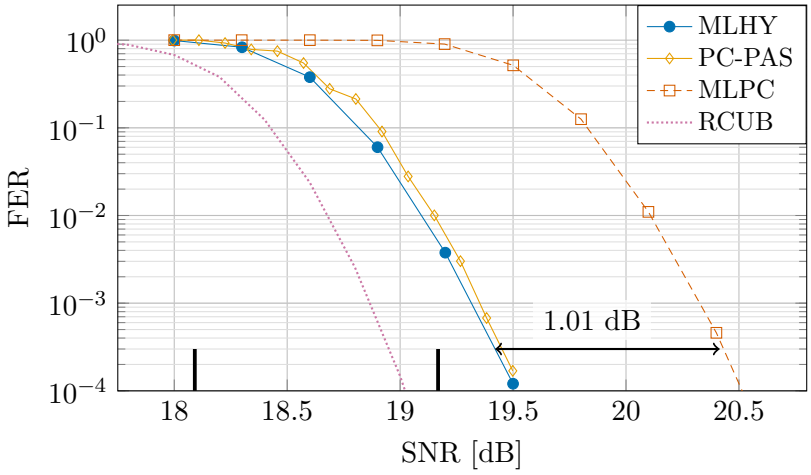
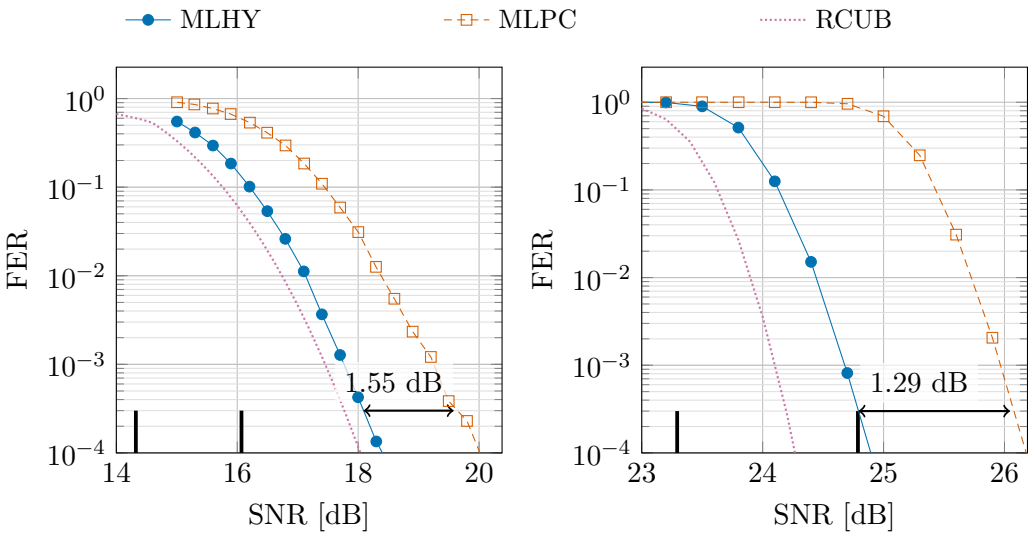


Figure 4.9: Performance of MLHY coding (dSNR = 19.8 dB, CRC-16) over the AWGN channel compared to PC-PAS (CRC-8 with type check), uniform MLPC (dSNR = 20.7 dB, CRC-16) and the RCUB, with a 16-ASK constellation, $N = 1024$, $L = 32$, and at a SE = 3 bpcu.

PC-PAS. Both achieve the theoretical shaping gain of 1.077 dB up to less than 0.1 dB over large parts of the waterfall region.

4.4.4 Coding for Channels with Asymmetric Input Distributions

We also present results for PAM transmission over the IM channel. For this scenario, PAS is not applicable as there are no symmetries in the optimal P_X. Figure 4.10 depicts performance curves for shaped MLHY, uniform MLPC as well as the achievable RCUB. Codes and RCUB are designed identically as in the previous section. Again, the shaping gains of MLHY coding almost attains the ideal ones of 1.746 dB at a SE of 1.75 bpcu and 1.495 dB at a SE of 3 bpcu. For both cases, 8-PAM at 1.75 bpcu at $N = 64$ and 16-PAM at 3 bpcu at $N = 1024$, the shaping gain lies approximately 0.2 dB below the theoretical gain for $N \to \infty$. While MLHY coding for 8-PAM and short block length follows the RCUB closely, for 16-PAM and longer block length there is still some gap. The slopes of MLHY coding and the RCUB are comparable.
We also observe a decreased slope for the IM channel compared to the results for the AWGN channel. This can also be observed from the RCUB for both block lengths.



(a) 8-PAM, $N = 64$, CRC-8, 1.75 bpcu
shaped dSNR = 16.5 dB
uniform dSNR = 18.5 dB

(b) 16-PAM, $N = 1024$, CRC-16, 3 bpcu
shaped dSNR = 24.0 dB
uniform dSNR = 25.5 dB

Figure 4.10: Performance of MLHY codes over the IM channel compared to uniform MLPC and the RCUB. Both polar codes have $L = 32$.

5 Coding with Side Information at the Transmitter

5.1 The Gelfand-Pinsker Channel

Consider a channel $W(y|x,s)$ with input X, output Y and state S. The state $\boldsymbol{S}_{[\![N]\!]}$ may be unknown to the receiver, but is known to the transmitter ahead of time.
Gelfand and Pinsker [GP80] showed that the capacity of such a channel can be achieved by introducing an auxiliary random variable Ξ, so that $x_i = f(\xi_i, s_i)$ with some function $f : \Xi \times \mathcal{S} \to \mathcal{X}$. Using a random coding-based argument, they showed that the achievable rate for this channel is equal to

$$R_{\mathrm{GP}} = \mathbb{I}(\Xi; Y) - \mathbb{I}(\Xi; S) = \mathbb{H}(\Xi|S) - \mathbb{H}(\Xi|Y). \tag{5.1}$$

From this, the channel capacity is

$$C_{\mathrm{GP}} = \max_{P_{\Xi|S}, f} R_{\mathrm{GP}}. \tag{5.2}$$

A block diagram of the channel model and the coding scheme used for their proof can be seen in Figure 5.1. By [EK11, Theorem 7.3, Appendix C], one can upper-bound the size of the alphabet of the auxiliary random variable Ξ required to achieve capacity by

$$|\Xi| \leq \min\{|\mathcal{X}| \cdot |\mathcal{S}|, |\mathcal{Y}| + |\mathcal{S}| - 1\}. \tag{5.3}$$

The expression for the achievable rate in Equation (5.1) can be interpreted as follows. Traditionally, information is encoded directly into the transmission sequence $\boldsymbol{x}$. By intro-

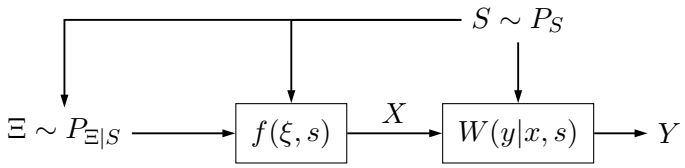


Figure 5.1: Gelfand-Pinsker channel.

