UNIVERSITY OF CALIFORNIA SAN DIEGO

Polar Codes for Data Storage and Communication Network Applications

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Electrical Engineering
(Communication Theory and Systems)

by

Karthik Nagarjuna Tunuguntla

Committee in charge:

Professor Paul H. Siegel, Chair
Professor Patrick J. Fitzsimmons
Professor Alireza S. Golsefidy
Professor Alon Orlitsky

2022

The Dissertation of Karthik Nagarjuna Tunuguntla is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To my parents

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I owe sincere gratitude to the people who helped me conduct research during my PhD program at UCSD. I would like to take the opportunity to thank each one of them here.

Firstly, I express my deepest gratitude to my advisor Prof. Paul H. Siegel for his excellent guidance majorly contributing to my thesis. Without his experience and advice, it would not have been possible for me to finish my dissertation. Prof. Siegel has given me a lot of freedom in everything from choosing the topic of dissertation to selection of individual research problems. This provided me the opportunity to explore my curiosity and make significant contributions to the field of my study. Before starting my research with Prof. Siegel, I took three coding theory classes offered by him which I found to be really fascinating. I admired him a lot as a teacher while taking those classes, which led to my interest in joining his research group. He has provided me a lot of his valuable time for discussing the research ideas, reviewing my papers and presentations in the course of study. His feedback often helped me enhance my research ideas. I specifically learnt how important writing and reading skills are for technical study and research while working with him.

I would like to thank Prof. Fitzsimmons Patricks, Prof. Alireza Salehi Golsefidy, and Prof. Alon Orlitsky for being on my defense committee.

I improved my mathematical thinking by taking algebra, real analysis and probability theory graduate courses from Math department at UCSD.

I would like to thank NSF projects CCF-1415109, CCF-1619053, and Samsung Electronics Co., Ltd. for supporting my research work contributing to the dissertation.

Chapter 2 in part is a reprint of material in the paper: Karthik Nagarjuna Tunuguntla, Paul H. Siegel, "Universal polar coding for asymmetric channels," *2018 IEEE Information Theory Workshop (ITW)*, pp. 1-5, Guangzhou, China, November 2018. The dissertation author was the primary investigator and author of the paper.

Chapter 3 in part is a reprint of material in the paper: Karthik Nagarjuna Tunuguntla, Paul H. Siegel, "Polar shaping codes for costly noisy and noiseless channels," *2021 International*

Symposium on Information Theory (ISIT), pp. 2560-2565, Melbourne, Australia, June 2021. The dissertation author was the primary investigator and author of the paper.

Chapter 4 in part is a reprint of material in the paper: Karthik Nagarjuna Tunuguntla, Paul H. Siegel, "Slepian-Wolf polar coding with unknown correlation,"*2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 132-139, October, 2019. The dissertation author was the primary investigator and author of the paper.

Chapter 5 in part is a reprint of material in the paper: Karthik Nagarjuna Tunuguntla, Paul H. Siegel, "Polar coding for multi-level 3-receiver broadcast channels," *2020 Information Theory Workshop (ITW)*, pp. 1-5, Riva Del Garda, Italy, April 2021. The dissertation author was the primary investigator and author of the paper.

VITA

2006–2010 B.Tech. in Electronics and Communications Engineering, Jawaharlal Nehru Technological University, Hyderabad (JNTUH), India

2010–2012 M.Tech. in Electrical Engineering (Communication and Signal Processing), Indian Institute of Technology, Bombay (IITB), India

2015–2018 M.S. in Electrical Engineering (Communication Theory and Systems), University of California San Diego

2015–2022 Ph.D. in Electrical Engineering (Communication Theory and Systems), University of California San Diego

PUBLICATIONS

Karthik Nagarjuna Tunuguntla, Paul H. Siegel, "Polar shaping codes for costly noisy and noiseless channels," *2021 International Symposium on Information Theory (ISIT)*, pp. 2560-2565, Melbourne, Australia, June 2021.

Karthik Nagarjuna Tunuguntla, Paul H. Siegel, "Polar coding for multi-level 3-receiver broadcast channels," *2020 Information Theory Workshop (ITW)*, pp. 1-5, Riva Del Garda, Italy, April 2021.

Karthik Nagarjuna Tunuguntla, Paul H. Siegel, "Slepian-Wolf polar coding with unknown correlation,"*2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 132-139, October 2019.

Karthik Nagarjuna Tunuguntla, Paul H. Siegel, "Universal polar coding for asymmetric channels," *2018 IEEE Information Theory Workshop (ITW)*, pp. 1-5, Guangzhou, China, November 2018.

ABSTRACT OF THE DISSERTATION

Polar Codes for Data Storage and Communication Network Applications

by

Karthik Nagarjuna Tunuguntla

Doctor of Philosophy in Electrical Engineering
(Communication Theory and Systems)

University of California San Diego, 2022

Professor Paul H. Siegel, Chair

The dissertation provides polar coding techniques for a variety of source and channel models with applications to storage and communication networks.

We first provide universal polar codes for asymmetric compound channels that avoid common randomness. A staircase alignment of polar blocks is considered in the code construction. An MDS code is used in each column achieving the universality and a scrambling technique is implemented for each column helping avoid common randomness. These compound asymmetric channels are used for modelling flash-memories, such as MLCs (multi-level cell flash memories), and TLCs (three-level cell flash memories) memories. Hence the proposed universal polar codes

for asymmetric channels can be used for flash memory error correction.

The costly noiseless channel model was used to model a flash memory device. Each of the voltage levels to which a flash memory cell can be programmed has an associated wear cost which reflects the damage caused to the cell by repeated programming to that level. Shaping codes that minimize the average cost per channel symbol for a specified rate and shaping codes that minimize the average cost per source symbol (i.e., the total cost) have been shown to reduce cell wear and increase the lifetime of the memory. Hence, we study polar shaping codes for costly noiseless channels minimizing total cost. We also study polar shaping codes for costly noisy channels for the design of efficient codes that combine wear reduction and error correction for use in a noisy flash memory device.

A novel scheme based on polar codes is proposed to compress a uniform source when a side information correlated with the source is available at the receiver while the conditional distribution of the side information given the source is symmetric and unknown to the source. An adaptation of universal polar codes with an incorporation of the linear code duality between channel coding and Slepian-Wolf coding is used in the design of those codes. Optimal rate is achieved through the proposed codes for the source model. These codes can be used in a wireless sensor network where the measurements tracked at two different nodes are correlated and the correlation may not always be fixed due to environmental changes such as weather. The nodes communicate the information sensed or measured by them to a central location.

Finally, we provide a capacity-achieving polar coding strategy on a multi-level 3-receiver broadcast channel in which the second receiver is degraded (stochastically) from the first receiver for the transmission of a public message intended for all the receivers and a private message intended for the first receiver. A chaining strategy, translating the ideas of superposition coding, rate-splitting and indirect coding into polar coding, is used in the construction. The codes designed for such a channel model and setting can be used for video and audio file transfer in a client-server network where the individual clients are a computer and two mobile phones.

Chapter 1

Introduction

1.1 Polar codes

Polar codes, introduced in 2009 by Arikan [1], are provably capacity-achieving codes for binary-input symmetric channels. They have an explicit construction and admit low-complexity encoding and decoding. These properties have made them the subject of extensive investigation since their introduction in 2009. The polar code construction is based on the phenomenon of polarization, whereby independent copies of a discrete memoryless channel are recursively transformed into channels that have nearly full capacity or nearly zero capacity. The channels with nearly full capacity are used to transmit information reliably, while the channels with nearly zero capacity are provided with predefined dummy bits.

When the recursive polar transformation is adapted to independent, identical random variables, the resulting phenomenon is referred to as source polarization [3]. Source polarization has led to code constructions for lossless compression, Slepian-Wolf source coding, lossy source coding, and Wyner-Ziv coding [3], [26], [27] that achieve fundamental rate limits. Capacity-achieving polar coding schemes are developed for asymmetric channels [24], [11], [33].

Polar codes are specifically designed for a given channel to achieve its capacity. Hence they are not universally applicable to any DMC for achieving its capacity. Some constructions based on polar codes have been proposed to achieve the universality [22], [23], [48]. Polar coding schemes for networks, such as Gilfand-Pinsker coding [27], broadcast channels [17], [32],

interference channels [54], relay channels [53], multiple-access channels [40], [5], and wiretap channels [30], [12], [39] have been investigated. Polar code constructions based on monotone chain rules are used for the Slepian-Wolf problem [2] and the multiple description problem [6] to attain all points in the rate region without time-sharing. All the polar code constructions for these multi-terminal settings have low-complexity encoding and decoding methods.

Polar code construction requires determination of good bit-channels of almost full capacity and bad bit-channels of almost zero capacity obtained after polarization. Efficient low-complexity algorithms to determine these bit-channel sets used for polar code constructions have been proposed [35], [50], [41]. The polarization phenomenon extends to arbitrary alphabets, such as finite fields [47], [36]. Asymptotic polarization behaviour in large block regimes [4], [25], [36] and its dependency on transmission rate [21] have been studied. Finite length scaling of polar codes – the required scaling of block length to maintain a fixed probability of decoding error as the code rate approaches capacity – has been studied for different polarization kernels and alphabet sizes [20], [19], [18], [42]. A generalization of the polar coding scheme exploiting several homogeneous kernels over alphabets of different sizes and its polarization behaviour [43] is studied.

In this dissertation, we construct polar codes for settings that apply to storage devices and communication networks. Specific contributions include: universal codes for asymmetric channels, suitable for error correction in flash memories; shaping codes to for costly noiseless and noisy channels, intended to enhance the reliability and extend the lifetime of flash memory devices; source coding schemes with correlated side information at the receiver but with unknown correlation at the transmitter, applicable to communication between nodes in a sensor network; and a capacity-achieving coding scheme for a three-receiver broadcast channel.

1.2 Polarization

Polarization process is used in all the code constructions provided in this dissertation as they are developed based on polar codes. We now review the source polarization results briefly.

We express any set of random variables $X_i, X_{i+1}, \ldots, X_j$ $(i < j)$ by a row vector $(X_i, X_{i+1}, \ldots, X_j)$ which is denoted by $X^{i:j}$. We denote the set $\{1, 2, 3, \ldots, N\}$ by $[N]$. If $U^{1:N}$ is a row vector and $\mathscr{A} \subset [N]$, then $U^{\mathscr{A}}$ denotes the row vector consisting of elements in $U^{1:N}$ corresponding to the subset of positions $\mathscr{A}$ in the same order.

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)$ be i.i.d. random tuples distributed according to $p(x)p(y|x)$ over $\mathscr{X} \times \mathscr{Y}$ and $N = 2^n$. Let $\mathscr{X} = \{0, 1\}$. Let G_N be the conventional polar transformation [1], represented by a binary matrix of dimension $N \times N$. If $U^{1:N} = X^{1:N} G_N$, then we denote $\mathbb{P}(U^{1:N} = u^{1:N})$ by $P_{U^{1:N}}(u^{1:N})$ and similarly we denote $\mathbb{P}(U_i = u_i | U^{1:i-1} Y^{1:N} = u^{1:i-1} y^{1:N})$ by $P_{U_i|U^{1:i-1}Y^{1:N}}(u_i | u^{1:i-1} y^{1:N})$.

For two random variables (X, Y) distributed as $p(x)p(y|x)$, the Bhattacharya parameter is defined as

$$Z(X|Y) = 2 \sum_y P_Y(y) \sqrt{P_{X|Y}(1|y) P_{X|Y}(0|y)}.$$

Let $\beta < 0.5$ and define the following subsets obtained by polarization, with notation adapted from [15].

$$\mathscr{H}_X = \{i \in [N] : Z(U_i | U^{1:(i-1)}) \geq 1 - 2^{-N^\beta}\}.$$
$$\mathscr{L}_X = \{i \in [N] : Z(U_i | U^{1:(i-1)}) \leq 2^{-N^\beta}\}.$$
$$\mathscr{H}_{X|Y} = \{i \in [N] : Z(U_i | U^{1:(i-1)} Y^{1:N}) \geq 1 - 2^{-N^\beta}\}.$$
$$\mathscr{L}_{X|Y} = \{i \in [N] : Z(U_i | U^{1:(i-1)} Y^{1:N}) \leq 2^{-N^\beta}\}.$$

Note that $\mathscr{L}_X \subseteq \mathscr{L}_{X|Y}$. From Theorem 1 in [24] we have the following results.

$$\lim_{N\to\infty}\frac{1}{N}|\mathscr{H}_X| = H(X), \lim_{N\to\infty}\frac{1}{N}|\mathscr{L}_X| = 1-H(X),$$
$$\lim_{N\to\infty}\frac{1}{N}|\mathscr{H}_{X|Y}| = H(X|Y),$$
$$\lim_{N\to\infty}\frac{1}{N}|\mathscr{L}_{X|Y}| = 1-H(X|Y).$$

We define several other subsets of bit-channels as follows:

$$I = \mathscr{H}_X \cap \mathscr{L}_{X|Y}, F = \mathscr{H}_X \cap \mathscr{L}^c_{X|Y}, R = (\mathscr{H}_X \cup \mathscr{L}_X)^c.$$

We refer to these as good, bad, and not completely polarized bit-channels respectively. We refer to bit-channels in $\mathscr{H}_X$ and bit-channels in $\mathscr{L}_X$ as high-entropy bit-channels and low-entropy bit-channels respectively. The fraction of bit-channels in set R with respect to the block length is vanishing as N increases due to polarization. From [24, Theorem 1],

$$\lim_{N\to\infty}\frac{|I|}{N} = I(X;Y). \tag{1.1}$$

A capacity-achieving polar code design for a binary-input DMC involves providing information bits in set I and frozen bits known to decoder in F.

1.3 Dissertation overview

The dissertation is organised as follows.

In Chapter 2, we provide universal polar code construction for a compound channel containing finite number of asymmetric DMCs avoiding common randomness. A staircase of polar blocks are considered in the code construction. The high-entropy bit-channels in each full-height column of staircase are provided with an MDS codeword encoded from information bits to achieve universality while randomly generated not-completely polarized bits are stored in good

bit-channels of all DMCs of the compound channel via scrambling for each full-height column, which helps avoid common randomness. A thorough probability of decoding error analysis is studied following the scheme. Multi-level cell flash memories are modelled as Truncated support beta-binomial channel model, which is a compound channel containing asymmetric channels [52]. Hence the proposed codes are useful for flash memory error correction in practice.

In Chapter 3, we propose shaping codes for costly noiseless and noisy channels based on polar codes. Costly channel has a cost associated to each symbol of the alphabet. Shaping codes encode information on costly channels subject to an average cost constraint. Our proposed polar codes for costly noiseless channels achieve minimum possible average cost per information bit. The proposed polar shaping codes for costly noisy channels achieves its minimum average cost per information bit ensuring reliable transmission of information by using common randomness. We also alternatively prove that a polar shaping code construction avoiding common randomness also achieves the optimal total cost. In practice, application of shaping codes include data transmission with a power constraint [16], and data storage on flash memories [29] and efficient strand synthesis for DNA-based storage [28].