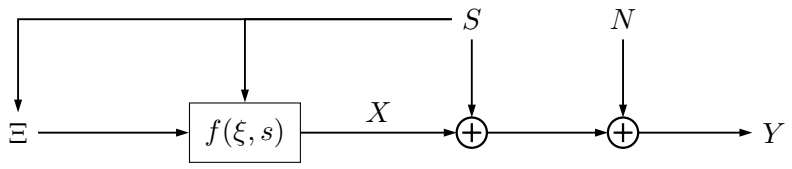


Figure 5.2: Dirty paper channel.

ducing the mapping f, the label Ξ is separated from the symbol X, introducing additional degrees of freedom for the coding scheme. With this additional freedom, interference cancellation can be decoupled from information encoding. When the encoding is constructed accordingly, there is no need to decode the pairs (X, S) at the receiver. Instead, the information sequence $\boldsymbol{\xi}$ is decoded directly. The state-dependent labelling rule effectively abstracts away the channel state.

We can compare Equation (5.1) to the well known capacity equation for channels without state, $C = \max_{P_X} \mathbb{H}(X) - \mathbb{H}(X|Y)$. There, $\mathbb{H}(X)$ quantifies the degrees of freedom per symbol which the transmitter can use to encode information into and $\mathbb{H}(X|Y)$ quantifies the remaining uncertainty that the receiver has to guess and which thus cannot carry information.

Gelfand-Pinsker channels pose an additional problem to practical communication system design. Not only is the capacity argument based on random codes infeasible for actual implementation, but the optimization problem governing the capacity is non-convex. Therefore, in addition to the search for good channel codes, also the search for good input distributions $P_{\Xi|S}$, good constellations $\mathcal{X}(s)$ and good labelling rules $f : \Xi \times \mathcal{S} \to \mathcal{X}$ is non-trivial.

5.2 The Dirty Paper Channel

One important instance of the Gelfand-Pinsker channel is the dirty paper channel. The dirty paper channel is a channel with additive interference and AWGN as shown in Figure 5.2 and can be modelled as $Y = X + S + N$ with $N \sim \mathcal{N}(0, \sigma_N^2)$, $S \sim P_S$, $X = f(\Xi, S)$ and $\Xi \sim P_{\Xi|S}$.

For a Gaussian scenario with $X \sim \mathcal{N}(0, \sigma_X^2)$ and $S \sim \mathcal{N}(0, \sigma_S^2)$ and therefore $Y \sim \mathcal{N}(0, \sigma_Y^2)$ with σ_Y^2 depending on σ_X^2, σ_S^2 and σ_N^2, Costa [Cos83] was able to show that the channel capacity is equal to the AWGN capacity without interference.

For this, Ξ is chosen as $\Xi = X + \alpha S$ with a scalar α and X independent of S. For $\alpha = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2}$, the achievable rate $R_{\mathrm{GP}} = \mathbb{H}(\Xi|S) - \mathbb{H}(\Xi|Y)$ is equal to the AWGN capacity $R_{\mathrm{AWGN}} = \frac{1}{2}\log(1 + \frac{\sigma_X^2}{\sigma_N^2})$ that assumes no interference.

Conceptually, this can be interpreted as leaving the distribution of X constant, while shifting the information-carrying label Ξ in the direction of the interference symbol.
A practical implementation of dirty paper coding poses additional problems. As this scheme assumes Gaussian signalling, the design of practically implementable schemes is again non-trivial. Apart from the difficulty of finding an optimal distribution $P_{\Xi|S}$ and an optimal symbol mapping f, implementing the desired state-dependent shaping is non-trivial as well. A search for such schemes is very well motivated, though, by the astonishing result of $R_{\mathrm{GP}} = R_{\mathrm{AWGN}}$ which suggests the possibility of communication over channels with known interference without any sacrifice in rate. A more extensive review of practical approaches to dirty paper coding can be found in [Len18] and references therein. We remark that most practical dirty paper coding schemes are based on the lattice strategies first proposed by [ZSE02], [EtB05], [ESZ05]. As those hardly compare to the scheme proposed in this work and require the framework of lattice codes [FTC00], we refrain from in-depth explanations.

5.3 Finite Constellation Dirty Paper Coding

One approach to design a practical dirty paper coding scheme is to assume finite constellations. For this, the channel input constellation $\mathcal{X}$ as well as the interference alphabet $\mathcal{S}$ are assumed to be discrete and finite. Assuming both alphabets to be finite can then be exploited for the design of a dirty paper signalling scheme [Len18]. The motivation to consider finite alphabet interference comes from one of the main applications of dirty paper coding, coding for so-called broadcast channels. In this scenario, a transmitter wants to send separate messages to multiple independent receivers over the same medium. There, the transmission to the first user can be considered as known interference for the transmission to the second user. Assuming digital modulation, this interference has a finite alphabet. For infinite alphabet interference, e. g., Gaussian interference, the interference can be quantized if methods for finite constellation interference should be applied.
The state-dependence can pose significant problems to schemes based on many-to-one mappings, as this would require a mapping for each candidate distribution. Given that such schemes often already suffer from too large alphabets, this becomes infeasible easily. Also, schemes based on DM such as PAS and its extensions [BSS15], [BLCS19] need additional complexity to provide a distribution matcher which is able to match symbols x_i according to arbitrary distributions P_X without any exploitable symmetries.
Furthermore, a possible transmission scheme also needs to be able to shape symbols according to asymmetric distributions [Len18]. Figure 5.3 shows an exemplary constellation for dirty paper coding which requires asymmetric shaping. This constellation is optimized

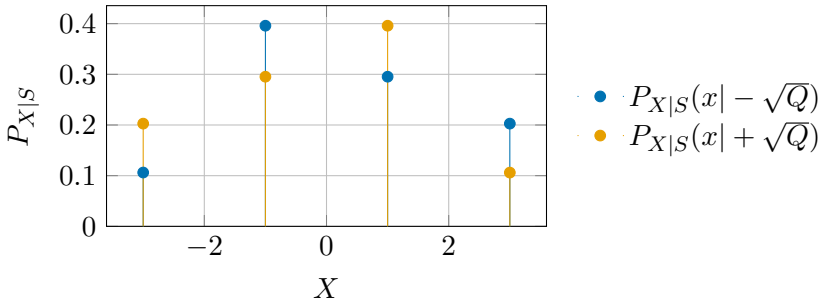


Figure 5.3: Exemplary $P_{X|S}$ dirty paper constellation with $1/Q = \text{SIR} = 5\,\text{dB}$.

for a 4-ASK channel input constellation with a uniform binary phase-shift keying (BPSK) interference, an SNR of 3 dB and a signal-to-interference ratio (SIR) of 5 dB. Depending on if the interference symbol s_i is positive or negative, the channel input symbol x_i should be distributed either according to the blue or the yellow distribution, respectively.

6 State-Dependent Shaping using Polar-Coded Modulation

This chapter discusses HY coding for channels with state.
We assume a Gelfand-Pinsker channel $W(y|x,s)$ with input X and output Y, where the state $\boldsymbol{S}_{[\![N]\!]}$ is known at the transmitter ahead of time. To achieve capacity for such a channel, separation of the label ξ_i and the transmission symbol $x_i = f(\xi_i, s_i)$ as well as a state-dependent distribution $\Xi \sim P_{\Xi|S}$ is necessary.
Before we adapt the scheme proposed in Chapter 4 to state-dependent shaping, we begin by discussing existing literature regarding state-dependent shaping for binary HY coding. After this, we proceed with demonstrating a simple extension of the proposed MLHY coding scheme to state-dependent shaping. We also discuss aspects of construction and present numerical results of the state-dependent shaping scheme.