In Chapter 4, we provide coding scheme for compressing a memoryless uniform source when a side information correlated to source is available at the receiver and correlation is unknown to the source. The proposed scheme involves adaption of universal polar coding schemes for compound channels and use of linear code duality between channel and source coding. The code construction designed based staircase of polar blocks only applies to a class of symmetric conditional distributions of the side information given the source. In wireless sensor networks, individual nodes monitor data related to physical conditions, such as temperature, sound, humidity, wind and forward to central node. The data monitored at any two nodes in such a sensor network can be assumed to have some correlation which is not always fixed and changes. Therefore, data monitored at one node can act as a correlated side information of data monitored at another node if the later node intends to receive information from the former. Hence the proposed schemes are useful for communication between nodes in a wireless sensor network.

In Chapter 5, we propose a scheme based on chaining polar blocks for a 3-receiver broadcast channel when transmitter is required to send a public message to all receivers and private message to receiver-1 reliably. The output at receiver-2 is degraded from the output at receiver-1 for the broadcast channel. The proposed coding scheme essentially translates the ideas of rate-splitting, superposition coding, and indirect decoding achieving all the rate pairs in the capacity region. We also provide the motivation for our interest in considering the problem through a file transfer application in a client-server network.

Chapter 2

Universal Asymmetric Channel Polar Coding without Common Randomness

2.1 Introduction

Arikan [1] constructed capacity-achieving codes for binary-input symmetric channels. A capacity-achieving coding scheme based on source and channel polarization for binary-input asymmetric channels was proposed by Honda and Yamamoto [24]. Following Mondelli et al. [33], we refer to the scheme as the integrated scheme. In this scheme, complex boolean functions are shared between encoder and decoder for non-information carrying bit-channels. The use of common randomness is proposed to avoid these complex boolean functions [24]. En Gad et al. [15] used randomized rounding for low entropy and not-completely polarized bit-channels. In addition, a side channel was used to reliably transmit bits corresponding to not-completely polarized bit-channels, whose fraction is vanishing with respect to the block length. This reduces the storage requirement to polynomial complexity. It was noted in [33] that better simulation results were achieved when an argmax rule was used in place of randomized rounding to encode low-entropy bits. A proof that argmax can be used to encode low-entropy bits is given by Chou and Bloch [11].

A compound channel is a set of discrete-memoryless channels (DMCs), $(\mathscr{X}, \{p_l(y|x) : l \in S\}, \mathscr{Y})$ where $y \in \mathscr{Y}$ for every state l in the set S. The compound channel can be looked at as a DMC with state, where the state is arbitrarily selected and fixed for the transmission

of an entire block. The assumption is that the decoder knows the channel state. Hassani and Urbanke [22], [23] presented two "polar-like" universal coding schemes to achieve rates close to the compound capacity of binary-input symmetric DMCs. Sasoglu and Wang [48] introduced a method to polarize symmetric channels universally. In this paper, we present a universal polar coding scheme for the *asymmetric* setting that achieves the compound capacity of any finite set of binary-input asymmetric channels. We can get such a scheme by a combination of Honda and Yamamoto's polar coding scheme for asymmetric channels [24] and a universal polar coding scheme for symmetric channels proposed by Hassani and Urbanke [22], [23]. But a direct use of Honda and Yamamoto scheme in the universal schemes presented in [22], [23] requires either high-complexity boolean functions or common randomness [24], or a side-channel [15] in its implementation.

Our main technical contribution in this paper is that we give a new coding strategy exploiting the structure of staircase, originally proposed in the universal scheme for symmetric channels [22], that eliminates the need to use storage-intensive shared boolean functions, a separate side channel to transmit bits corresponding to bit-channels that are not-completely polarized, and common randomness. We initially assume a condition on the polar block that the number of the bit-channels which are good for all the DMCs of the compound channel is greater than the number of not-completely polarized bit-channels in order to implement our staircase scheme. The key idea behind the scheme is that we use randomized rounding to encode not-completely polarized bit-channels that lie in a full-height column of the staircase and we store them in the good-bit channel set for all DMCs in the same column so that they are reliably decoded at the decoder.

Our solution addresses the technical challenges to construct the desired coding scheme based on this idea. One of the main challenges of our random code construction method is that the we should ensure that each block in the staircase has the same average distribution as required for a single asymmetric channel code [24], [11] to get a reliable code. To do so, we implement the following novel elements in the code construction. We add a designated

information bit of each full-height column to the encoded not-completely polarized bits of that column and store these resulted bits in the good bit-channels of all DMCs in the same column. To fill the positions in the non-full-height columns on the left, we use the frozen vector, which is generated randomly according to the distribution requirements. Finally, the bit-channels of the non-full-height columns on the right are encoded using randomized rounding with the required distribution. These do not require decoding and are ignored at the decoder. We rigorously prove that following these novel steps satisfies the average distribution requirement and provide the decoding error analysis for the proposed staircase scheme. We also propose a continuous encoding and decoding method for the staircase scheme that saves delay in communication by a factor of the width of the staircase.

If the assumed condition is not true, we propose to make use of another universalization method based on bit-channel combining [22] to produce a hybrid polar block that satisfies the required condition so that we can apply the proposed staircase scheme using such a hybrid block. We refer to this scheme as the hybridized staircase scheme. We define the bit-channels of the hybrid polar block and then we establish the average distribution requirement for each hybrid polar block used in the staircase to get a reliable code. This will give us the conditional distribution requirement of each bit-channel of the hybrid polar block while encoding. Then we provide steps for encoding and decoding by adapting the encoding and decoding methods of the proposed staircase scheme for the original block. We show that the encoding method ensures the average distribution requirement for each hybrid block and provide the rigorous decoding error analysis for the hybridized staircase scheme. We present a new algorithm that efficiently produces a hybrid polar block, satisfying the desired condition, which is at most 2^{s-1} times the original polar block length, where s is the number of DMCs in the compound channel.

In practice, common randomness is implemented by use of pseudo random number generators with a common seed. They are not truly random in nature, and they can suffer from shorter than expected period for weak seed states, correlation of successive values, or lack of uniformity of distribution for large quantities of generated numbers. Our scheme only needs

random number generation at the encoder to implement the randomized rounding rule. We can use truly generated random variables at the encoder side which can be practically realized by hardware random generators like thermal noise and clock drift, for example. Some of these methods may be limited by the rate of random variable generation. However, we note that we need to only produce random variables for not-completely polarized bit-channels, which are diminishing in fraction, and therefore do not require a high rate of random number generation. Moreover, if the rate of random number generation is not sufficient, pseudo random generators with periodically refreshed random keys generated by a hardware random generator can be used. The capacity of some network settings such as an arbitrarily varying channel depends on the availability of common randomness between the encoder and the decoder [14, p.172]. So, implementing a code for a network without common randomness and still achieving the rates that can be achieved with the availability of common randomness is also a theoretically interesting result.

The paper is organized as follows. In Section 2.2, we introduce some notations and recall some background results. In Section 2.3, we review the integrated scheme that achieves the capacity of binary-input asymmetric channels along with the enhancements in [24], [15]. In Section 2.4, we describe our new code construction for the universal polar coding scheme for binary-input asymmetric DMCs. This section highlights the main contribution of the paper. The result is that there exists a sequence of low-complexity universal codes, achieving the capacity of a compound channel comprising a finite set of binary-input asymmetric DMCs without using high-complexity boolean functions or common randomness or a side-channel in its implementation. We also provide a continuous encoding and decoding method for the staircase scheme in this section. In Section 2.5, we provide the hybridized staircase scheme with detailed decoding error analysis and also provide our new algorithm to efficiently produce a hybrid polar block with the desired condition satisfied. In Section 2.6, we describe the code construction in detail, including encoding and decoding methods directly using the hybrid block produced after combining two polar blocks by the universalization technique based on bit-channel combining.

The description of the scheme in this section helps provide a clear view of the order of the bit-channels of the hybrid block produced after combining two blocks. In Section 2.7, we conclude the paper and also pose a few open problems.

2.2 Preliminaries

We denote random variables by upper-case letters, such as X, Y, and their realizations by lower-case letters, such as x, y. We denote the input alphabet of the compound channel by $\mathscr{X} = \{0,1\}$ and the output alphabet by $\mathscr{Y}$. We express any set of random variables $X_i, X_{i+1}, \ldots, X_j$ $(i < j)$ by a row vector $(X_i, X_{i+1}, \ldots, X_j)$ which is denoted by $X^{i:j}$. We denote the set $\{1,2,3,\ldots,N\}$ by $[N]$. We denote the set $\{i, i+1, \cdots, j\}$ by $[i:j]$ $(i < j)$. Let $U^{1:N}$ be a row vector and let $\mathscr{A} \subset [N]$. The row vector consisting of elements in $U^{1:N}$ corresponding to the positions in $\mathscr{A}$ is denoted by $U^{\mathscr{A}}$. We denote the vector $\{U_{j1}, U_{j2}, \ldots, U_{jN}\}$ by an indexed vector $U_j^{1:N}$. We use the abbreviation "w.p." for "with probability". Let P and Q be any two distributions on an arbitrary discrete alphabet $\mathscr{Z}$. We denote the total variation distance between the two distributions P and Q as $||P-Q||$. Therefore $||P-Q|| = \sum_{z \in \mathscr{Z}} \frac{1}{2}|P(z)-Q(z)| = \sum_{z:P(z)>Q(z)} P(z) - Q(z)$.

Let the finite state set of the compound channel be $S = \{1,2,\ldots,s\}$ for some $s \in \mathbb{N}$. We refer to the DMC associated with state $l \in S$ as DMC l. Let $(X_1,Y_1),(X_2,Y_2),\ldots,(X_N,Y_N)$ be independent and identically distributed (i.i.d.) random tuples distributed according to $P_X(x)p_l(y|x)$, where $l \in S$ and $N = 2^n$. Let G_N be the conventional polar transformation [1], represented by a binary matrix of dimension $N \times N$. If $U^{1:N} = X^{1:N}G_N$, then we denote $\mathbb{P}(U^{1:N} = u^{1:N})$ by $P_{U^{1:N}}(u^{1:N})$ and similarly we denote $\mathbb{P}(U_i = u_i | U^{1:i-1}Y^{1:N} = u^{1:i-1}y^{1:N})$ by $P_{U_i|U^{1:i-1}Y^{1:N}}(u_i|u^{1:i-1}y^{1:N})$. For two random variables (X, Y^l) distributed as $P_X(x)p_l(y|x)$, the Bhattacharyya parameter is defined as

$$Z(X|Y^l) = 2\sum_y P_{Y^l}(y)\sqrt{P_{X|Y^l}(1|y)P_{X|Y^l}(0|y)}.$$

Let $0 < \beta < 0.5$ and define the following bit-channel subsets, with notation adapted from [15]

augmented by a subscript l to address the selected DMC in S.

$$\mathscr{H}_X = \{i \in [N] : Z(U_i|U^{1:i-1}) \geq 1 - 2^{-N^\beta}\}.$$
$$\mathscr{L}_X = \{i \in [N] : Z(U_i|U^{1:i-1}) \leq 2^{-N^\beta}\}.$$
$$\mathscr{H}_{(X|Y)_l} = \{i \in [N] : Z(U_i|U^{1:i-1}Y^{1:N}) \geq 1 - 2^{-N^\beta}\}.$$
$$\mathscr{L}_{(X|Y)_l} = \{i \in [N] : Z(U_i|U^{1:i-1}Y^{1:N}) \leq 2^{-N^\beta}\}.$$

Note that $\mathscr{L}_X \subseteq \mathscr{L}_{(X|Y)_l}$, $l \in S$. From [24, Theorem 1], we have the following polarization results:

$$\lim_{N\to\infty} \frac{1}{N}|\mathscr{H}_X| = H(X).$$
$$\lim_{N\to\infty} \frac{1}{N}|\mathscr{L}_X| = 1 - H(X).$$
$$\lim_{N\to\infty} \frac{1}{N}|\mathscr{H}_{(X|Y)_l}| = H(X|Y^l).$$
$$\lim_{N\to\infty} \frac{1}{N}|\mathscr{L}_{(X|Y)_l}| = 1 - H(X|Y^l).$$

We define several other subsets of bit-channels as follows:

$$I_l = \mathscr{H}_X \cap \mathscr{L}_{(X|Y)_l},$$
$$F_l = \mathscr{H}_X \cap \mathscr{L}^c_{(X|Y)_l},$$
$$R = (\mathscr{H}_X \cup \mathscr{L}_X)^c.$$

We refer to these as good, bad, and not-completely polarized bit-channels respectively. We often refer to bit-channels in $\mathscr{H}_X$ as high-entropy bit-channels. We refer to bit-channels in $\mathscr{L}_X$ as deterministic bit-channels or low-entropy bit-channels. The size of set R is a vanishing fraction with respect to the block length as N increases due to polarization. The capacity of a compound

channel is well known [14, p. 170] and is given by

$$C_c = \max_{P_X(x)} \min_{l \in S} I(X;Y^l). \tag{2.1}$$

where (X,Y^l) is distributed as $P_X(x)p_l(y|x)$, whether or not channel state is available to the decoder [14, p. 170]. From [24, Theorem 1], we have

$$\lim_{N\to\infty} \frac{|I_l|}{N} = I(X;Y^l). \tag{2.2}$$

The compound capacity-achieving distribution could be non-uniform even if some of the DMCs in S are binary-input symmetric channels. However, there has to be at least one binary-input asymmetric DMC in S to get a non-uniform capacity-achieving distribution. The scheme that we give in this paper is applied whenever the input distribution is non-uniform.

Example:

Let $S = \{1,2\}$. Let DMC 1 be a Z-channel with cross-over probabilities $p(0|1) = 0.5$, $p(1|1) = 0.5$ and $p(0|0) = 1$. Let DMC 2 be a binary erasure channel with erasure probability 0.5. Let i_l be the mutual information $I(X;Y^l)$. If X is distributed as Bernoulli(α), then mutual information $i_1 = H(Y^1) - H(Y^1|X) = H(\frac{\alpha}{2}) - \alpha$ and mutual information $i_2 = H(X) - H(X|Y^2) = \frac{H(\alpha)}{2}$. The derivative of the mutual information i_1 w.r.t α, that is $\frac{di_1}{d\alpha}$, becomes $\frac{1}{2}\log(\frac{1-\alpha/2}{\alpha/2}) - 1$. By equating the derivative to zero, we get $\alpha = 2/5$. This gives the capacity-achieving distribution for DMC 1 as mutual information is concave in α. At $\alpha = 2/5$, notice that $i_2 = 0.4855$ is greater than $i_1 = 0.3219$. Hence, from equation (2.1), this will also give the compound capacity-achieving distribution for the compound channel. Therefore the capacity of the compound channel is $H(1/5) - 2/5 = 0.322$.