6.1 State-Dependent Honda-Yamamoto Coding

Extending HY coding to state-dependent shaping in a theoretically sound way is not as straightforward as the extension to MLC and remains an open problem.
For HY coding, the decoder at the transmitter is initialized with the symbol prior P_X. In our extension to MLHY, we initialized the decoder for each bitlevel with the conditioned and marginalized probabilities $\mathcal{P}(x_i^{\mathrm{B},\ell}|\boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]})$. This extension introduced a state dependency to the provided probabilities. Although this changes the initial probabilities P_X – which are constant for binary HY coding – to varying initial probabilities and conditions the considered bitchannel entropies, we showed that this construction maintains polarization convergence.
The naive way to extend HY coding to a state-dependent shaping scenario suggests to

1. apply the (multilevel) polar transform to Ξ instead of X,

2. consider $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S})$ and $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y})$ for construction instead of $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]})$ and $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y})$ and

3. initialize the decoder at the transmitter using $P_{X|S}(x_i|s_i)$, $i \in [\![N]\!]$, instead of just $P_X(x_i)$.

Compared to the achievable rate without interference $\mathbb{I}(X;Y)$, there is a reduction by $\mathbb{I}(\Xi;S) = \mathbb{H}(\Xi) - \mathbb{H}(\Xi|S)$ in rate for Gelfand-Pinsker coding required to perform interference cancellation. This is also reflected in the construction of the DM code, where $\mathbb{H}(\Xi|S)$ instead of $\mathbb{H}(X)$ is polarized in order to realize conditional target distributions $P_{\Xi|S}$. Due to $\mathbb{H}(\Xi|S) \leq \mathbb{H}(\Xi) \leq \log|\Xi|$, this imposes additional constraints onto the encoding process and may result in additional bitchannels with transmitter entropy $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S}) \approx 0$. Therefore, the DM rate is expected to be decreased by an accordingly constructed HY-based shaping scheme.

The main difficulty in applying this extension can be seen when coming back to the proofs of the original HY coding [HY13] as well as of Lemma 2. For the allocation of bitchannels to result in the Gelfand-Pinsker rate, i. e., $R_{\mathrm{GP}} = \lim_{N\to\infty} \frac{1}{N}\left|\mathcal{H}_{U|S} \cap \mathcal{L}_{U|Y}\right| = \mathbb{H}(\Xi|S) - \mathbb{H}(\Xi|Y)$, the condition $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y}) \leq \mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S})$ needs to hold. If this condition does not hold, the set $\mathcal{L}_{U|S} \cap \mathcal{H}_{U|Y}$ may be non-empty. This would mean that there exists at least one bitchannel i, which cannot be decoded at the receiver and thus needs to be frozen for decoding but which at the same time is also deterministic with respect to $\boldsymbol{S}_{[\![N]\!]}$ and thus needs to be available for the shaping code in order to achieve the desired distribution. In this case, the additional constraints from $\mathbb{H}(\Xi|S)$ do not allow the direct use of such a HY-based coding scheme.

Handling such a case or showing that it does not occur is subject of several publications which we review hereinafter.

We remark that $\mathcal{L} \subseteq \mathcal{H}^C$ and $\mathcal{H} \subseteq \mathcal{L}^C$. In particular in the finite block length case depending on the threshold δ, these sets may be proper subsets. To account for poorly polarized bitchannels in the construction, most authors consider the set $\mathcal{H}_{U|Y} \setminus \mathcal{H}_{U|S} \supseteq \mathcal{H}_{U|Y} \cap \mathcal{L}_{U|S}$ for analysis. In this interpretation, $\mathcal{H}_{U|Y}$ is used for FEC, $\mathcal{H}^C_{U|S} \cap \mathcal{H}^C_{U|Y}$ is used for DM and $\mathcal{H}_{U|S} \setminus \mathcal{H}_{U|Y}$ is used for data transmission.

The first to discuss the application of polar codes to Gelfand-Pinsker channels was [KU10b], [KU10a]. There, the authors consider a binary dirty paper channel of the form

$$Y = X \oplus S \oplus Z, \text{ with } \mathcal{X} = \mathcal{S} = \mathcal{Z} = \mathbb{F}_2. \tag{6.1}$$

They argue that for $P_{S|\Xi} \preceq P_{Y|\Xi}$, $\left|\mathcal{H}_{U|Y} \setminus \mathcal{H}_{U|S}\right| = o(N)$, i. e., for any $\eta > 0$, there exists a N_0 such that for all $N > N_0$, $\frac{1}{N}\left|\mathcal{H}_{U|Y} \setminus \mathcal{H}_{U|S}\right| \leq \eta$.

To handle the remaining bit indices that are required for DM as well as FEC at the same time, they propose to use these conflicting bits for DM and collect them. After a sufficient number of conflicting bits are collected, they are transmitted using some (other)

channel code. From the view of the polar code, this creates a side channel for the receiver that aids in decoding the conflicting bits. The proposed two-phase scheme has two main problems. First, the use of multiple codes increases code design complexity. Second, the system either requires large block lengths or the receiver needs to store a large number of blocks before it can start decoding.
Using a construction similar to HY coding, [GAG14] propose polar coding schemes for broadcast channels. They use the following result.

Proposition 4 ([GAG14, Lemma 7; Kor09, Lemma 1.8, Lemma 4.7])**.** *Let* $\Xi, Y, S \sim P_{\Xi,Y,S}$ *be jointly distributed random variables with* $P_{S|\Xi} \preceq P_{Y|\Xi}$*. Let* $\boldsymbol{\Xi} = \boldsymbol{U}\boldsymbol{F}^{\otimes n}$ *be the polar transform of the random vector* $\boldsymbol{U}$*. Then,* $Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y}) \leq Z(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S})$ *and thus* $\mathcal{L}_{U|S} \subseteq \mathcal{L}_{U|Y}$*.*

By combining these broadcast coding schemes with the two-phase idea explained above, the authors of [MHSU14] propose a chained broadcast coding scheme that does not require stochastic degradation. The authors of [SC16] apply this chained coding scheme to broadcast channels with side information.
Finally, the authors of [BB19], [BB20] improve on the bounds of $\frac{1}{N}\left|\mathcal{H}_{U|Y} \setminus \mathcal{H}_{U|S}\right|$ for binary DPC. They find that

$$\left|\mathcal{H}_{U|Y} \setminus \mathcal{H}_{U|S}\right| = \mathcal{O}(N^{\nu}) \tag{6.2}$$

with arbitrarily small $\nu > 0$.
A similar problem occurs when constructing polar codes for so-called wiretap channels which are dual to Gelfand-Pinsker channels in certain sense [GP19]. Suppose data $\boldsymbol{u}$ is encoded to codeword $\boldsymbol{x}$ and sent to the intended receiver $\boldsymbol{y}$ over a channel $W(y|x)$. An eavesdropper may observe $\boldsymbol{z}$ with channel $W(z|x)$. For a respective polar coding scheme, bitchannels are evaluated based on their reliability at the receiver $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y})$ and their reliability at the eavesdropper $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Z})$. Ideally, one chooses $\mathcal{L}_{U|Y} \cap \mathcal{H}_{U|Z}$ for data transmission, $\mathcal{H}_{U|Y}$ for FEC and $\mathcal{L}_{U|Z}$ for random bits. The set $\mathcal{H}_{U|Y} \cap \mathcal{L}_{U|Z}$ is desired to be empty, as it is required to be frozen for reliable decoding at the intended receiver, but can also be decoded at the eavesdropper, which would leak information about the used code.
In the literature discussing polar coding for wiretap channels, the same approaches exist as discussed above. The authors of [MV11], [ŞV13], [LYL18] assume stochastically degraded channels. Using a chaining construction similar to the one proposed in [KU10a], the authors of [ŞV13], [GB16] propose a multi-block coding scheme for general wiretap channels. We remark that the assumption of stochastically degraded channels is a common assumption when discussing wiretap channels.

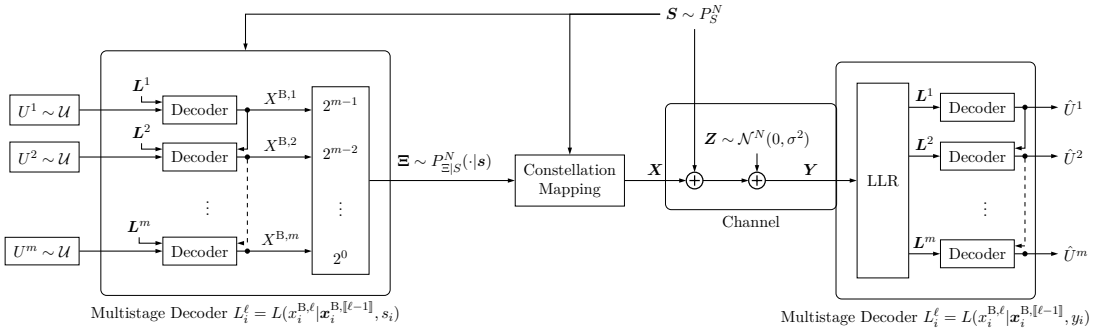


Figure 6.1: Multilevel Honda-Yamamoto Coding for Dirty Paper Coding.

Furthermore, there is a second issue within the proof of converging decoding error probabilities. The original proof uses a twofold decoder, where data bits are decoded based on the channel outputs $\boldsymbol{y}$ and DM bits are decoded based on the known prior. This is not possible in the Gelfand-Pinsker case, as the prior depends on the state S. The idea behind Gelfand-Pinsker coding is to decode Ξ without the need of decoding (S, X). Thus, the receiver does not know S and is not able to compute the prior. Although in practice the decoder decodes all bits based on $\boldsymbol{y}$, the known proofs for the convergence of $\mathcal{P}(\mathcal{E})$ require the two-fold decoding.

6.2 State-Dependent Multilevel Honda-Yamamoto Coded Modulation

In this section, we extend the MLHY transmission scheme proposed in Chapter 4 to state-dependent shaping for DPC. For this construction, we assume that a MLHY code with sufficient reliability and shaping capability is used and no conflicting bits exist, so that $\mathcal{L}_{U|S} \cap \mathcal{H}_{U|Y} = \emptyset$.

To achieve the capacity with DPC, the state-dependent labelling rule $f : \Xi \times \mathcal{S} \to \mathcal{X}$ is required as well as some state-dependent distribution $\Xi \sim P_{\Xi|S}$. The overall transmission scheme is depicted in Figure 6.1. In comparison to Figure 4.1, Figure 6.1 introduces an additional interference S which is also used to determine the likelihoods at the encoder as well as the constellation mapping. Otherwise, the same MLC structure as already known is used. For each bitlevel, the multistage decoder at the encoder provides the respective HY encoder with a target distribution, so that eventually $\widehat{P}_{\Xi|S} = P_{\Xi|S}$. At the decoder, a multistage decoder with polar decoders for each component is used.

The labelling rule can be readily implemented in MLHY. As in the case without state, a label ξ_i is represented by each bit group $\boldsymbol{x}_i^{\mathrm{B},[\![m]\!]}$ which consists of the bits at index i of

each codeword $\boldsymbol{x}^{\mathrm{B},\ell}$ at the ℓ-th bitlevel. Where in Chapter 4, we assumed a one-to-one mapping between x and ξ, this is now implemented by the state-dependent mapping f, which can be a look-up table or a computation.

In order to realize the distribution $P_{\Xi|S}$, the multistage decoder calculates the likelihoods $\mathcal{P}(x_i^{\mathrm{B},\ell}|\boldsymbol{x}_i^{\mathrm{B},[\![\ell-1]\!]}, s_i)$ and provides them to the polar decoders at each bitlevel ℓ. Again, $\mathcal{P}(x^{\mathrm{B},\ell}|\boldsymbol{x}^{\mathrm{B},[\![\ell-1]\!]}, s_i)$ conditions $P_{\Xi|S}$ on all bitlevels $\ell' < \ell$ and marginalizes $P_{\Xi|S}$ over all bitlevel $\ell' > \ell$. With this, the polar decoder together with the multistage decoder finds the most likely label sequence $\boldsymbol{\xi}$ given $\boldsymbol{s}$, so that $\boldsymbol{\xi}$ is a codeword of the MLHY code.

6.3 Simulation Results

6.3.1 Selection of Bitchannels

As with the unmodified MLHY transmission scheme, we construct the MLHY code based on two bitchannel entropies. Using a MC simulation, we calculate $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{S})$ and $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y})$, where U is sampled conditional on S so that $\Xi, S, Y \sim P_{\Xi,S,Y}$.

Conditional Polarization

Although $\mathcal{L}_{U|S} \subseteq \mathcal{L}_{U|Y}$ may not be guaranteed, we find that

$$\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{S}) \geq \mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y}) \tag{6.3}$$

may still hold in some practically relevant cases. Therefore, whether a chaining construction such as the one proposed by [KU10a] is necessary, has to be determined for each specific channel individually. In case the computed uncertainties satisfy Equation (6.3), the MLHY code can be used as-is for state-dependent shaping without any further modifications.

We observed that Equation (6.3) holds for the DPC scenarios discussed in [Len18, Section 5.2]. There, achievable rates for DPC with a uniform BPSK interference and a SIR of 5 dB are considered. DPC schemes and in particular the respective optimized conditional distributions are proposed for BPSK signalling as well as 4-ASK signalling with SNRs reaching from -5 dB to 20 dB. For these scenarios, the rate is strictly positive, i. e., $\mathbb{H}(\Xi|S) > \mathbb{H}(\Xi|Y)$. Using the optimized symbol distributions presented there, Equation (6.3) holds for both constellation sizes over the considered SNR range and block lengths $2^8 \leq N \leq 2^{13}$.

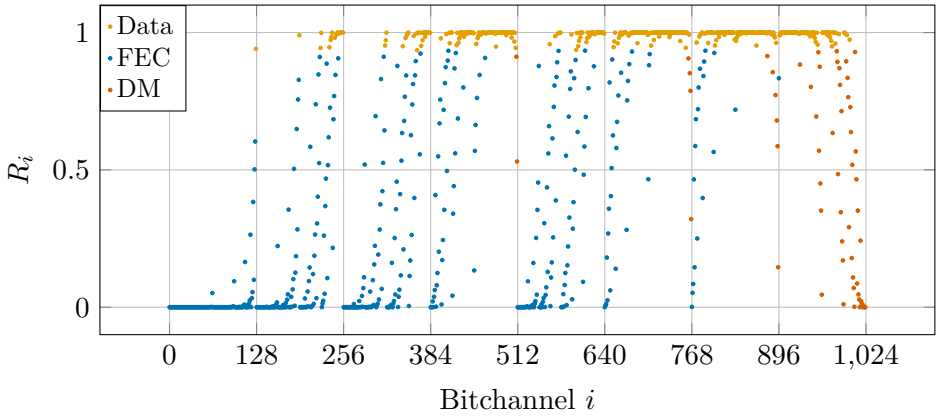


Figure 6.2: Bitchannel $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S}) - \mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y})$ for DPC with $M = 2$, $N = 1024$ and SNR $= 2.5\,\mathrm{dB}$.