2.3 Integrated polar coding for binary-input asymmetric channels

In this section, we present the capacity-achieving asymmetric channel coding scheme based upon [24], [15], [11], which is used as a building block in our proposed universal polar coding scheme. Let the asymmetric DMC be characterized by $p(y|x)$ and let $p(x)$ be the non-uniform capacity-achieving input distribution. We use the same notations as in Section 2.2 with the substitution of I and F for I_l and F_l, respectively, as we are considering the single channel case. Now we describe the encoding and decoding procedure.

2.3.1 Code construction

We first generate random function $f : F \to \{0,1\}$, where each $f(j)$, $j \in F$, is chosen independently and uniformly. These frozen bits are shared between encoder and decoder.

We also generate independent random boolean functions $\lambda_i : \{0,1\}^{i-1} \to \{0,1\}$ for each $i \in R$ by using the following probability rule:

$$\lambda_i(u^{1:i-1}) = u \text{ w.p. } P_{U_i|U^{1:i-1}}(u|u^{1:i-1}), \text{ for } u \in \{0,1\}$$

independently for each $u^{1:i-1}$. Let the set of random functions be denoted by λ_R. These functions are shared between encoder and decoder, which can require exponential storage complexity. The encoding algorithm is described as follows:

Encoding

Input: uniformly distributed message $M^{1:|I|}$

Output: codeword $X^{1:N}$

for $i = 1 : N$, encode U_i as follows:

1. If $i \in I$, the value of U_i is given by setting $U^I = M^{1:|I|}$.

2. If $i \in F$, we set $U_i = f(i)$.

3. If $i \in \mathscr{L}_X$, we encode U_i using the **argmax rule** [11],

$$U_i = \underset{x\in\{0,1\}}{\operatorname{argmax}} P_{U_i|U^{1:i-1}}(x|U^{1:i-1}). \tag{2.3}$$

4. If $i \in R$, we set $U_i = \lambda_i(U^{1:i-1})$.

end

Transmit $X^{1:N} = U^{1:N}G_N$.

The decoding algorithm is as follows.

Decoding

Input: received vector $Y^{1:N}$

Output: message estimate $\hat{M}^{1:|I|}$

for $i = 1 : N$

1. If $i \in F$, set $\hat{U}_i = f(i)$.

2. If $i \in \mathscr{L}_X \cup I$, set

$$\hat{U}_i = \underset{x\in\{0,1\}}{\operatorname{argmax}} P_{U_i|U^{1:i-1},Y^{1:N}}(x|\hat{U}^{1:i-1}, Y^{1:N}).$$

3. If $i \in R$, set $\hat{U}_i = \lambda_i(\hat{U}^{1:i-1})$.

end

Decode $\hat{M}^{1:|I|} = \hat{U}^I$.

For $i \in \mathscr{L}_X$, the induced conditional distribution $\delta_i(u|u^{1:i-1})$ on U_i given $U^{1:i-1}$ satisfies $\delta_i(u|u^{1:i-1}) = 1$ and $\delta_i(u+1|u^{1:i-1}) = 0$ where

$$u = \underset{x\in\{0,1\}}{\operatorname{argmax}} P_{U_i|U^{1:i-1}}(x|u^{1:i-1}).$$

The ensemble average distribution of $U^{1:N}$ induced by the code construction is as follows:

$$\mathbb{E}_{(\lambda_R,f)}[\mathbb{P}(U^{1:N}=u^{1:N}|(\lambda_R,f))] = 2^{-|\mathscr{H}_X|}\Pi_{i\in R}P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1})\Pi_{i\in\mathscr{L}_X}\delta_i(u_i|u^{1:i-1}).$$

This average distribution is $O(2^{-N^{\beta'}})$ close in total variation distance to the distribution induced when $X^{1:N}$ is an i.i.d. random vector with distribution $p(x)$, for $\beta' < \beta < 0.5$. This makes the decoding method reliable, with average probability of error $\mathbb{E}_{(\lambda_R,f)}[P_e(\lambda_R,f)] = O(2^{-N^{\beta'}})$ [24]. Common randomness, using pseudo random numbers, can be used for encoding and decoding these bit-channels in R [24]. In [15], use of a side-channel is proposed for bit-channels $(\mathscr{H}_X \cup \mathscr{L}_{X|Y})^c$ as an alternative to sharing boolean functions. We chose to use boolean functions or common-randomness to decode all bits in R rather than just to decode bits in $(\mathscr{H}_X \cup \mathscr{L}_{X|Y})^c$, a subset of R, for the purposes of applying this to universal coding that we propose in this paper. The quantities $P_{U_i|U^{1:i-1}}(u|u^{1:i-1})$ and $P_{U_i|U^{1:i-1},Y^{1:N}}(u|u^{1:i-1},y^{1:N})$ used during encoding and decoding can be computed in $O(N\log N)$ real operations [24].

2.4 Universal scheme for asymmetric channels without common randomness

In this section, we provide our new staircase construction for asymmetric compound channels. Let $p(x)$ be the non-uniform compound capacity-achieving distribution for compound channel S.

Let $L = \min\{|I_1|,|I_2|,\ldots,|I_s|\}$. Clearly, $\lim_{N\to\infty}\frac{|I_l|}{N} = i_l$ from equation (2.2), where i_l is the mutual information $I(X;Y^l)$ when DMC l is selected in the compound channel. For any $\varepsilon > 0$, there exists a large enough $N_l(\varepsilon)$, such that $i_l - \varepsilon \leq \frac{|I_l|}{N} \leq i_l + \varepsilon$ for all $N \geq N_l(\varepsilon)$. Therefore we have $\min\{i_1-\varepsilon, i_2-\varepsilon,\ldots,i_s-\varepsilon\} \leq \min\{\frac{|I_1|}{N},\frac{|I_2|}{N},\ldots,\frac{|I_s|}{N}\} \leq \min\{i_1+\varepsilon,i_2+\varepsilon,\ldots,i_s+\varepsilon\}$ for all $N \geq \max\{N_1(\varepsilon),N_2(\varepsilon),\ldots,N_s(\varepsilon)\}$. This implies that $\min\{i_1,i_2,\ldots,i_s\} - \varepsilon \leq \frac{\min\{|I_1|,|I_2|,\ldots,|I_s|\}}{N} \leq \min\{i_1,i_2,\ldots,i_s\}+\varepsilon$ for all $N \geq \max\{N_1(\varepsilon),N_2(\varepsilon),\ldots,N_s(\varepsilon)\}$. Therefore, $\lim_{N\to\infty}\frac{L}{N} = C_c$.

If the inequality $|I_1 \cap I_2 \cap \ldots \cap I_s| \leq L$ is strict, by assigning message bits to indices in

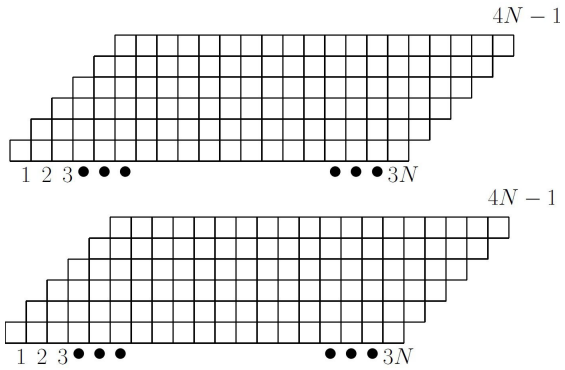

Figure 2.1. Extended staircases with $k = 3, N = 6$ and $p = 2$

$I_1 \cap I_2 \cap \ldots \cap I_s$, assigning uniform random frozen bits to indices in $\mathscr{H}_X - (I_1 \cap I_2 \cap \ldots \cap I_s)$, and using the same coding scheme as in Section 2.3 to encode the bits in other indices, we can get a reliable code, but it will not be capacity-achieving.

As $F_l \cup I_l = \mathscr{H}_X$ for all $l \in S$, for any channels $l, m \in S$, $l \neq m$, a bit-channel which is good for DMC l and not good for DMC m will be a bad bit-channel for DMC m. This fact will enable us to adapt the universal coding scheme for symmetric channels [22] to the asymmetric case and to construct codes that achieve rates close to $\frac{L}{N}$. As in [22], [23], we use a staircase scheme composed of polar blocks to achieve rates close to $\frac{L}{N}$. In our scheme, we exploit the staircase structure to give a new coding strategy that avoids storage-intensive boolean functions, common randomness, and a side-channel for encoding bits in R. To do so, we initially assume $|I_1 \cap I_2 \cap .... \cap I_s| \geq |R|$, an assumption that will be relaxed in Section 2.5.

2.4.1 Code construction

In order to achieve universality, we will require the use of a linear maximum distance separable (MDS) code $\mathscr{M}$ with block length $|\mathscr{H}_X| - |R|$. We let $p \in N$ be the smallest integer for which such a code exists over $GF(2^p)$. We arrange polar blocks of size N, for N sufficiently large, in a staircase with height N. We extend the staircase by placing $k \in \mathbb{N}$ such staircases side-by-side. Now take p such extended staircases, graphically placed one above the other, as illustrated in Fig. 2.1 for the case $N = 6$, $k = 3$, and $p = 2$. In our code construction, while

encoding, we fill $U^{1:N}$s of all the polar blocks column-by-column from left to right, and we follow the same order while decoding. Hence we encode/decode different polar blocks in parallel while encoding/decoding a column. The total number of columns is $(k+1)N-1$, and we label them with indices $1:(k+1)N-1$ from left to right. While encoding, we need to make sure that each of the polar blocks in the staircase has the same ensemble average distribution as in the single asymmetric channel case so that decoding will be reliable.

Before we give the details of our code construction that avoids common randomness and boolean functions, we briefly describe an elementary staircase scheme with non-uniform input distributions which is adapted from the symmetric case [22] by directly using the integrated scheme with common randomness. We need an MDS code with block length $|\mathscr{H}_X|$ in this elementary staircase construction. Let p' be the smallest integer for which such an MDS code exists over $GF(2^{p'})$. For the sake of exposition, let us assume p' to be 1, in which case the number of extended staircases will be 1. In any column, we set any bit-channel not in $\mathscr{H}_X$ according to $P_{U_i|U^{1:i-1}}$. This is done using common randomness shared between encoder and decoder, so the decoder will know these bits precisely. Then, we set bit-channels in $\mathscr{H}_X$ by using the MDS code that encodes L bits into $|\mathscr{H}_X|$ bits if it is a full-height column. Note that no channel information is used in the encoding step. Then, we set bit-channels in $\mathscr{H}_X$ in non-full-height columns according to independent uniform distribution, which is known to the decoder by the common randomness. This encoding method satisfies the ensemble average distribution for each block in the staircase. The decoder has full channel state information, so it knows the channel used. Thus, precisely L indices from $\mathscr{H}_X$ in every full-height column can be decoded with negligible error. Finally, the remaining $|\mathscr{H}_X|-L$ bits are obtained by erasure decoding of the MDS code.

When $p'>1$, we use p' staircases to store the MDS codeword in a full-height column in its binary format. As we store the MDS codeword in the high-entropy bit-channels of a full-height column, we need to make sure that each bit of the codeword in its binary format satisfies a uniform distribution. In the code construction that we propose, we also use the MDS

code over field $GF(2^p)$ and store it in high-entropy bit-channels that are required to satisfy a uniform distribution. Lemma 1 below guarantees that this distribution requirement will be satisfied.

Lemma 1. *Let G be the generator matrix of the linear MDS code $\mathscr{M}$ over $GF(2^p)$. If G does not have a zero column, then we have an equal number of codewords with zero and codewords with one in any given position in the binary representation of $\mathscr{M}$.*

Proof: Let j be any column of G. Since it is non-zero, it has a non-zero entry $g_{ij} \in GF(2^p)$. The jth position of the codeword corresponding to message $[0,\ldots,m_i,\ldots,0]$ will be $m_i g_{ij}$. As m_i ranges over all elements of $GF(2^p)$, $m_i g_{ij}$ also does. Therefore the binary representation of this codeword entry ranges over all possible binary p-tuples. This ensures that for any position in the binary representation of $\mathscr{M}$ there exists a codeword which has the value 1 in that position. Due to linearity of the equivalent binary representation, we must have an equal number of codewords with zero and codewords with one in any given position. □

We now turn to the discussion of our new code construction that avoids the use of both common randomness and high-complexity boolean functions. The construction introduces the following novel elements. We use randomized rounding for not-completely polarized bit-channels in a full-height column and store them in a bit-channel set that is good for all the DMCs. In order to satisfy the distribution for the high-entropy bit-channels, we add a designated information bit in the column to these stored bits. To fill the positions in the non-full-height columns on the left, we use the frozen vector, which is generated randomly according to the distribution requirements. Finally, the bit-channels of the non-full-height columns on the right are encoded using randomized rounding with the required distribution. These do not require decoding and are ignored at the decoder. These encoding steps ensure that the required distribution is satisfied for all of the blocks.

We generate a random frozen vector $W^{1:N}$ such that

$$\mathbb{P}(W^{1:N} = u^{1:N}) = 2^{-|\mathscr{H}_X|}\Pi_{i\in R}P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1})\Pi_{i\in\mathscr{L}_X}\delta_i(u_i|u^{1:i-1}). \tag{2.4}$$

The vector $W^{1:N}$ is shared between encoder and decoder, which is used to fill the non-full-height columns on the left as mentioned. Let I' be a subset of $I_1 \cap I_2 \cap \ldots \cap I_s$ such that $|I'| = |R|$, to store the bits encoded in the not-completely polarized bit-channels of a full-height column. Let $g : I' \to R$ be an arbitrary bijection. Let $L' = L - |I'|$.

Now we are ready to present our code construction illustrating the encoding and decoding schemes in detail.

Encoding

Input: pL' information bits for each full-height column.

Output: $U^{1:N}$, to which we apply G_N to get $X^{1:N}$, of each polar block in the staircase.

- To encode non-full-height columns on the left from $t = 1 : N-1$, we assign $U_i = W_i$ for the block with channel index i in that column. We repeat this for all p staircases. This step ensures that the prefix part, which lies in the non-full-height region, of the polar blocks satisfies the required ensemble average distribution.

- To encode full-height columns from $t = N \leq i \leq kN$:

  - First, encode the blocks with index $i \in \mathscr{L}_X$ in column t using the argmax rule (2.3). Repeat this for all p staircases. This maintains the required conditional distribution for these indices.

  - Second, encode the blocks with index $i \in R$ in column t using the randomized rounding rule, i.e., $U_i = u$ w.p. $P_{U_i|U^{1:i-1}}(u|U^{1:i-1})$ for $u \in \{0,1\}$. Repeat this for all p staircases. This will maintain the required conditional distribution. Since these are randomly generated, we use the inverse function g^{-1} to store these bits in

$I' \subseteq I_1 \cap I_2 \cap \ldots \cap I_s$ where they can be reliably decoded.

– Third, encode the blocks with index $i \in I'$ by assigning $U_i = H \oplus \tilde{U}_{g(i)}$, where $\tilde{U}_{g(i)}$ is the bit copied from the block with index $g(i) \in R$ and H is a designated information bit corresponding to that column. Repeat the same for all p staircases. This maintains the distribution of the indices in $\mathscr{H}_X$ and also ensures the independence from previously encoded bits of the polar block. This is the key step of the construction, since the direct use of $\tilde{U}_{g(i)}$ to encode U_i would not satisfy the required distribution. Let $\overline{P}$ represents the bits encoded in the previous columns. Let $\mathscr{P} = \{\mathbf{z_1}, \mathbf{z_2}, \mathbf{z_3}, \ldots \mathbf{z}_{|\mathscr{P}|}\}$ be the set of possible bit strings that can be encoded in previous columns. The intuition of the scrambling that we did in this step is shown through a picture of the joint distribution on $(\overline{P}, \tilde{U}_{g(i)}, H, U_i)$ given in Fig. 2.2. As the designated information bit H is independent of both $\overline{P}$ and $\tilde{U}_{g(i)}$, for a given pair of $\overline{P}$ and $\tilde{U}_{g(i)}$, the conditional distribution of H will be uniform, which can be noticed in Fig. 2.2. For a given $\overline{P}$ and when $\tilde{U}_{g(i)} = 0$, $U_i = H$. So, the conditional distribution of U_i is uniform, given $\overline{P}$ when $\tilde{U}_{g(i)} = 0$. For a given $\overline{P}$ and when $\tilde{U}_{g(i)} = 1$, $U_i = H + 1$. So, again the conditional distribution of U_i is uniform, given $\overline{P}$ when $\tilde{U}_{g(i)} = 1$. Hence U_i is also uniform and independent of both $\overline{P}$ and $\tilde{U}_{g(i)}$, which can be noticed in Fig. 2.2.

– Fourth, encode the blocks with indices $i \in \mathscr{H}_X - I'$.

  * Encode pL' information bits (equivalent to L' symbols over $GF(2^p)$) into codeword m in the binary representation of $\mathscr{M}$.

  * Fill blocks with indices in $i \in \mathscr{H}_X - I'$ in all p staircases with codeword m as shown in Fig. 2.3. By Lemma 1, a uniform distribution is guaranteed for these positions, as required for indices in $\mathscr{H}_X$. Since m depends only on the information bits of the current column, independence from previously encoded bits of the polar block is also guaranteed.

– The layout of coding a full-height column is shown in Fig. 2.3.

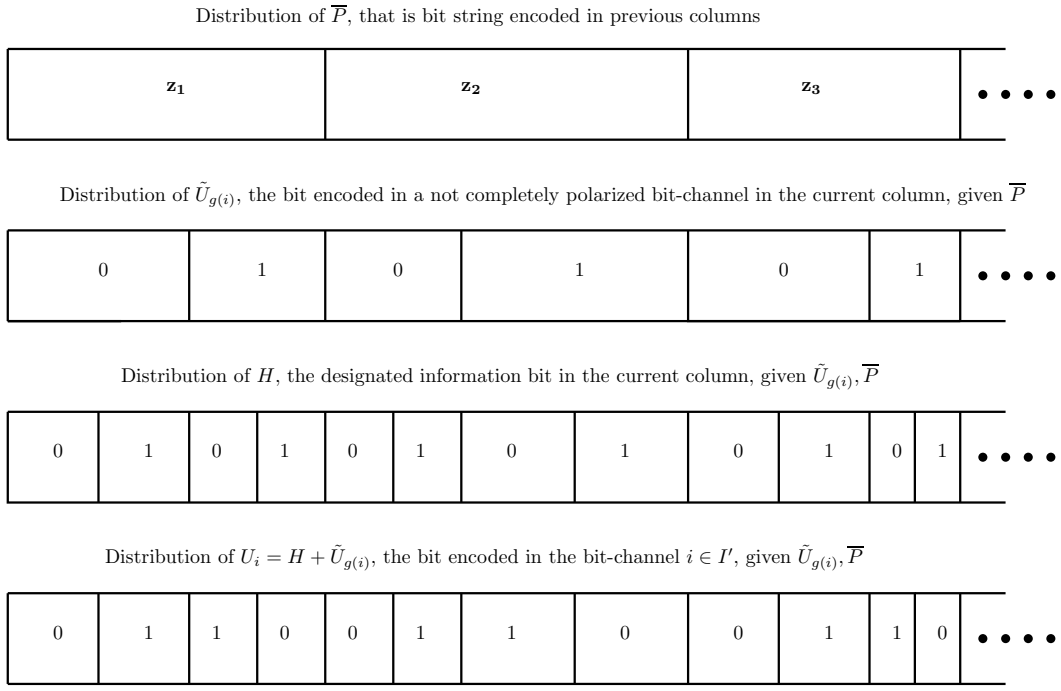

Figure 2.2. Joint distribution of ($\overline{P}$, $\tilde{U}_{g(i)}$, H and U_i) where the width of each symbol scales according to the probability of the occurrence of the symbol.

  - Hence all U_is corresponding to all polar blocks in the column are encoded in all p staircases. This enables the continuation of SC encoding of the polar blocks to encoding U_is corresponding to the next column.

- To encode non-full-height columns $t = kN+1 : (k+1)N-1$ on the right, we generate all bits randomly to satisfy the distribution of the polar block. This is done for all p staircases as follows:

  - For blocks with index $i \in \mathscr{H}_X$, generate U_i independently and uniformly.
  - For blocks with index $i \in R$, generate $U_i = u$ w.p. $P_{U_i|U^{1:i-1}}(u|U^{1:i-1})$, for $u \in \{0,1\}$.
  - For blocks with index $i \in \mathscr{L}_X$ use argmax rule.

- Transmit $X^{1:N} = U^{1:N}G_N$ for each polar block.

Decoding

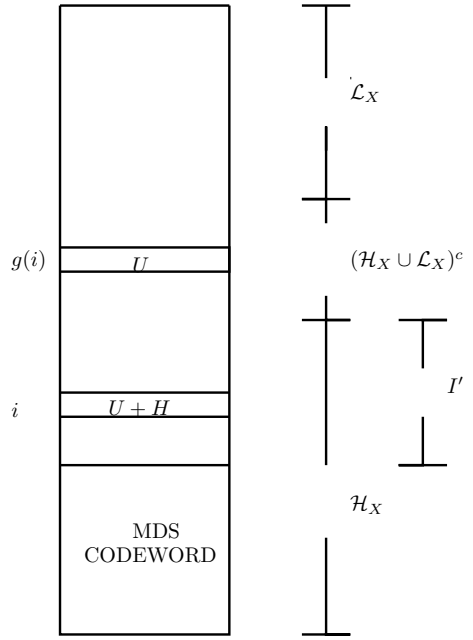

Figure 2.3. Coding a full-height column: H is the designated information bit in the the column and U is the encoded bit in the block with index $g(i)$ in the column.

Input: Received vector $Y^{1:N}$ for each block.

Output: Estimates of encoded information bits.

- To decode non-full-height columns on the left from $t = 1 : N-1$, we recover $\hat{U}_i = W_i$ for the block with channel index i in that column. Repeat this for all p staircases.

- To decode full-height columns from $t = N \leq i \leq kN$:

  - First, decode the blocks with index $i \in \mathscr{L}_X \cup I'$ in column t using the following decision rule:

    $$\hat{U}_i = \underset{x \in \{0,1\}}{\operatorname{argmax}} P_{U_i|U^{1:i-1},Y^{1:N}}(x|\hat{U}^{1:i-1}, Y^{1:N}).$$

    This is possible since these indices are either good for all channels or deterministic. We repeat this for all p staircases.

  - Second, decode the blocks with index in $\mathscr{H}_X - I'$:

    * Decode the L' symbols from the good indices based on the channel that is

selected using the argmax rule above. Let C be the partially recovered codeword.

* The codeword $\hat{m}$ can be recovered from C by erasure decoding since it is an MDS codeword, providing an estimate of pL' information bits corresponding to the column.

– Last, decode blocks with index $i \in R$ by estimating $\hat{U}_i = \hat{H} \oplus \hat{U}'_{g^{-1}(i)}$ where $\hat{U}'_{g^{-1}(i)}$ is the already decoded bit corresponding to the block with index $g^{-1}(i) \in I'$ in the same column and $\hat{H}$ is the recovered information bit which was designated in the column during the encoding procedure. We repeat this for all p staircases.

– Hence all $\hat{U}_i$s corresponding to all polar blocks in the column are decoded in all p staircases. This enables the continuation of SC decoding of the polar blocks to recover $\hat{U}_i$s corresponding to the next column.

- Ignore and do not decode non-full-height columns $t = kN+1 : (k+1)N-1$ on the right. Note that this will not be a problem as information bits have already been fully recovered from full-height columns.

Note that we encoded $L'q$ information bits only in full-height columns. Hence we get the rate $\frac{L'}{N}$ for each full-height column. Since $\frac{|I'|}{N}$ is diminishing, the rate for each such column will be close to $\frac{L}{N}$. Also, as k increases, the full-height columns will constitute a significant fraction of the total block length and the overall rate approaches $\frac{L}{N}$. The exact relation between the achievable rate and k can be found in [22], [23].

We used a linear MDS code in our asymmetric staircase construction. Notice that in the symmetric channel construction, linearity is not required. Now we derive an upper bound on p, which upper bounds the total number of polar blocks in the staircase structure. If we consider a Reed-Solomon (RS) code as the linear MDS code over $GF(2^p)$, the block length of the code should divide $2^p - 1$ [7, p. 174]. We bound p as follows:

- If $|\mathscr{H}_X| - |R|$ is odd:

  By Euler's Theorem, p can take value $\phi(|\mathscr{H}_X| - |R|)$ where ϕ is Euler's totient function. Therefore $p \leq \phi(|\mathscr{H}_X| - |R|) \leq |\mathscr{H}_X| - |R| \leq N$.

- If $|\mathscr{H}_X| - |R|$ is even:

  Use a RS code of block length $(|\mathscr{H}_X| - |R|) - 1$. Then $p \leq N$ since the block length is odd. Fill the remaining position with the parity of the information bits to maintain the required distribution of the high-entropy bit-channel in all p staircases and modify the scheme accordingly.

The following theorem computes the ensemble average distribution of each polar block, the overall decoding probability of error and the encoding/decoding complexity of the scheme, in detail.

Theorem 1.

1. For every polar block encoded in the staircase

$$\mathbb{E}_{W^{1:N}}[\mathbb{P}(U^{1:N} = u^{1:N}|W^{1:N})] = 2^{-|\mathscr{H}_X|}\Pi_{i\in R}P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1})\Pi_{i\in\mathscr{L}_X}\delta_i(u_i|u^{1:i-1}).$$

2. Let $P_{e,l}(W^{1:N})$ be the probability of error when DMC l is selected in S for a given code in the above random code construction. The average probability of error is $\mathbb{E}_{W^{1:N}}[P_{e,l}(W^{1:N})] = O(Npk2^{-N^{\beta'}})$ for $\beta' < \beta < 0.5$ for each $l \in S$.

3. Encoding and decoding take $O((\log_2 N)p^{\log_2 3-1})$ and $O((\log_2 N)^2(\log_2(\log_2 N))p^{\log_2 3-1})$ binary operations per bit, respectively. Encoding and decoding also take $O(\log_2 N)$ real operations per bit.

Proof: Refer to the Appendix.

2.4.2 Existence of universal code with high probability

Theorem 1 states that the average probability of error over the random ensemble is $O(Npk2^{-N^{\beta'}})$ for any DMC that gets selected in S. We now show the existence of universal codes with high probability in the random ensemble of codes. Let K be a positive constant. By the Markov inequality, we get the following:

$$\mathbb{P}_{W^{1:N}}(P_{e,l} > Ks\mathbb{E}_{W^{1:N}}[P_{e,l}(W^{1:N})]) < 1/Ks.$$

By using the union bound, we get the following:

$$\mathbb{P}_{W^{1:N}}(\cup_{l\in S}(P_{e,l} > Ks\mathbb{E}_{W^{1:N}}[P_{e,l}(W^{1:N})])) < 1/K.$$

Therefore by taking the probability of the complementary event, we get

$$\mathbb{P}_{W^{1:N}}(\cap_{l\in S}(P_{e,l} \leq Ks\mathbb{E}_{W^{1:N}}[P_{e,l}(W^{1:N})])) \geq 1 - 1/K.$$

By substituting $K = N$ in the above equation, we get that universal codes that have the probability of error $O(sN^2qk2^{-N^{\beta}})$ exist with high probability $1 - 1/N$ in the random ensemble of codes. This kind of analysis is not needed in the symmetric channel case as we have an explicit universal code construction without randomization.

2.4.3 Application to single asymmetric channel

Note that the staircase alignment of blocks played an essential role in satisfying the distribution requirement of each block without common randomness or complex boolean functions. Hence we can use the staircase scheme for capacity-achieving single asymmetric channel code construction without common randomness or complex boolean functions. Note that we do not require to use an MDS code for each full-height column for the single channel case. Hence we

will only need one extended staircase in this case. In fact, our staircase scheme proposed is an appropriate solution to achieving asymmetric channel capacity without common randomness, complex boolean functions, or a side channel, independent of the compound channel setting.

For the single asymmetric channel case, we can alternatively use a chaining construction to avoid complex boolean functions or common randomness. While encoding, we copy the not-completely polarized bit-channels of a polar block into good bit-channels of its successive block in the chaining construction. We add randomly chosen uniform i.i.d. bits that are known to the decoder in advance to these not-completely polarized bits before copying them to the successive block to satisfy the distribution requirement. The last polar block of the chaining construction uses polarization with uniform distribution so that the block does not have not-completely polarized bit-channels. This will enable the decoder to recover blocks in reverse order while decoding. This is similar to the multi-coding implementation without the degraded assumption [15, construction 16]. Polar code construction with uniform input distribution is needed in case of chaining, whereas it is not needed in the staircase construction.

Gallager's scheme [24] uses alphabet extension, which leads to the asymptotic complexity $O((\Delta I)^{-0.5} N \log N)$ where ΔI is the gap of the achievable rate from the capacity of the asymmetric channel. In our staircase scheme and in the chaining construction, the asymptotic complexity does not depend on the gap of the achievable rate from the capacity as the two schemes are based on the integrated scheme [24] that already avoids alphabet extension.

2.4.4 Continuous encoding and decoding for staircase scheme

We have k sub-staircases adjoined to form the extended staircase, the block length of which becomes the overall block length. We need to increase the width of the staircase k to infinity to achieve rates arbitrary close to $\frac{L}{N}$. Such a large k contributes to a very large block length, leading to a large delay on the order of the block length. We optimize the delay in communication by overcoming the width factor k by continuous encoding and decoding of sub-staircases sequentially. For the sake of exposition, we consider p to be 1. This idea is general

and applied to both symmetric and asymmetric staircase schemes. We index such sub-staircases as $m = 1, 2, \cdots, k$ from left to right. Now we provide steps for the continuous encoding and decoding of sub-staircases.