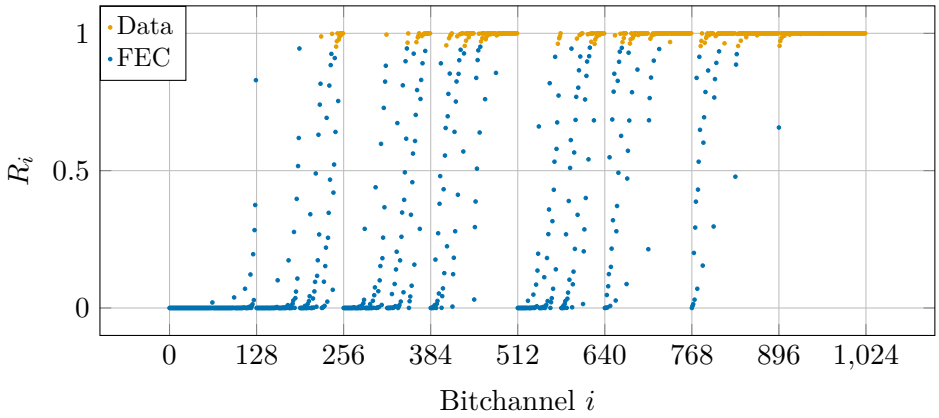


Figure 6.3: Bitchannel $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}) - \mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y})$ for AWGN with $M = 2$, $N = 1024$ and SNR $= 0.9\,\mathrm{dB}$.

We remark that the bitchannel entropies and thus also whether Equation (6.3) holds depends solely on $P_{S,\Xi,Y}$ and N. Therefore, it is a property of the channel together with the used symbol constellation and distribution. Whether Equation (6.3) holds or not can not be manipulated by code design, i. e., bitchannel selection alone.

Speed of Polarization

We further analyze whether the additional conditioning decreases finite length polarization effects. For this, we compare the number of weakly polarized bitchannels for the AWGN channel and for the dirty paper channel.

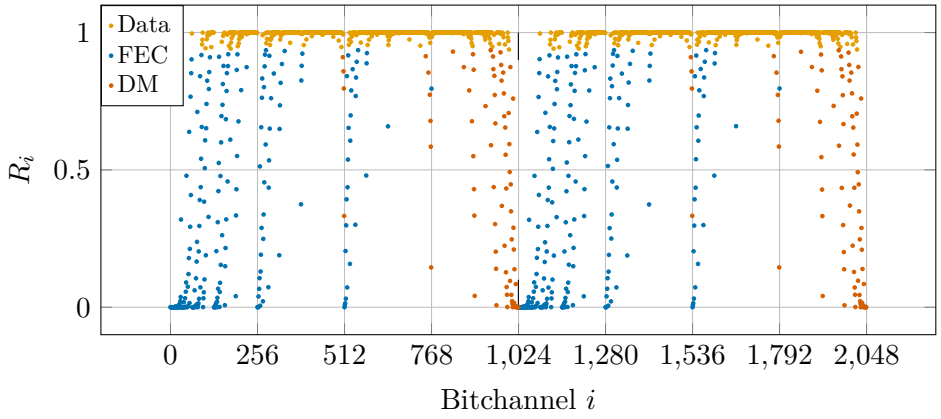


Figure 6.4: Bitchannel $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{S}) - \mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y})$ for DPC with $M = 4$, $N = 1024$ and SNR $= 10\,\mathrm{dB}$.

Figure 6.2 and Figure 6.3 show

$$R_i = \mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S}) - \mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y}) = \mathbb{I}(U_i; \boldsymbol{Y}|\boldsymbol{U}_{[\![i-1]\!]}) - \mathbb{I}(U_i; \boldsymbol{S}|\boldsymbol{U}_{[\![i-1]\!]}) \tag{6.4}$$

for each bitchannel for BPSK over the dirty paper channel and over the AWGN channel, respectively. Analogously to the code construction explained in Section 4.3, the bitchannels with the highest R_i are used for data transmission and the remaining bitchannels are used for DM and FEC. In the figures, blue (·) bitchannels have high $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{Y})$ and are used for FEC, yellow (·) bitchannels have high R_i and are used for data transmission and the remaining green (·) bitchannels have low $\mathbb{H}(U_i|\boldsymbol{U}_{[\![i-1]\!]}, \boldsymbol{S})$ and are used for DM. Both, the code for DPC as well as the code for the AWGN channel without interference, have an identical number of data bitchannels and an identical threshold for R_i. The constellation-constrained channel capacities for the corresponding SE are 0 dB for the dirty paper channel and $-1.2\,\mathrm{dB}$ for the AWGN channel. We define weakly polarized bitchannels as bitchannels for which $\varepsilon < R_i < 1 - \varepsilon$ with threshold ε.

To achieve 457 out of $mN = 1024$ bitchannels for data transmission with a threshold of $\varepsilon \approx 0.064$, the code for DPC needs a back-off from capacity of 2.5 dB. For polar coding over the AWGN channel, a back-off of 2.3 dB is enough to reach 457 data bitchannels with this ε. At these parameters, the HY code for DPC has 240 weakly polarized bitchannels and the polar code for the AWGN channel 219.

Figure 6.4 and Figure 6.5 show the bitchannel R_i for 4-ASK MLHY codes for transmission over the dirty paper channel and over the AWGN channel, respectively. Both codes are designed to have 1450 out of $mN = 2048$ bitchannels for data transmission and $\varepsilon \approx 0.064$. For the corresponding SE, the capacities are 8.5 dB for the dirty paper channel and 8 dB

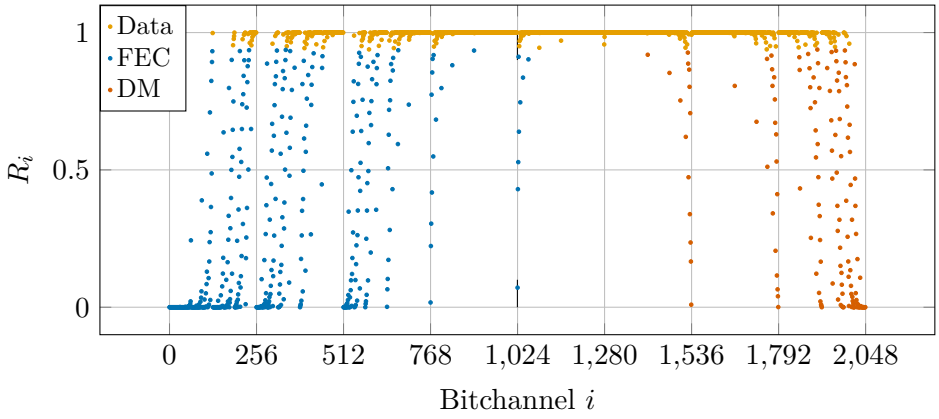


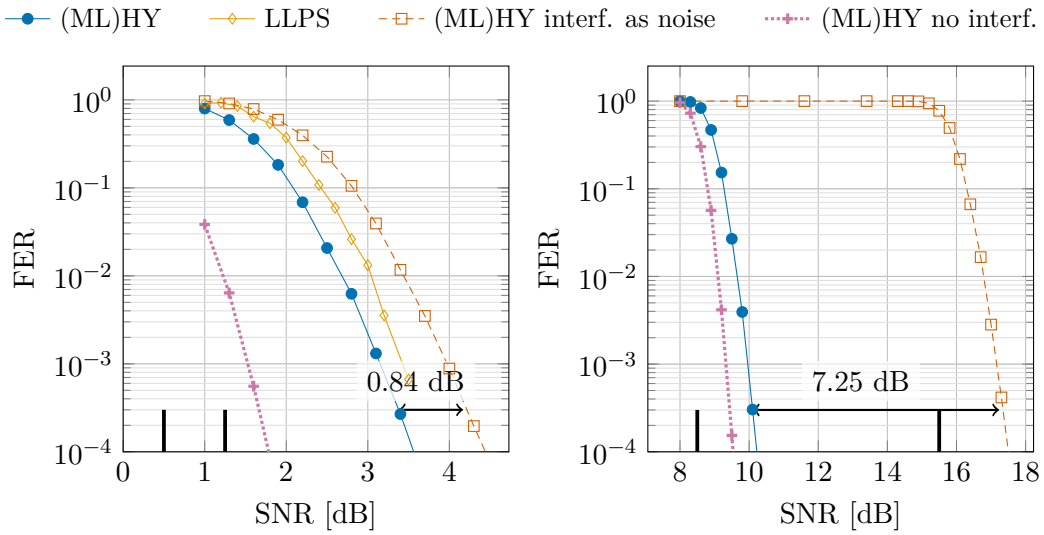
Figure 6.5: Bitchannel $\mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}) - \mathbb{H}(U_i^\ell|\boldsymbol{U}_{[\![i-1]\!]}^\ell, \boldsymbol{U}^{[\![\ell-1]\!]}, \boldsymbol{Y})$ for AWGN with $M = 4$, $N = 1024$ and SNR $= 9.5\,\mathrm{dB}$.

for the AWGN channel. The MLHY code for DPC has 372 weakly polarized bitchannels whereas the MLHY for the AWGN channel has 300. Both codes have a back-off from capacity of 1.5 dB.

The examples for BPSK and 4-ASK show that DPC using MLHY codes is possible without incurring overly large overhead. In the cases we analyze, the difference in back-off from the respective capacity is below 0.5 dB for communication with and without interference. Similarly, the increase in the number of weakly polarized bitchannels normalized to mN is below 5 % at $N = 1024$.

6.3.2 Coding for the Dirty Paper Channel

We compare performances for three cases, namely DPC, transmission over the same dirty paper channel but treating the interference as an additional noise that is not known to transmitter or receiver (interference as noise) and transmission over the respective AWGN without any interference. The HY and MLHY codes used are constructed identically to the ones in Section 4.4. The encoder decides deterministically for the minimum-energy codeword out of the $L = 32$ SCL candidate codewords. For the transmission with interference treated as noise, the receiver knows the distribution of the BPSK interference. Furthermore, we compare the scheme proposed herein with the DPC scheme using linear layered probabilistic shaping (LLPS) proposed by [BLCS19]. LLPS extends the PAS architecture to asymmetric target distributions by turning the linear FEC into a coset code, from which a codeword with the desired hamming weight is chosen. For their DPC results, they use a rate 1/2 LDPC code with block length $N = 1056$.



(a) BPSK, 0.4697 bpcu
dSNR = 2.5 dB

(b) 4-ASK, 1.4 bpcu
dSNR = 10.0 dB
interf. as noise dSNR = 16.5 dB

Figure 6.6: Performance of (ML)HY codes for DPC compared to LLPS [BLCS19], (ML)HY with interference as noise and (ML)HY without interference. The MLHY codes have $N = 1024$ and use an outer CRC-16.

Figure 6.6 shows FER performance vs SNR for communication over the dirty paper channel. The channel capacities for DPC and treating the interference as noise are denoted by the two black bars in each figure. The DPC gain is the difference in SNR which is required to achieve the same SE and FER when treating the interference as noise. HY-based DPC for BPSK and MLHY-based DPC for 4-ASK constellations achieves DPC gains of 0.84 dB and 7.25 dB, respectively. We observe that these gains lie above the theoretical DPC gains of 0.75 dB and 7 dB. This is due to the more moderate slope of the interference as noise case compared to the DPC case. We also compare DPC results to coding over an AWGN channel without interference. Whilst for BPSK the finite-constellation DPC scheme has a distance of 1 dB to 2 dB to the AWGN case [Len18, Section 5.2.3], [BLCS19], DPC for 4-ASK transmission can achieve almost the same rates as coding without interference [Len18]. For finite-length HY, we observe a gap of 1.6 dB to AWGN performance and for MLHY a gap of 0.5 dB.

We further observed that similar to the LDPC codes used by [BLCS19], DPC with polar codes requires longer block lengths in order to achieve FER vs SNR slopes comparable to the case without interference.

7 Conclusion

In this thesis we present a joint FEC and DM scheme for arbitrary target distributions on higher-order modulation alphabets based on MLC and polar coding. We prove that the proposed MLHY scheme achieves the constellation-constrained capacity of a memoryless channel and show that the multilevel construction does not worsen the $\mathcal{O}(2^{-N^{\beta'}})$ error exponents of HY coding. We show an effective method for construction that only requires the dSNR and the SE as parameters. Using numerical simulations, we demonstrate performance results and observe that the theoretical shaping gain is recovered and that MLHY coding slightly outperforms PC-PAS for short block lengths while being less complex to implement and without any restrictions on the feasible target distributions. Additionally, we extend MLHY to DPC and show that this construction is employable in practice, even though a general theorem for its capacity achieving property for state-dependent DM is not known. Still, the extension recovers the DPC gains in simulations and almost achieves interference-free AWGN performance for 4-ASK transmission.

Future research might improve the theoretical foundation of HY-based DPC by deriving necessary conditions under which $\mathcal{L}_{U|S} \cap \mathcal{H}_{U|Y} = \emptyset$, so that this construction is capacity achieving. This might connect to the design of capacity achieving constellations and distributions for finite-constellation DPC in general. Similarly, the existence of a good higher-order modulation Marton-coding scheme based on the scheme proposed herein seems likely and might motivate further works.

The implementation of MLHY coding requires the receiver to know the generated symbol distribution $\widehat{P}_X$. In case of deterministic encoding, this is not determined directly by the prior P_X fed to the DM SC decoder as with probabilistic encoding. Instead, $\widehat{P}_X$ also depends heavily on the bitchannel selection and needs to be measured during code construction. Future work might improve on this by finding more analytical or more efficient approaches to determine $\widehat{P}_X$ and improved algorithms to select the DM bitchannels.

While polar codes were the first codes to be described via the polar transform, they are not the only class of codes that can be constructed this way. By choosing different bitchannels for DM or FEC, setting frozen bits depending on the previous bits [TM15] or employing different decoding rules, the class of codes and coset codes that can be

constructed using the polar transform includes many more interesting codes. Their applicability to a MLHY-based shaping scheme might be explored by further research.

Bibliography

[Arı09] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, Jun. 2009. DOI: 10.1109/TIT.2009.2021379.