- Set $j = 1$. We encode U_is of columns $t = 1 : 2N - 1$. We will now have $U^{1:N}$ ready for all blocks in sub-staircase $m = 1$. So we can apply the polar transform to $U^{1:N}$ to get $X^{1:N}$ for all the blocks in sub-staircase $m = 1$. We transmit $X^{1:N}$ corresponding to the polar blocks in sub-staircase $m = 1$. The decoder receives vectors $Y^{1:N}$ corresponding to the polar blocks in sub-staircase $m = 1$. Now we can decode $\hat{U}_i$s for columns $t = 1 : N$. Increment j by 1.

- Now we encode U_is of the next N columns $t = jN : (j+1)N - 1$. We will then have $U^{1:N}$ ready for all blocks in sub-staircase $m = j$ as in the above step. We then transmit $X^{1:N} = U^{1:N} G_N$ and decoder receives $Y^{1:N}$ corresponding to the polar blocks in sub-staircase $m = j$. Now we will be able to decode $\hat{U}_i$s for columns $t = (j-1)N + 1 : jN$. Once we do that, we will have $\hat{U}^{1:N}$ ready for all the blocks in the sub-staircase $m = j - 1$. Then we apply polar transform to $\hat{U}^{1:N}$ to get $\hat{X}^{1:N}$ corresponding to the polar blocks in sub-staircase $j - 1$. Increment j by 1.

- We repeat the above step until $j = k - 1$ sequentially.

- Finally we finish encoding U_is until the last column. We will then have $U^{1:N}$ ready for all blocks in the last sub-staircase $m = k$. We then transmit $X^{1:N} = U^{1:N} G_N$ and the decoder receives $Y^{1:N}$ corresponding to the polar blocks in sub-staircase $m = k$. Now we can decode $\hat{U}_i$s until the last column. We will then have $\hat{U}^{1:N}$ ready for all the blocks in remaining sub-staircases $m = k - 1$ and $m = k$. Then recover the $\hat{X}^{1:N}$ corresponding to the polar blocks in sub-staircases $k - 1$ and k.

Fig. 2.4 illustrates these sequential steps of continuous encoding and decoding of sub-staircases for $N = 3, k = 5$ and $p = 1$.

Remember that we do not decode $\hat{U}_i$s of non-full-height columns on the right when applying the proposed continuous encoding and decoding method to our asymmetric staircase scheme. Hence we do not have complete $\hat{U}^{1:N}$ decoded for the last sub-staircase k. So we can recover $\hat{X}^{1:N}$ only for the blocks until sub-staircase $k-1$.

2.5 Hybridized staircase scheme

If the required condition $|I_1 \cap I_2 \cap \ldots \cap I_s| \geq |R|$ does not hold, we can use the universalizing procedure based on bit-channel combining [22] to produce a partially universalized block that satisfies the desired condition. We propose to apply the staircase scheme using such a hybrid polar block. Now we discuss the idea of the universalizing procedure based on bit-channel combining [22] while applying it to the asymmetric case.

2.5.1 Idea of universal method based on bit-channel combining

The idea of the universalizing method can be explained by considering two independent polar blocks. Let S_1 and S_2 be two non-intersecting subsets of S. If we combine (standard polar combining operation [1]) a bit-channel of the first polar block which is good for all DMCs in S_1 and bad for at least one DMC in S_2 with a bit-channel of the second polar block that is good for all DMCs in S_2 and bad for at least one DMC in S_1, then we get two new bit-channels. This combining of bit-channels governs a new order of decoding for the combined polar block. At this point, we get one new bit-channel that is good for all DMCs in $S_1 \cup S_2$ of the combined polar block. We validate this fact shortly when we provide the probability of error analysis for the hybridized staircase scheme described in this section. Note that a good/bad bit-channel in an updated block means that its Bhattacharyya parameter defined with the received vector given is low/high under the distribution that is induced when the codeword components of the original blocks involved are i.i.d. according to the non-uniform input distribution we are working with.

We can combine many such bit-channels at a time with these two independent polar blocks to achieve universalization. We consider bit-channel sets $\mathscr{A}$, which has bit-channels that

are good for all DMCs in S_1 and bad for at least one DMC in S_2, and $\mathscr{B}$, which has bit-channels that are good for all DMCs in S_2 and bad for at least one DMC in S_1. As already explained, if we combine a bit channel in $\mathscr{A}$ of the first polar block, say x, with a bit channel in $\mathscr{B}$ of the second polar block, say y, then we get two new bit channels, one of which is good for all DMCs in $S_1 \cup S_2$. In a polar block, a later bit-channel output has the previous bit-channel input as one of its components. So a next valid bit-channel combining could be any bit-channel later than x in $\mathscr{A}$ of the first polar block with any bit-channel later than y in $\mathscr{B}$. Hence a best way to combine bit-channels $\mathscr{A}$ of the first block with bit-channels $\mathscr{B}$ of the second block is to combine them in order without missing any bit-channels in between. Let $\mathscr{A} = \{x_1, x_2, \ldots, x_{|\mathscr{A}|}\}$ where $x_{j_1} < x_{j_2}$ for $j_1 < j_2$, $\mathscr{B} = \{y_1, y_2, \ldots, y_{|\mathscr{B}|}\}$ where $y_{j_1} < y_{j_2}$ for $j_1 < j_2$ and $G = \min\{|\mathscr{A}|, |\mathscr{B}|\}$. We do bit-channel combinings x_j with y_j for each $1 \leq j \leq G$ as shown in Fig. 2.5. Now we get G new bit-channels that are good for all DMCs in $S_1 \cup S_2$ in the combined block. These combinings of bit-channels will create a specific order of decoding of bit-channels in the resultant block produced. The description of the scheme in Section 2.6 will help provide a clear view of the order of bit-channels in the block that is generated after combining two polar blocks.

We can consider two such universalized blocks produced in this manner and apply this procedure again with any two non-intersecting subsets of S. This procedure can be done recursively multiple times. So a block obtained after t steps will be of size $N \cdot 2^t$.

We are combining bit-channels in $\mathscr{H}_X$ in the first step of the recursive procedure to produce new bit-channels, which can be either good or bad for a DMC in S. We keep combining these bit-channels in the recursive procedure while we are leaving the low entropy bit-channels $\mathscr{L}_X$ and not-completely polarized bit-channels R of the original polar blocks in the recursive procedure as is. So we have good bit-channels and bad bit-channels defined for each DMC in S for the hybrid polar block. We also have low entropy bit-channels as well as not-completely polarized bit-channels for the hybrid block, which are from the original blocks involved in the hybrid polar block. Now we propose to apply our staircase scheme in Section 2.4 by using the hybrid polar block with the desired condition instead of original polar block. The encoding and

decoding procedure of the staircase scheme with the hybrid block will be identical to our staircase scheme with the original polar block presented in Section 2.4. We refer to such a staircase scheme as a hybridized staircase scheme, which is described in the following subsection.

2.5.2 Code construction for hybridized staircase scheme

We now define the vectors associated with a hybrid polar block generated after t steps. Let $\{X_j^{1:N}\}_{j=1}^{2^t}$ be the codeword component vectors of length N corresponding to each of the original polar blocks indexed $j = 1,2,\ldots,2^t$ in the hybrid polar block and $\{Y_j^{1:N}\}_{j=1}^{2^t}$ be the corresponding received word components when passed through DMC l selected in S. Let $U_j^{1:N} = X_j^{1:N} G_N$, for $j = 1,2,\ldots,2^t$. So $\{U_j^{1:N}\}_{j=1}^{2^t}$ are bit-channel vectors of length N of each original polar block in the hybrid polar block. Let $\{U_j'^{1:N}\}_{j=1}^{2^t}$ be the bit-channel vectors of the hybrid polar block generated after the recursive combining of the original blocks whose bit-channel vectors are $\{U_j^{1:N}\}_{j=1}^{2^t}$. Let $U'^{1:N\cdot 2^t}$ be the permutation of $\{U_j'^{1:N}\}_{j=1}^{2^t}$, according to the order of bit-channels in the hybrid block, governed by the recursive combining procedure. Let the bijective transform that transforms $U'^{1:N\cdot 2^t}$ to $\{U_j^{1:N}\}_{j=1}^{2^t}$ be $\mathscr{H}$.

We refer to the union of good bit-channel set and bad bit-channel set of the hybrid polar block (of any DMC in S) as the high entropy set of the hybrid polar block. We denote these high-entropy bit-channels of the hybrid polar block generated after t steps in the recursive procedure by $\mathscr{H}_{X_t}$. We use the notation I_i^t for DMC i's good bit-channel set of the hybrid polar block. So the DMC i's bad bit-channel set of the hybrid block will be $\mathscr{H}_{X_t} - I_i^t$. We denote the low entropy bit-channel set of the hybrid polar block by $\mathscr{L}_{X_t}$. We denote the not-completely polarized bit-channel set of the hybrid polar block by R_t.

While encoding, we need to ensure that, for each hybrid polar block in the staircase, the original polar blocks involved in the hybrid block have the same ensemble average distribution as in the single asymmetric channel case and also these original blocks involved are independent. This defines an average distribution requirement for each hybrid polar block in the staircase. Let Q be the measure on the hybrid polar block, which is according to the average distribution

requirement of each hybrid block in the staircase. Therefore,

$$Q_{\{U_j^{1:N}\}_{j=1}^{2^t}}(\{u_j^{1:N}\}_{j=1}^{2^t}) = \Pi_{j=1}^{2^t}(2^{-|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}\delta_i(u_{ji}|u_j^{1:i-1})\Pi_{i\in R}P_{U_i|U^{1:i-1}}(u_{ji}|u_j^{1:i-1})).$$

This implies that

$$Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t}) = \Pi_{j=1}^{2^t}(2^{-|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}\delta_i(u_{ji}|u_j^{1:i-1})\Pi_{i\in R}P_{U_i|U^{1:i-1}}(u_{ji}|u_j^{1:i-1})). \tag{2.5}$$

where $u'^{1:N\cdot 2^t}$ is obtained by applying the bijective transform $\mathscr{H}^{-1}$ to $\{u_j^{1:N}\}_{j=1}^{2^t}$.

So, to fill the non-full-height columns in the left, we generate $W'^{1:N\cdot 2^t}$ randomly according to measure $Q_{U'^{1:N\cdot 2^t}}$ in the code construction as follows:

$$\mathbb{P}(W'^{1:N\cdot 2^t} = u'^{1:N\cdot 2^t}) = Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t}). \tag{2.6}$$

We now compute $Q_{U'_{i'}|U'^{1:i'-1}}(u'_{i'}|u'^{1:i'-1})$ for each $i' \in [N\cdot 2^t]$.

For $i' \in \mathscr{L}_{X_t}$, there exist $i \in \mathscr{L}_X$, $j \in [2^t]$ such that $U'_i = U_{ji}$. Then,

$$\begin{aligned} Q_{U'_{i'}|U'^{1:i'-1}}(u'_{i'}|u'^{1:i'-1}) &= \frac{Q_{U'^{1:i'-1}}(u'^{1:i'-1})}{Q_{U'^{1:i'}}(u'^{1:i'})} \\ &= \frac{\sum_{u'^{i'}:N\cdot 2^t} Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t})}{\sum_{u'^{i'+1:N\cdot 2^t}} Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t})} \overset{(a)}{=} \delta_i(u_{ji}|u_j^{1:i-1}). \end{aligned}$$

Identity (a) follows by substituting (2.5) in both numerator and denominator. Note that $u_j^{1:i-1}$ is a function of $u'^{1:i'-1}$. So, to satisfy this conditional distribution, we use argmax rule for a bit-channel $i' \in \mathscr{L}_{X_t}$ that lies either in the full-height column regime or a non-full-height column on the right side of a hybrid polar block in the staircase.

For $i' \in R_t$, there exist $i \in R$, $j \in [2^t]$ such that $U'_i = U_{ji}$. Then,

$$Q_{U'_{i'}|U'^{1:i'-1}}(u'_{i'}|u'^{1:i'-1}) = \frac{Q_{U'^{1:i'-1}}(u'^{1:i'-1})}{Q_{U'^{1:i'}}(u'^{1:i'})}$$
$$= \frac{\sum_{u'^{i':N\cdot 2^t}} Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t})}{\sum_{u'^{i'+1:N\cdot 2^t}} Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t})} \overset{(a)}{=} P_{U_i|U^{1:i-1}}(u_{ji}|u_j^{1:i-1}).$$

Identity (a) follows by substituting (2.5) in both numerator and denominator. Note that $u_j^{1:i-1}$ is a function of $u'^{1:i'-1}$. So, to satisfy this conditional distribution, we use the randomized rounding rule for a bit-channel $i' \in R_t$ that lies either in the full-height column regime or a non-full-height column on the right side of a hybrid polar block in the staircase.

For $i' \in \mathscr{H}_{X_t}$,

$$Q_{U'_{i'}|U'^{1:i'-1}}(u'_{i'}|u'^{1:i'-1}) = \frac{Q_{U'^{1:i'-1}}(u'^{1:i'-1})}{Q_{U'^{1:i'}}(u'^{1:i'})}$$
$$= \frac{\sum_{u'^{i':N\cdot 2^t}} Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t})}{\sum_{u'^{i'+1:N\cdot 2^t}} Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t})} \overset{(a)}{=} 0.5.$$

Identity (a) follows by substituting (2.5) in both numerator and denominator. The conditional distribution of $U'_{i'}$ given $U'^{1:i'-1}$ is always uniform, which also means $U'_{i'}$ is independent of $U'^{1:i'-1}$. Hence, for a bit-channel $i' \in \mathscr{H}_{X_t}$ that lies either in the full-height column regime or a non-full-height column on the right side, we use the same encoding method as we use for high entropy bit-channels in the staircase scheme with the original block, presented in Section 2.4. This maintains the distribution requirement for high-entropy bit-channels in $\mathscr{H}_{X_t}$ that lie in the full-height column regime or the non-full-height column regime on the right side, of a hybrid polar block in the staircase.

We then use an appropriate decoding method, corresponding to encoding method. The decoding rule for the bit-channels $i' \in I_l^t \cup \mathscr{L}_{X_t}$ that lie in the full-height columns will be as follows:

$$\hat{U}'_{i'} = \underset{x\in\{0,1\}}{\operatorname{argmax}} P_{U'_{i'}|U'^{1:i'-1},\{Y_j^{1:N}\}_{j=1}^{2^t}}(x|\hat{U}'^{1:i'-1}, \{Y_j^{1:N}\}_{j=1}^{2^t}).$$

We now provide the probability of decoding error analysis and show that average probability of error diminishes as block length grows for the proposed hybridized staircase scheme.

2.5.3 Probability of decoding error analysis for hybridized staircase scheme

We provide the probability of decoding error analysis for the hybridized staircase scheme with the hybrid block produced after t steps in the recursive procedure. We now compute the ensemble average distribution for each of the hybrid polar blocks used in the staircase.