[Arı10] ——, "Source polarization," in *IEEE International Symposium on Information Theory (ISIT)*, Austin, TX, Jun. 2010, pp. 899–903, ISBN: 978-1-4244-7892-7. DOI: 10.1109/ISIT.2010.5513567.

[AT09] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, Jun. 2009, pp. 1493–1495, ISBN: 978-1-4244-4312-3. DOI: 10.1109/ISIT.2009.5205856.

[BB19] B. Beilin and D. Burshtein, "On polar coding for binary dirty paper," in *IEEE International Symposium on Information Theory (ISIT)*, Paris, France, Sep. 2019, pp. 1402–1406, ISBN: 978-1-5386-9291-2. DOI: 10.1109/ISIT.2019.8849333.

[BB20] ——, "On polar coding for side information channels," *IEEE Transactions on Information Theory*, p. 1, Nov. 2020. DOI: 10.1109/TIT.2020.3035658.

[BBB15] A. Balatsoukas-Stimming, M. Bastani Parizi, and A. Burg, "LLR-based successive cancellation list decoding of polar codes," *IEEE Transactions on Signal Processing*, vol. 63, no. 19, pp. 5165–5179, Jun. 2015. DOI: 10.1109/TSP.2015.2439211.

[BG16] G. Böcherer and B. C. Geiger, "Optimal quantization for distribution synthesis," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6162–6172, Sep. 2016. DOI: 10.1109/TIT.2016.2610433.

[BİX20] R. Böhnke, O. İşcan, and W. Xu, "Multi-level distribution matching," *IEEE Communications Letters*, vol. 24, no. 9, pp. 2015–2019, May 2020. DOI: 10.1109/LCOMM.2020.2993929.

[BLCS19] G. Böcherer, J. D. Lentner Ibañez, A. Cirino, and F. Steiner, *Probabilistic parity shaping for linear codes*, Feb. 2019. [Online]. Available: `http://arxiv.org/pdf/1902.10648v1`.

[BPYS17] G. Böcherer, T. Prinz, P. Yuan, and F. Steiner, "Efficient polar code construction for higher-order modulation," in *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, San Francisco, CA, Mar. 2017, pp. 1–6, ISBN: 978-1-5090-5908-9. DOI: `10.1109/WCNCW.2017.7919039`.

[BSS15] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4651–4665, Oct. 2015. DOI: `10.1109/TCOMM.2015.2494016`.

[CB15] R. A. Chou and M. R. Bloch, "Using deterministic decisions for low-entropy bits in the encoding and decoding of polar codes," in *Allerton Conference on Communication, Control, and Computing*, Monticello, IL: IEEE, Oct. 2015, pp. 1380–1385, ISBN: 978-1-5090-1824-6. DOI: `10.1109/ALLERTON.2015.7447169`.

[CDJ+19] M. C. Coşkun, G. Durisi, T. Jerkovits, G. Liva, W. Ryan, B. Stein, and F. Steiner, "Efficient error-correcting codes in the short blocklength regime," *Physical Communication*, vol. 34, pp. 66–79, 2019. DOI: `https://doi.org/10.1016/j.phycom.2019.03.004`.

[CO90] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the gaussian channel," *IEEE Transactions on Information Theory*, vol. 36, no. 4, pp. 726–740, Jul. 1990. DOI: `10.1109/18.53734`.

[Cos83] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983. DOI: `10.1109/TIT.1983.1056659`.

[CT06] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. Hoboken, N.J.: Wiley, 2006, ISBN: 9780471241959.

[DPN21] J. Dai, J. Piao, and K. Niu, "Progressive rate-filling: A framework for agile construction of multilevel polar-coded modulation," *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 1123–1127, Feb. 2021. DOI: `10.1109/LWC.2021.3059841`.

[EK11] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge and New York: Cambridge University Press, 2011, ISBN: 9781139190916.

[ESZ05] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3820–3833, Oct. 2005. DOI: 10.1109/TIT.2005.856935.

[EtB05] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3417–3432, Sep. 2005. DOI: 10.1109/TIT.2005.855586.

[FHW98] R. F. Fischer, J. B. Huber, and U. Wachsmann, "On the combination of multilevel coding and signal shaping," in *ITG Conference on Source and Channel Coding*, 1998, pp. 273–278.

[For88] G. Forney, "Coset codes. I. introduction and geometrical classification," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 1123–1151, Sep. 1988. DOI: 10.1109/18.21245.

[For92] ——, "Trellis shaping," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 281–300, Mar. 1992. DOI: 10.1109/18.119687.

[FTC00] G. D. Forney, M. D. Trott, and S.-Y. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 820–850, May 2000. DOI: 10.1109/18.841165.

[GAG14] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for broadcast channels," *IEEE Transactions on Information Theory*, vol. 61, no. 2, pp. 758–782, Dec. 2014. DOI: 10.1109/TIT.2014.2378172.

[Gal68] R. G. Gallager, *Information theory and reliable communication*. New York [u.a.]: Wiley, 1968, ISBN: 978-0-471-29048-3.

[GB16] T. C. Gulcu and A. Barg, "Achieving secrecy capacity of the wiretap channel and broadcast channel with a confidential component," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 1311–1324, Nov. 2016. DOI: 10.1109/TIT.2016.2631223.

[GFAW20] 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. DOI: 10.3390/e22050581.

[GP19] Z. Goldfeld and H. H. Permuter, "Wiretap and gelfand-pinsker channels analogy and its applications," *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 4979–4996, Apr. 2019. DOI: 10.1109/TIT.2019.2910106.

[GP80] S. I. Gel'fand and M. S. Pinsker, "Coding for channels with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, Jan. 1980.

[HY13] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7829–7838, Sep. 2013. DOI: 10.1109/TIT.2013.2282305.

[İBX17] O. İşcan, R. Böhnke, and W. Xu, "Shaped polar codes for higher order modulation," *IEEE Communications Letters*, vol. 22, no. 2, pp. 252–255, Oct. 2017. DOI: 10.1109/LCOMM.2017.2766621.

[İBX19] ——, "Probabilistic shaping using 5G new radio polar codes," *IEEE Access*, vol. 7, pp. 22 579–22 587, Feb. 2019. DOI: 10.1109/ACCESS.2019.2898103.

[KK93] A. Khandani and P. Kabal, "Shaping multidimensional signal spaces. I. optimum shaping, shell mapping," *IEEE Transactions on Information Theory*, vol. 39, no. 6, pp. 1799–1808, Nov. 1993. DOI: 10.1109/18.265491.

[Kor09] S. B. Korada, "Polar codes for channel and source coding," Doctoral Thesis, École polytechnique fédérale de Lausanne, Jul. 2009. DOI: 10.5075/epfl-thesis-4461.

[KP93] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for gaussian channels," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 913–929, May 1993. DOI: 10.1109/18.256499.

[KŞU10] S. B. Korada, E. Şaşoğlu, and R. L. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6253–6264, Nov. 2010. DOI: 10.1109/TIT.2010.2080990.

[KU10a] S. B. Korada and R. L. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1751–1768, Mar. 2010. DOI: 10.1109/TIT.2010.2040961.

[KU10b] ——, "Polar codes for slepian-wolf, wyner-ziv, and gelfand-pinsker," in *IEEE Information Theory Workshop (ITW)*, Cairo, Egypt, Jan. 2010, pp. 1–5, ISBN: 978-1-4244-6372-5. DOI: 10.1109/ITWKSPS.2010.5503220.