For a hybrid block that lies completely in the full-height column regime:

For a bit-channel $i' \in \mathscr{H}_{X_t}$, $\mathbb{P}(U'_{i'} = u'_{i'} | W'^{1:N\cdot 2^t} U'^{1:i'-1} = u'^{1:i'-1}) = 0.5$, because we use the same encoding rule as we use for high-entropy bit-channels in the staircase scheme with the original block. It is shown in Theorem 1 that the encoded bit in a high-entropy bit-channel will be uniform and is independent of bits encoded in the previous columns. So we get the same conditional distribution here as well. Therefore, $\mathbb{P}(U'_{i'} = u'_{i'} | W'^{1:N\cdot 2^t} U'^{1:i'-1} = u'^{1:i'-1}) = 0.5 = Q_{U'_{i'}|U'^{1:i'-1}}(u'_{i'}|u'^{1:i'-1})$.

For a bit-channel $i' \in \mathscr{L}_{X_t} \cup R_t$,

$$\mathbb{P}(U'_{i'} = u'_{i'} | W'^{1:N\cdot 2^t} U'^{1:i'-1} = u'^{1:i'-1}) = Q_{U'_{i'}|U'^{1:i'-1}}(u'_{i'}|u'^{1:i'-1}),$$

as described in the hybridized staircase scheme.

By the chain rule of conditional probability, we get

$$\begin{aligned}
\mathbb{P}(U'^{1:N\cdot 2^t} = u'^{1:N\cdot 2^t} | W'^{1:N\cdot 2^t}) &= \Pi_{i' \in [N\cdot 2^t]} \mathbb{P}(U'_{i'} = u'_{i'} | W'^{1:N\cdot 2^t} U'^{1:i'-1}) \\
&= \Pi_{i' \in [N\cdot 2^t]} Q_{U'_{i'}|U'^{1:i'-1}}(u'_{i'}|u'^{1:i'-1}) \\
&= Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t}).
\end{aligned}$$

By taking expectations on both the sides, we get

$$\mathbb{E}_{W'^{1:N\cdot 2^t}}[\mathbb{P}(U'^{1:N\cdot 2^t} = u'^{1:N\cdot 2^t} | W'^{1:N\cdot 2^t})] = Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t}).$$

For a hybrid block that lies partly in the full-height column regime and partly in the non-full-height column regime on the right:

For a high entropy bit-channel $i' \in \mathscr{H}_{X_t}$ that lies in the non-full-height column regime on the right, we generate an independent uniform random variable. For the low entropy bit-channels and not-completely polarized bit-channels that lie in the non-full-height regime on the right, we use the same rule as in the case where they lie in the full-height column regime. So, for the bit-channels that lie in the non-full-height column regime on the right, we get same conditional distribution as in the previous case where they lie in the full-height column regime.

By the chain rule of conditional probability, we get

$$\begin{aligned}
\mathbb{P}(U'^{1:N\cdot 2^t} = u'^{1:N\cdot 2^t} | W'^{1:N\cdot 2^t}) &= \Pi_{i' \in [N\cdot 2^t]} \mathbb{P}(U'_{i'} = u'_{i'} | W'^{1:N\cdot 2^t} U'^{1:i'-1}) \\
&= \Pi_{i' \in [N\cdot 2^t]} Q_{U'_{i'}|U'^{1:i'-1}}(u'_{i'} | u'^{1:i'-1}) \\
&= Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t}).
\end{aligned}$$

By taking expectations on both the sides, we get

$$\mathbb{E}_{W'^{1:N\cdot 2^t}}[\mathbb{P}(U'^{1:N\cdot 2^t} = u'^{1:N\cdot 2^t} | W'^{1:N\cdot 2^t})] = Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t}).$$

For a hybrid block that lies partly in the full-height column regime and partly in the non-full-height column regime on the left:

Suppose that bit-channels $\{1 : m\}$ of the block lie in the non-full-height column regime on the left side and the remaining bit-channels $\{m+1 : N \cdot 2^t\}$ lie in the full-height column

regime. Then, by the code construction and by the chain rule of conditional probability,

$$\begin{aligned}&\mathbb{P}(U'^{1:N\cdot 2^t} = u'^{1:N\cdot 2^t} | W'^{1:N\cdot 2^t}) \\ &\qquad = \mathbb{1}(\cap_{i'=1}^{m}(u'_{i'} = W'_{i'}))\Pi_{i'=m+1:N\cdot 2^t}\mathbb{P}(U'_{i'} = u'_{i'} | W'^{1:N\cdot 2^t} U'^{1:m-1} = u'^{1:m-1}).\end{aligned}$$

By taking expectation on both sides and by the linearity of expectation, we get

$$\begin{aligned}&\mathbb{E}_{W'^{1:N\cdot 2^t}}[\mathbb{P}(U'^{1:N\cdot 2^t} = u'^{1:N\cdot 2^t} | W'^{1:N\cdot 2^t})] \\ &\qquad = \mathbb{E}[\mathbb{1}(\cap_{i'=1}^{m}(u'_{i'} = W'_{i'}))]\Pi_{i'=m+1}^{N\cdot 2^t}\mathbb{P}(U'_{i'} = u'_{i'} | W'^{1:N\cdot 2^t} U'^{1:m-1} = u'^{1:m-1}) \qquad (2.7) \\ &\qquad \overset{(a)}{=} \mathbb{E}[\mathbb{1}(\cap_{i'=1}^{m}(u'_{i'} = W'_{i'}))]\Pi_{i'=m+1}^{N\cdot 2^t} Q_{U'_{i'}|U'^{1:i'-1}}(u'_{i'} | u'^{1:i'-1}).\end{aligned}$$

Identity (a) follows as the conditional probabilities of $U'_{i'}$ given $U'^{1:i'-1}$ for the bit-channels $\{m+1 : N\cdot 2^t\}$ in the full-height column are according to the measure $Q_{U'^{1:N\cdot 2^t}}$.

We evaluate $\mathbb{E}_{W'^{1:N\cdot 2^t}}[\mathbb{1}(\cap_{i'=1}^{m}(u'_{i'} = W'_{i'}))]$ as below:

$$\begin{aligned}\mathbb{E}_{W'^{1:N\cdot 2^t}}[\mathbb{1}(\cap_{i'=1}^{m}(u'_{i'} = W'_{i'}))] &= \mathbb{E}_{W'^{1:N\cdot 2^t}}[\sum_{u'_{i'}\in\{0,1\}:i\in\{m+1:N\cdot 2^t\}} \mathbb{1}(\cap_{i\in[N\cdot 2^t]}(u'_{i'} = W'_{i'}))] \\ &\overset{(a)}{=} \sum_{u'_{i'}\in\{0,1\}:i\in\{m+1:N\cdot 2^t\}} \mathbb{E}_{W^{1:N}}[\mathbb{1}(\cap_{i\in[N\cdot 2^t]}(u'_{i'} = W'_{i'}))] \\ &= \sum_{u'_{i'}\in\{0,1\}:i\in\{m+1:N\cdot 2^t\}} \mathbb{P}(W'^{1:N\cdot 2^t} = u'^{1:N\cdot 2^t}) \\ &\overset{(b)}{=} \sum_{u'_{i'}\in\{0,1\}:i\in\{m+1:N\cdot 2^t\}} Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t}) \\ &= Q_{U'^{1:m}}(u'^{1:m}).\end{aligned}$$

Identity (a) follows by linearity of expectation and identity (b) follows from equation (2.6). Substituting this back into equation (2.7), we get

$$\mathbb{E}_{W'^{1:N\cdot 2^t}}[\mathbb{P}(U'^{1:N\cdot 2^t} = u'^{1:N\cdot 2^t} | W'^{1:N\cdot 2^t})] = Q_{U'^{1:m}}(u'^{1:m})\Pi_{i'=m+1}^{N\cdot 2^t} Q_{U'_{i'}|U'^{1:i'-1}}(u'_{i'} | u'^{1:i'-1})$$

$$= Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t}).$$

We have shown that the ensemble average distribution of each of the hybrid polar blocks in the staircase is according to measure $Q_{U'^{1:N\cdot 2^t}}$. Now we provide lemmas and propositions which will be used in the probability of decoding error analysis.

Lemma 2. *Let $P^j_{X,Y}(x,y)$ be a joint distribution on (X,Y) supported on $\mathscr{X}\times\mathscr{Y}$ for each $j\in\mathscr{J}$. Let $Q(j)$ be the distribution on $\mathscr{J}$. Define $P_{X,Y}(x,y)=\sum_{j\in\mathscr{J}}Q(j)P^j_{X,Y}(x,y)$. Then $Z(X|Y)\geq\sum_{j\in\mathscr{J}}Q(j)Z^j(X|Y)$ where $Z^j(X|Y)=2\sum_{y\in\mathscr{Y}}\sqrt{P^j_{X,Y}(0,y)P^j_{X,Y}(1,y)}$.*

Proof: Refer to the Appendix.

Lemma 2 is used in the proof of the following proposition.

Proposition 1. *Let (X_1,Y_1) and (X_2,Y_2) be independent random variable pairs which may not be identically distributed. X_1 and X_2 are defined over $\mathscr{X}=\{0,1\}$, where Y_1 and Y_2 are distributed over alphabets $\mathscr{Y}_1$ and $\mathscr{Y}_2$. Let $U_1=X_1+X_2$ and $U_2=X_2$. Then*
1. $Z(U_1|Y_1Y_2)\leq Z(X_1|Y_1)+Z(X_2|Y_2)$ and $Z(U_1|Y_1Y_2)\geq\max\{Z(X_1|Y_1),Z(X_2|Y_2)\}$.
2. $Z(U_2|U_1Y_1Y_2)=Z(X_1|Y_1)Z(X_2|Y_2)$ and hence $Z(U_2|U_1Y_1Y_2)\leq\min\{Z(X_1|Y_1),\quad Z(X_2|Y_2)\}$.

Proof: Refer to the Appendix.

Lemma 3. *Let $P(x_1,x_2)=P_1(x_1)P_2(x_2)$ and $Q(x_1,x_2)=Q_1(x_1)Q_2(x_2)$ be two joint distributions on random variables X_1 and X_2 so that the random variables are mutually independent over both the joint distributions P and Q. The marginals of random variable X_i will be $P_i(x_i)$ and $Q_i(x_i)$ over the distributions P and Q, respectively, for $i=1,2$. Assume that the total variation distance between P_i and Q_i is ε_i, for $i=1,2$. Then the total variation distance between the distributions P and Q is at most $\varepsilon_1+\varepsilon_2$.*

Proof: Refer to the Appendix.

Lemma 4. *Let the (X,Y) random variable pair have two measures defined as $Q_{X,Y}(x,y)=Q_X(x)p(y|x)$ and $P_{X,Y}(x,y)=P_X(x)p(y|x)$, respectively. So the conditional distributions $Q_{Y|X}(y|x)$*

and $P_{Y|X}(y|x)$ *are both equal to* $p(y|x)$*. The total variation between the joint distributions* $||Q_{X,Y} - P_{X,Y}||$ *becomes* $||Q_X - P_X||$*.*

Proof: Refer to the Appendix.

The analysis of probability of decoding error is given in the following steps:

- We index each hybrid polar block in the staircase as $b = 1, 2, \ldots, (2^t Npk)$. Let P be the measure on a hybrid polar block induced when $X_j^{1:N}$ is i.i.d. distributed according to $p(x)$ for each j and the vectors $\{X_j^{1:N}\}$, $\{X_k^{1:N}\}$ are independent for $j \neq k$.

- Let $\mathscr{E}_{i'b}$ be the error event of bit-channel i' for hybrid polar block b in the staircase, which is defined as follows:

$$\begin{aligned}\mathscr{E}_{i'b} = \{&(u'^{1:N\cdot 2^t}, \{y_j^{1:N}\}_{j=1}^{2^t}) \text{ tuples of blocks } \tilde{b} \in [2^t Npk] : \\ &P_{U'_{i'}|U'^{1:i'-1},\{Y_j^{1:N}\}_{j=1}^{2^t}}(u'_{i'}+1|u'^{1:i'-1}, \{y_j^{1:N}\}_{j=1}^{2^t}) \geq \\ &\qquad P_{U'_{i'}|U'^{1:i'-1},\{Y_j^{1:N}\}_{j=1}^{2^t}}(u'_{i'}|u'^{1:i'-1}, \{y_j^{1:N}\}_{j=1}^{2^t}) \\ &\qquad\qquad \text{holds for } (u'^{1:N\cdot 2^t}, \{y_j^{1:N}\}_{j=1}^{2^t}) \text{ of block } b\}.\end{aligned}$$

  As mentioned in the Theorem 1, the overall error event satisfies

$$\mathscr{E} \subset \cup_{b\in[2^t Npk]} \cup_{i'\in\mathscr{L}_{X_t}\cup I_l^t} \mathscr{E}_{i'b}. \tag{2.8}$$

  Let $\mathscr{E}_b = \cup_{i'\in I_l^t\cup\mathscr{L}_{X_t}} \mathscr{E}_{i'b}$ be an error event corresponding to hybrid polar block b in the staircase. We also define the error event $\mathscr{E}_{i'}$ of bit-channel i' below for the hybrid polar block when the hybrid block is directly used for code construction:

$$\begin{aligned}\mathscr{E}_{i'} = \{&(u'^{1:N\cdot 2^t}, \{y_j^{1:N}\}_{j=1}^{2^t}) : \\ &P_{U'_{i'}|U'^{1:i'-1},\{Y_j^{1:N}\}_{j=1}^{2^t}}(u'_{i'}+1|u'^{1:i'-1}, \{y_j^{1:N}\}_{j=1}^{2^t}) \geq \\ &\qquad P_{U'_{i'}|U'^{1:i'-1},\{Y_j^{1:N}\}_{j=1}^{2^t}}(u'_{i'}|u'^{1:i'-1}, \{y_j^{1:N}\}_{j=1}^{2^t})\}.\end{aligned}$$

- We apply the union bound to equation (2.8) and take the expectation under the measure induced by the random ensemble of codes. That gives the following upper-bound for the ensemble average probability of error as shown in Theorem 1:

$$\mathbb{E}_{W'^{1:N\cdot 2^t}}[\mathbb{P}(\mathscr{E}|W'^{1:N\cdot 2^t})] \leq \sum_{b\in[2^tNpk]} \mathbb{E}_{W'^{1:N\cdot 2^t}}[\mathbb{P}(\cup_{i'\in\mathscr{L}_{X_t}\cup I_l^t}\mathscr{E}_{i'b}|W'^{1:N\cdot 2^t})]. \tag{2.9}$$

- Now

$$\begin{aligned}
&\mathbb{P}(\mathscr{E}_b|W'^{1:N\cdot 2^t}) \\
&= \sum_{((u'^{1:N\cdot 2^t},\{y_j^{1:N}\}_{j=1}^{2^t}) \text{ tuples of all the blocks})\in\mathscr{E}_b} \\
&\quad \mathbb{P}(\cap_{\tilde{b}\in[2^tNpk]}(U'^{1:N\cdot 2^t}=u'^{1:N\cdot 2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}=\{y_j^{1:N}\}_{j=1}^{2^t} \text{ of block } \tilde{b})|W'^{1:N\cdot 2^t}).
\end{aligned}$$

From the definitions of $\mathscr{E}_{i'}$ and $\mathscr{E}_b$, we get

$$\begin{aligned}
&\mathbb{P}(\mathscr{E}_b|W'^{1:N\cdot 2^t}) \\
&= \sum_{((u'^{1:N\cdot 2^t},\{y_j^{1:N}\}_{j=1}^{2^t}) \text{ of block } b)\in\cup_{i'\in\mathscr{L}_{X_t}\cup I_l^t}\mathscr{E}_{i'}} \ \sum_{((u'^{1:N\cdot 2^t},\{y_j^{1:N}\}_{j=1}^{2^t}) \text{ tuples of all the blocks } \tilde{b}\neq b)} \\
&\quad \mathbb{P}(\cap_{\tilde{b}\in[2^tNpk]}(U'^{1:N\cdot 2^t}=u'^{1:N\cdot 2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}=\{y_j^{1:N}\}_{j=1}^{2^t} \text{ of block } \tilde{b})|W'^{1:N\cdot 2^t}),
\end{aligned}$$

By marginalizing over $(U'^{1:N\cdot 2^t},\{Y_j^{1:N}\}_{j=1}^{2^t})$ tuples of blocks $[2^tNpk]-\{b\}$, we get

$$\begin{aligned}
&\mathbb{P}(\mathscr{E}_b|W'^{1:N\cdot 2^t}) \\
&= \sum_{((u'^{1:N\cdot 2^t},\{y_j^{1:N}\}_{j=1}^{2^t}) \text{ of block } b)\in\cup_{i'\in\mathscr{L}_{X_t}\cup I_l^t}\mathscr{E}_{i'}} \\
&\qquad \mathbb{P}(\ (U'^{1:N\cdot 2^t}=u'^{1:N\cdot 2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}=\{y_j^{1:N}\}_{j=1}^{2^t} \text{ of block } b)|W'^{1:N\cdot 2^t}).
\end{aligned}$$

By the chain rule of conditional probability and also by the fact that

$$\mathbb{P}(\{Y_j^{1:N}\}_{j=1}^{2^t} = \{y_j^{1:N}\}_{j=1}^{2^t} \text{ of block } b | U'^{1:N\cdot 2^t} = u'^{1:N\cdot 2^t} \text{ of block } b, W'^{1:N\cdot 2^t}) = \Pi_{j=1}^{2^t}\Pi_{i=1}^{N} p_l(y_{ji}|x_{ji}),$$

where $\{x_j^{1:N}\}_{j=1}^{2^t}$ are the codeword component vectors of original blocks corresponding to bit-channel vector $u'^{1:N\cdot 2^t}$ of hybrid block b and $\{y_j^{1:N}\}_{j=1}^{2^t}$ is also of hybrid block b, we get

$$\mathbb{P}(\mathscr{E}_b|W'^{1:N\cdot 2^t}) = \sum_{((u'^{1:N\cdot 2^t}, \{y_j^{1:N}\}_{j=1}^{2^t}) \text{ of block } b) \in \cup_{i' \in \mathscr{L}_{X_t} \cup I_l^t} \mathscr{E}_{i'}} \mathbb{P}(U'^{1:N\cdot 2^t} = u'^{1:N\cdot 2^t} \text{ of block } b)|W'^{1:N\cdot 2^t})\Pi_{j=1}^{2^t}\Pi_{i=1}^{N} p_l(y_{ji}|x_{ji}).$$

By taking expectation on both sides followed by applying the linearity of expectation, we get,

$$\begin{aligned}
&\mathbb{E}_{W'^{1:N\cdot 2^t}}[\mathbb{P}(\mathscr{E}_b|W'^{1:N\cdot 2^t})] \\
&= \sum_{((u'^{1:N\cdot 2^t}, \{y_j^{1:N}\}_{j=1}^{2^t}) \text{ of block } b) \in \cup_{i' \in \mathscr{L}_{X_t} \cup I_l^t} \mathscr{E}_{i'}} \mathbb{E}_{W'^{1:N\cdot 2^t}}[\mathbb{P}(U'^{1:N\cdot 2^t} = u'^{1:N\cdot 2^t}|W'^{1:N\cdot 2^t})]\Pi_{j=1}^{2^t}\Pi_{i=1}^{N} p_l(y_{ji}|x_{ji}) \\
&= \sum_{((u'^{1:N\cdot 2^t}, \{y_j^{1:N}\}_{j=1}^{2^t}) \text{ of block } b) \in \cup_{i' \in \mathscr{L}_{X_t} \cup I_l^t} \mathscr{E}_{i'}} Q_{U'^{1:N\cdot 2^t}}(u'^{1:N\cdot 2^t} \text{ of block } b)\Pi_{j=1}^{2^t}\Pi_{i=1}^{N} p_l(y_{ji}|x_{ji}) \\
&= \sum_{(((u'^{1:N\cdot 2^t}, \{y_j^{1:N}\}_{j=1}^{2^t}) \text{ of block } b) \in \cup_{i' \in \mathscr{L}_{X_t} \cup I_l^t} \mathscr{E}_{i'}} Q_{U'^{1:N\cdot 2^t}, \{Y_j^{1:N}\}_{j=1}^{2^t}}((u'^{1:N\cdot 2^t}, \{y_j^{1:N}\}_{j=1}^{2^t}) \text{ of block } b)
\end{aligned}$$

$$= Q_{U'^{1:N\cdot 2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}}(\cup_{i'\in\mathscr{L}_{X_t}\cup I_l^t}\mathscr{E}_{i'})$$
$$= Q_{\{X_j^{1:N}\}_{j=1}^{2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}}(\cup_{i'\in\mathscr{L}_{X_t}\cup I_l^t}\mathscr{E}_{i'}),$$

where $Q_{U'^{1:N\cdot 2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}}$ is measure induced when $\{X_j^{1:N}\}_{j=1}^{2^t}$ vectors distributed under measure $Q_{U'^{1:N\cdot 2^t}}$ is transmitted over the DMC l selected and $\{Y_j^{1:N}\}_{j=1}^{2^t}$ vectors are received.

Therefore,

$$\mathbb{E}_{W'^{1:N\cdot 2^t}}[\mathbb{P}(\cup_{i'\in\mathscr{L}_{X_t}\cup I_l^t}\mathscr{E}_{i'b}|W'^{1:N\cdot 2^t})] = Q_{\{X_j^{1:N}\}_{j=1}^{2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}}(\cup_{i'\in\mathscr{L}_{X_t}\cup I_l^t}\mathscr{E}_{i'}). \tag{2.10}$$

- Now $Q_{\{X_j^{1:N}\}_{j=1}^{2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}}(\cup_{i'\in\mathscr{L}_{X_t}\cup I_l^t}\mathscr{E}_{i'})$ can be bounded as the sum of two entities as follows:

$$Q_{\{X_j^{1:N}\}_{j=1}^{2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}}(\cup_{i'\in\mathscr{L}_{X_t}\cup I_l^t}\mathscr{E}_{i'})$$
$$\leq P(\cup_{i'\in\mathscr{L}_{X_t}\cup I_l^t}\mathscr{E}_{i'}) + ||P_{\{X_j^{1:N}\}_{j=1}^{2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}} - Q_{\{X_j^{1:N}\}_{j=1}^{2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}}||.$$

- Note that there is a possibility that we combine two good bit-channels of DMC l in many of the t recursive steps while generating the hybrid block. In that case the Bhattacharyya parameter of a bit-channel $i' \in I_l^t$, $Z(U'_{i'}|U'^{1:i'-1}\{Y_j^{1:N}\}_{j=1}^{2^t})$, should be upper bounded as $2^t 2^{-N^\beta}$ since the upper bound of the Bhattacharyya parameter of one of the produced bit-channels after each combining is the sum of the Bhattacharyya parameters of the input bit-channels of the combining by Proposition 1.

- The $P(\cup_{i'\in\mathscr{L}_{X_t}\cup I_l^t}\mathscr{E}_{i'})$ is upper bounded by $\sum_{i'\in I_l^t\cup\mathscr{L}_{X_t}} P(\mathscr{E}_{i'})$ by using the union bound. Note that $P(\mathscr{E}_{i'})$ is upper bounded by $Z(U'_{i'}|U'^{1:i'-1}\{Y_j^{1:N}\}_{j=1}^{2^t})$ as the decision rule for these bit-channels is MAP (Maximum A Posteriori) decision rule under measure P [46, Proposition 2.7]. Therefore $P(\cup_{i'\in\mathscr{L}_{X_t}\cup I_l^t}\mathscr{E}_{i'})$ is upper bounded by the sum of Bhattacharyya parameters, i.e. $\sum_{i'\in I_l^t\cup\mathscr{L}_{X_t}} Z(U'_{i'}|U'^{1:i'-1}\{Y_j^{1:N}\}_{j=1}^{2^t})$, which will be $O(2^{2t}N2^{-N^\beta})$ from the above step.

- Now, the total variation distance satisfies

$$||P_{\{X_j^{1:N}\}_{j=1}^{2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}} - Q_{\{X_j^{1:N}\}_{j=1}^{2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}}|| \overset{(a)}{\leq} \sum_{j=1}^{2^t} ||P_{X_j^{1:N},Y_j^{1:N}} - Q_{X_j^{1:N},Y_j^{1:N}}|| \overset{(b)}{=} \sum_{j=1}^{2^t} ||P_{X_j^{1:N}} - Q_{X_j^{1:N}}||.$$

  The identity (a) follows by application of Lemma 3 using the fact that $\{X_j^{1:N}, Y_j^{1:N}\}_{j=1}^{2^t}$ vector tuples are i.i.d. distributed in both the measures P and Q . The identity (b) is true by the application of Lemma 4 using the fact that the conditional measure of $Y_j^{1:N}$ given $X_j^{1:N}$ in both the P and Q measures is induced by the selected DMC in S. Now the total variation distance $||P_{X_j^{1:N}} - Q_{X_j^{1:N}}||$ is $O(2^{-N^{\beta'}})$, as we mentioned in Section 2.3 for the single asymmetric channel case. Overall $||P_{\{X_j^{1:N}\}_{j=1}^{2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}} - Q_{\{X_j^{1:N}\}_{j=1}^{2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}}||$ is upper bounded by $O(2^t 2^{-N^{\beta'}})$ for $\beta' < \beta < 0.5$.

- Hence $Q_{\{X_j^{1:N}\}_{j=1}^{2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}}(\cup_{i' \in \mathscr{L}_{X_t} \cup I_l^t} \mathscr{E}_{i'})$ is upper bounded by $O(2^{2t} N 2^{-N^{\beta'}})$.

- From equations (2.9) and (2.10), the overall average error probability is upper bounded by $\sum_{b \in [2^t Npk]} Q_{\{X_j^{1:N}\}_{j=1}^{2^t},\{Y_j^{1:N}\}_{j=1}^{2^t}}(\cup_{i' \in \mathscr{L}_{X_t} \cup I_l^t} \mathscr{E}_{i'})$. Hence the overall average error probability of the hybridized staircase scheme will become $O(N^2 pk 2^{3t} 2^{-N^{\beta'}})$ for any DMC l in S.

2.5.4 Algorithm to produce hybrid polar block, to be used in the staircase scheme

In this sub-section, we provide an efficient recursive method in Algorithm 1 to produce a hybrid polar block that satisfies the desired condition with a block length at most 2^{s-1} times the original polar block length.

We refer to the properties associated to a hybrid polar block, such as good bit-channel set, bad bit-channel set, low-entropy bit-channel set, not-completely polar bit-channel set, order of bit-channels, as the type of the hybrid block. The variables *hPolarBlock1* and *hPolarBlock2*

identify with the type of a hybrid polar block that gets updated in the course of the recursive procedure.

Algorithm 1: getHybridizedPolarBlock(S_1, γ, *hPolarBlock1*)

Input: $S_1 \subset S$, scale index γ and *hPolarBlock1* with $|\cap_{i\in S_1} I_i^1|_N < 2^\gamma |R|$ satisfied

Output: *hPolarBlock2* with $|\cap_{i\in S_1} I_i^2|_N \geq 2^\gamma |R|$ satisfied

1 set $S_2 = \underset{S_2\subset S_1 : |S_2|=|S_1|-1}{\operatorname{argmax}} |\cap_{i\in S_2} I_i^1|_N$

2 **if** $|\cap_{i\in S_2} I_i^1|_N < 2^{\gamma+1}|R|$ **then**

    `/* Recursive call */`

3   *hPolarBlock1* = getHybridizedPolarBlock(S_2, $\gamma+1$, *hPolarBlock1*)

`/* Now` $|\cap_{i\in S_2} I_i^1|_N \geq 2^{\gamma+1}|R|$ `is satisfied */`

`/* Combing step */`

4 consider two independent hybrid polar blocks of *hPolarBlock1* type

5 combine bit channels that are "good for all DMCs in S_2 and bad for DMC $S_1 - S_2$" of one block with bit channels that are "good for DMC $S_1 - S_2$ and bad for at least one DMC in S_2" of the other block to produce an updated hybrid polar block

6 set *hPolarBlock2* to the type of updated polar block

`/* Now` $|\cap_{i\in S_1} I_i^2|_N \geq 2^\gamma |R|$ `is satisfied */`

7 **return** *hPolarBlock2*

We refer to I_i^1 and I_i^2 for DMC i's good bit-channel set of hybrid polar blocks of *hPolarBlock1* type and *hPolarBlock2* type, respectively. We refer to $|\cap_{i\in S_1} I_i^1|_N$ and $|\cap_{i\in S_1} I_i^2|_N$ for the size of bit-channel set $\cap_{i\in S_1} I_i^1$ and $\cap_{i\in S_1} I_i^2$ per block length N. If the variable *hPolarBlock1* is a type of a hybrid polar block produced after k recursive combinings, then $|\cap_{i\in S_1} I_i^1|_N$ will be equal to $\frac{|\cap_{i\in S_1} I_i^1|}{2^k}$ as the overall block length then will be $2^k N$. $|R|$ is the size of the not-completely polarized bit-channel set of an original polar block.

Algorithm 1 is implemented through recursive procedure getHybridizedPolarBlock(),

which has three inputs, the first one of which is a subset S_1 of the compound channel set S. The second input is scale index γ and the third input is *hPolarBlock1* with $|\cap_{i\in S_1} I_i^1|_N < 2^\gamma |R|$ satisfied. getHybridizedPolarBlock() procedure performs recursive combining to produce a hybrid block and returns the variable *hPolarBlock2* updated to the type of the hybrid polar block produced with $|\cap_{i\in S_1} I_i^2|_N \geq 2^\gamma |R|$ satisfied.

To generate hybrid polar block, to be used in staircase scheme, from an original polar block, which does not satisfy the desired condition, we call getHybridizedPolarBlock() with input S_1 initialized to compound channel set S, the input γ initialized to 0 and the input $hPolarBlock1$ initialized to the original polar block type. Now we describe the execution of recursive procedure getHybridizedPolarBlock() in detail with these three particular inputs.