[Len18] J. D. Lentner Ibañez, "Dirty paper coding for higher-order modulation and finite constellation interference," Master's Thesis, Technische Universität München, Mar. 2018.

[Liu16] 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.

[LYL18] L. Liu, Y. Yan, and C. Ling, "Achieving secrecy capacity of the gaussian wiretap channel with polar lattices," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1647–1665, Jan. 2018. DOI: 10.1109/TIT.2018.2794327.

[LYLW18] L. Liu, Y. Yan, C. Ling, and X. Wu, "Construction of capacity-achieving lattice codes: Polar lattices," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 915–928, Oct. 2018. DOI: 10.1109/TCOMM.2018.2876113.

[LZH15] L. Li, W. Zhang, and Y. Hu, *On the error performance of systematic polar codes*, Apr. 2015. [Online]. Available: https://arxiv.org/pdf/1504.04133.

[MHSU14] M. Mondelli, S. H. Hassani, I. Sason, and R. L. Urbanke, "Achieving marton's region for broadcast channels using polar codes," *IEEE Transactions on Information Theory*, vol. 61, no. 2, pp. 783–800, Nov. 2014. DOI: 10.1109/TIT.2014.2368555.

[MHU18] M. Mondelli, S. H. Hassani, and R. L. Urbanke, "How to achieve the capacity of asymmetric channels," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3371–3393, Jan. 2018. DOI: 10.1109/TIT.2018.2789885.

[MKM+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, M4B.2, ISBN: 978-1-943580-53-8. DOI: 10.1364/OFC.2019.M4B.2.

[MT09] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," in *IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, Jun. 2009, pp. 1496–1500, ISBN: 978-1-4244-4312-3. DOI: 10.1109/ISIT.2009.5205857.

[MV11] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6428–6443, Oct. 2011. DOI: 10.1109/TIT.2011.2162275.

[PPV10] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, Apr. 2010. DOI: 10.1109/TIT.2010.2043769.

[PY18] T. Prinz and P. Yuan, "Successive cancellation list decoding of BMERA codes with application to higher-order modulation," in *IEEE International Symposium on Turbo Codes & Iterative Information Processing (ISTC)*, Hong Kong, China, Dec. 2018, pp. 1–5. DOI: 10.1109/ISTC.2018.8625293.

[PYB+17] T. Prinz, P. Yuan, G. Böcherer, F. Steiner, O. İşcan, R. Böhnke, and W. Xu, "Polar coded probabilistic amplitude shaping for short packets," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Sapporo, Japan, Jul. 2017, pp. 1–5, ISBN: 978-1-5090-3009-5. DOI: 10.1109/SPAWC.2017.8227653.

[RV19] M. Rowshan and E. Viterbo, "How to modify polar codes for list decoding," in *IEEE International Symposium on Information Theory (ISIT)*, Paris, France, Jul. 2019, pp. 1772–1776, ISBN: 978-1-5386-9291-2. DOI: 10.1109/ISIT.2019.8849539.

[SB15] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 430–434, Nov. 2015. DOI: 10.1109/TIT.2015.2499181.

[SC16] J. Sima and W. Chen, "Polar codes for broadcast channels with receiver message side information and noncausal state available at the encoder," in *IEEE International Symposium on Information Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 993–997. DOI: 10.1109/ISIT.2016.7541448.

[Sei15] M. Seidl, "Polar coding: Finite-length aspects," Doctoral Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, Apr. 2015. [Online]. Available: https://nbn-resolving.org/urn:nbn:de:bvb:29-opus4-62013.

[SG17] P. Schulte and B. C. Geiger, "Divergence scaling of fixed-length, binary-output, one-to-one distribution matching," in *IEEE International Symposium on Information Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 3075–3079. DOI: 10.1109/ISIT.2017.8007095.

[Sha48] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[SSSH13] M. Seidl, A. Schenk, C. Stierstorfer, and J. B. Huber, "Polar-coded modulation," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4108–4119, Sep. 2013. DOI: 10.1109/TCOMM.2013.090513.130433.

[ŞTA09] E. Şaşoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," in *IEEE Information Theory Workshop (ITW)*, Taormina, Italy, Oct. 2009, pp. 144–148, ISBN: 978-1-4244-4982-8. DOI: 10.1109/ITW.2009.5351487.

[Sto02] 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.

[ŞV13] E. Şaşoğlu and A. Vardy, "A new polar coding scheme for strong security on wiretap channels," in *IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 1117–1121, ISBN: 978-1-4799-0446-4. DOI: 10.1109/ISIT.2013.6620400.

[Tah17] B. Tahir, "Construction and performance of polar codes for transmission over the awgn channel," Master's Thesis, Technische Universität Wien, Oct. 2017. [Online]. Available: https://resolver.obvsg.at/urn:nbn:at:at-ubtuw:1-103376.

[TM15] P. Trifonov and V. Miloslavskaya, "Polar subcodes," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 2, pp. 254–266, Nov. 2015. DOI: 10.1109/JSAC.2015.2504269.

[TV13] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6562–6582, Jul. 2013. DOI: 10.1109/TIT.2013.2272694.

[TV15] ——, "List decoding of polar codes," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2213–2226, Mar. 2015. DOI: 10.1109/TIT.2015.2410251.

[Ung82] G. Ungerböck, "Channel coding with multilevel/phase signals," *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982. DOI: 10.1109/TIT.1982.1056454.

[WDY+20] T. Wiegart, F. Da Ros, M. P. Yankov, F. Steiner, S. Gaiarin, and R. D. Wesel, "Probabilistically shaped 4-pam for short-reach im/dd links with a peak power constraint," *Journal of Lightwave Technology*, vol. 39, no. 2, pp. 400–405, Oct. 2020. DOI: 10.1109/JLT.2020.3029371.

[WFH99] U. Wachsmann, R. F. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999. DOI: 10.1109/18.771140.

[WKP05] C.-C. Wang, S. R. Kulkarni, and H. V. Poor, "Density evolution for asymmetric memoryless channels," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4216–4236, Nov. 2005. DOI: 10.1109/TIT.2005.858931.

[WSSY19] T. Wiegart, F. Steiner, P. Schulte, and P. Yuan, "Shaped on–off keying using polar codes," *IEEE Communications Letters*, vol. 23, no. 11, pp. 1922–1926, Jul. 2019. DOI: 10.1109/LCOMM.2019.2930511.

[YPB+19] P. Yuan, T. Prinz, G. Böcherer, O. İşcan, R. Böhnke, and W. Xu, "Polar code construction for list decoding," in *International ITG Conference on Systems, Communications and Coding (SCC)*, Rostock, Germany: VDE, Feb. 2019, pp. 1–6. DOI: 10.30420/454862022.

[YS18] P. Yuan and F. Steiner, *Construction and decoding algorithms for polar codes based on* 2×2 *non-binary kernels*, Jul. 2018. [Online]. Available: http://arxiv.org/pdf/1807.03767v2.

[ZLJ+21] H. Zhou, Y. Li, X. Jia, C. Gao, Y. Liu, J. Qiu, X. Hong, H. Guo, Y. Zuo, and J. Wu, "Polar coded probabilistic shaping pam8 based on many-to-one mapping for short-reach optical interconnection," *Optics express*, vol. 29, no. 7, pp. 10 209–10 220, Mar. 2021. DOI: 10.1364/OE.418045.

[ZSE02] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, Aug. 2002. DOI: 10.1109/TIT.2002.1003821.