First, we set S_2 as a subset of S_1 with size one less than the size of S_1 such that $|\cap_{i\in S_2} I_i^1|_N$ is maximum. Note that the input S_1 is S. Then, we check the condition $|\cap_{i\in S_2} I_i^1|_N < 2|R|$, since γ is 0. The condition not being true means that we already have $|\cap_{i\in S_2} I_i^1|_N \geq 2|R|$ satisfied. If that checked condition is true, we call getHybridizedPolarBlock() a second time with inputs S_2, 1, *hPolarBlock1*, which is the original polar block type. So the call returns and updates the variable *hPolarBlock1* so that the condition $|\cap_{i\in S_2} I_i^1|_N \geq 2|R|$ is satisfied. Now we consider two independent hybrid polar blocks of *hPolarBlock1* type. We combine "bit channels that are good for all DMCs in S_2 and bad for DMC $S - S_2$" of one block with "bit channels that are good for DMC $S - S_2$ and bad for at least one DMC in S_2" of the other block to generate a new hybrid block. We then update the variable *hPolarBlock2* to the type of the new hybrid block generated. Notice that, $|\cap_{i\in S} I_i^2|_N$ gets updated as $|\cap_{i\in S} I_i^1|_N + \frac{\min\{|\cap_{i\in S_2} I_i^1|_N - |\cap_{i\in S} I_i^1|_N, |\cap_{i\in S-S_2} I_i^1|_N - |\cap_{i\in S} I_i^1|_N\}}{2}$, which is at least $|\cap_{i\in S} I_i^1|_N + \frac{2|R| - |\cap_{i\in S} I_i^1|_N}{2}$ as $|\cap_{i\in S_2} I_i^1|_N \geq 2|R|$. So $|\cap_{i\in S} I_i^2|_N$ will be at least $|R| + |\cap_{i\in S} I_i^1|_N/2$, which is greater than or equal to $|R|$. Now we return *hPolarBlock2*, which is the type of a desired hybrid polar block to be used in the staircase.

We now analyse the recursive flow of calls to getHybridizedPolarBlock() and the depth of recursion. As the number of recursive calls in our algorithm is not always fixed, we look at the worst case recursion depth. In the first call of getHybridizedPolarBlock(), we need

hPolarblock1 (initialized to original block type) such that the condition $|\cap_{i\in S_2} I_i^1|_N \geq 2|R|$ is satisfied for the combining step. If the condition does not hold, we invoke the second call to getHybridizedPolarBlock(), which updates the *hPolarblock1* to a hybrid polar block type such that the condition $|\cap_{i\in S_2} I_i^1|_N \geq 2|R|$ is satisfied. If the condition holds, we do not invoke call to getHybridizedPolarBlock() a second time, in which case the recursion depth is 1. In second call, we need *hPolarblock1* (initialized to original block type) such that the condition $|\cap_{i\in S_2} I_i^1|_N \geq 2^2|R|$ is satisfied. Note that size of S_2 is two less than size of S in this call. If the condition does not hold, we invoke the third call to getHybridizedPolarBlock(), which updates the *hPolarblock1* to a hybrid polar block type such that $|\cap_{i\in S_2} I_i^1|_N \geq 2^2|R|$ is satisfied. If the condition holds, we do not invoke call to getHybridizedPolarBlock() third time, in which case the recursion depth is 2. Similarly, in the kth call, we need *hPolarblock1* (initialized to original block type) such that the condition $|\cap_{i\in S_2} I_i^1|_N \geq 2^k|R|$ is satisfied. If the condition does not hold, we invoke a $(k+1)$th call to getHybridizedPolarBlock(), which updates the *hPolarblock1* to a hybrid polar block type such that $|\cap_{i\in S_2} I_i^1|_N \geq 2^k|R|$ is satisfied. If it holds, we do not invoke a $(k+1)$th call, in which case recursion depth is k. Note that the size of S_1 will be $k-1$ less than the size of S and the size of S_2 will be k less than the size of S in this call to getHybridizedPolarBlock(), since the size of S_1 keeps decreasing by one whenever we make an additional call to getHybridizedPolarBlock() in the recursive flow. So, if the algorithm happens to execute the $(s-1)$th call to getHybridizedPolarBlock(), then $|S_2|$ will be one and *hPolarblock1* (original block type) will be such that the condition $|\cap_{i\in S_2} I_i^1|_N \geq 2^{s-1}|R|$ is satisfied for any S_2 as $\min\{|I_1^1|, |I_2^1|, \ldots, |I_s^1|\} > 2^{s-1}|R|$ holds for sufficiently large N as the fraction of bit-channels in R of original polar block vanishes as block length grows due to polarization. So we never make sth call to getHybridizedPolarBlock(). Hence the maximum recursion depth that is possible is $s-1$.

Now we will look at what happens in the execution of the kth call to getHybridized-PolarBlock(), which is called from the $(k-1)$th call of getHybridizedPolarBlock() with the input *hybridBlock1* with the condition $|\cap_{i\in S_2} I_i^1|_N < 2^{k-1}|R|$ satisfied (where S_2 here is of the

$(k-1)$th call). S_2 of the $(k-1)$th call is passed as input S_1 to the kth call. Scale index input will be $k-1$ in that call. As we set S_2 as a subset of S_1 with size one less than the size of S_1 such that $|\cap_{i\in S_2} I_i^1|_N$ is maximum, $|S_2|$ will be $|S|-k$ and $|S_1|$ will be $|S|-(k-1)$ in this call to getHybridizedPolarBlock(). Now we check the condition if $|\cap_{i\in S_2} I_i^1|_N < 2^k|R|$, since scale index input is $k-1$ in this call to getHybridizedPolarBlock(). If that is true, we then call getHybridizedPolarBlock() a $(k+1)$th time with inputs S_2, k and *hPolarBlock1*, which will be of the original polar block type. So the recursive call returns and updates the variable *hPolarBlock1* that satisfies $|\cap_{i\in S_2} I_i^1|_N \geq 2^k|R|$. Now we consider two independent hybrid polar blocks of *hPolarBlock1* type. We combine "bit channels that are good for all DMCs in S_2 and bad for DMC S_1-S_2" of one block with "bit channels that are good for DMC S_1-S_2 and bad for at least one DMC in S_2" of the other block to generate a new hybrid block. We then update the variable *hPolarBlock2* as the type of the generated hybrid block. As we update *hPolarBlock2* after the combining step with the generated hybrid polar block type, $|\cap_{i\in S} I_i^2|_N$ gets updated as $|\cap_{i\in S_1} I_i^1|_N + \frac{\min\{|\cap_{i\in S_2} I_i^1|_N - |\cap_{i\in S_1} I_i^1|_N, |\cap_{i\in S_1-S_2} I_i^1|_N - |\cap_{i\in S_1} I_i^1|_N\}}{2}$ which is at least $|\cap_{i\in S_1} I_i^1|_N + \frac{2^k|R| - |\cap_{i\in S_1} I_i^1|_N}{2}$ as $|\cap_{i\in S_2} I_i^1|_N \geq 2^k|R|$. So $|\cap_{i\in S_1} I_i^2|_N$ will be at least $2^{k-1}|R| + |\cap_{i\in S_1} I_i|_N/2$, which is greater than or equal to $2^{k-1}|R|$. Now we return *hPolarBlock2* to the point of execution in the $(k-1)$th call, where the kth call is invoked from.

If the recursion depth is k, in the execution of the kth call to getHybridizedPolarBlock(), we combine original polar blocks and return *hPolarBlock2* updated as the type of hybrid polar block generated to the point of execution in the $(k-1)$th call, where the kth call to getHybridizedPolarBlock() is invoked from. Again we combine two independent hybrid blocks of the returned type in the execution of the $(k-1)$th recursive call and return *hPolarBlock2* updated as the type of hybrid polar block generated to the point of execution in the $(k-2)$th call, where the $(k-1)$th call is invoked from. We keep doing this until we return to the first call and finish the combining step in the first call to produce the hybrid block, to be used in the staircase. Hence the block length of hybrid block returned by the algorithm in this case will be 2^kN. Since the maximum recursion depth is $s-1$, the desired hybrid block returned by the algorithm is at

most $2^{s-1}N$.

We may improve the Algorithm 1 to have fewer recursive calls. Nevertheless, the mentioned recursive procedure guarantees a block length which is at most $2^{s-1}N$. Hence the overall block length of the hybridized staircase scheme becomes $O(2^{2(s-1)}pkN^2)$.

2.6 Universal scheme via combining bit-channels

In Section 2.5.1, we discussed an idea of universal procedure based on bit-channel combining to produce a hybrid block. But we do not describe the detail order of bit-channels of the hybrid block produced when combining two blocks. In this section, we describe the code construction, including encoding and decoding methods directly using the hybrid block produced after combining the two original blocks. We consider here S to be $\{1,2\}$. So we combine "bit-channels that are good for DMC 1 and bad for DMC 2" of one block with "bit-channels that are good for DMC 2 and bad for DMC 1" of the other block. The description of the scheme in this section helps provide a clear view of the bit-channel order of the hybrid block generated after combining the original blocks.

2.6.1 Combining two independent polar blocks to align good bit-channels of the two DMCs in S

Let $G = \min\{|I_1 \cap F_2|, |I_2 \cap F_1|\}$. Consider the sets $\mathscr{A}$ and $\mathscr{B}$ to be the first G indices in $I_1 \cap F_2$ and $I_2 \cap F_1$, respectively. Let $\mathscr{A} = \{x_1, x_2, \ldots, x_G\}$ where $x_1 < x_2 < \ldots < x_G$ and $\mathscr{B} = \{y_1, y_2, \ldots, y_G\}$ where $y_1 < y_2 < \ldots < y_G$. Consider two independent polar blocks and refer to them as block 1 and block 2. Let $X^{1:2N}$ be i.i.d. distributed according to non-uniform compound capacity-achieving distribution $p(x)$ and

$$U^{1:N} = X^{1:N}G_N, \quad V^{1:N} = X^{N+1:2N}G_N. \tag{2.11}$$

The vectors $X^{1:N}$ and $X^{N+1:2N}$ are codeword components of block 1 and block 2, respectively. $U^{1:N}$ and $V^{1:N}$ are bit-channel vectors of block 1 and block 2, respectively. For each $j \in [G]$, we combine bit-channel x_j of block 1 with bit-channel y_j of block 2, which produces two new bit-channels with inputs $U'_{x_j} = U_{x_j} + V_{y_j}$ and $V'_{y_j} = V_{y_j}$. We do not involve bit-channels, which are not in $\mathscr{A}$ of block 1 and not in $\mathscr{B}$ of block 2, in combining. The combining of bit-channels of two independent blocks is shown in Figure 2.5 of Section 2.5. Let $\{(X_i, Y_i)\}_{i=1}^{2N}$ be i.i.d. distributed according to $p(x)p_l(y|x)$ where $l \in S$.

Lemma 5. *For each $j \in [G]$, for any $\beta < 0.5$, for sufficiently large N*

1. $Z(U'_{x_j}|U^{1:x_j-1}V^{1:y_j-1}Y^{1:2N}) \geq 1 - 2^{-N^\beta}$.

(U'_{x_j} is almost uniform given $U^{1:x_j-1}V^{1:y_j-1}Y^{1:2N}$ for both DMCs $l = 1, 2$)

2. $Z(V'_{y_j}|U^{1:x_j-1}V^{1:y_j-1}U'_{x_j}Y^{1:2N}) \leq 2^{-N^\beta}$.

(V'_{y_j} is almost deterministic given $U^{1:x_j-1}U'_{x_j}V^{1:y_j-1}Y^{1:2N}$ for both DMCs $l = 1, 2$)

Proof: Refer to the Appendix.

It follows from Lemma 5 that the bit-channel combinings mentioned above give us a block of length $2N$ with $2|I_1 \cap I_2| + G$ good bit-channels for both the DMCs in S. In the random code construction that we propose, encoding method ensures that the ensemble average distribution of $(U'^{1:N}, V'^{1:N})$ is $O(2^{-N^{\beta'}})$ close to the distribution induced when the word $X^{1:2N}$ is i.i.d. distributed according to $p(x)$. Now we describe the code construction.

2.6.2 Code construction

We first generate random functions $f_1 : \mathscr{H}_X - (I_1 \cap I_2) \to \{0,1\}$ and $f_2 : \mathscr{H}_X - ((I_1 \cap I_2) \cup \mathscr{B}) \to \{0,1\}$ where each $f_i(j)$, $j \in F$ and $i \in \{1,2\}$, is chosen independently and uniformly. These frozen bits are shared between encoder and decoder.

For both the blocks, we generate independent random boolean functions $\lambda_i^b : \{0,1\}^{i-1} \to \{0,1\}$ for blocks $b = 1, 2$, and for each $i \in R$, by using the following probability rule:

$$\lambda_i^b(u^{1:i-1}) = u \text{ w.p. } P_{U_i|U^{1:i-1}}(u|u^{1:i-1}), \text{ for } u \in \{0,1\}$$

independently for each $u^{1:i-1}$. Let the set of random functions be denoted by λ_R^b. These functions are used to encode not-completely polarized bit-channels and are shared between the encoder and the decoder. We can alternatively use common randomness for encoding these bit-channels. As mentioned earlier in this section, the goal here is to help provide a clear view of the order of bit-channels as opposed to simplifying the encoding of bit-channels in R avoiding common randomness. Now we describe the encoding and decoding algorithms.

Encoding

Input: uniform message M of $2|I_1 \cap I_2| + G$ bits

Output: codeword $X^{1:2N}$

1. Partition M into M_1 and M_2 such that M_1 takes the first $|I_1 \cap I_2|$ bits of M and M_2 takes the last $|I_1 \cap I_2| + G$ bits of M. Set $U'^{I_1 \cap I_2} = M_1$ and $V'^{(I_1 \cap I_2) \cup \mathscr{B}} = M_2$.

2. Set $U_i' = f_1(i)$ for all $i \in \mathscr{H}_X - (I_1 \cap I_2)$ in block 1 and $V_i' = f_2(i)$ for all $i \in \mathscr{H}_X - ((I_1 \cap I_2) \cup \mathscr{B})$ in block 2.

3. Set $U_{x_j} = U'_{x_j} + V'_{y_j}$ and $V_{y_j} = V'_{y_j}$ for all $j \in [G]$.

4. Set $U_i = U_i'$ for all $i \in \mathscr{H}_X - \mathscr{A}$ and $V_i = V_i'$ for all $i \in \mathscr{H}_X - \mathscr{B}$.

5. For all $i \in \mathscr{L}_X$, set U_i and V_i using the following argmax rules:

$$U_i = \underset{x \in \{0,1\}}{\operatorname{argmax}} P_{U_i|U^{1:i-1}}(x|U^{1:i-1}),$$

$$V_i = \underset{x \in \{0,1\}}{\operatorname{argmax}} P_{U_i|U^{1:i-1}}(x|V^{1:i-1}).$$

6. For all $i \in R$, we assign $U_i = \lambda_i^1(U^{1:i-1})$ and $V_i = \lambda_i^2(V^{1:i-1})$.

7. Set $U_i' = U_i$ and $V_i' = V_i$ for all $i \notin \mathscr{H}_X$.

8. Transmit $X^{1:N} = U^{1:N} G_N$ and $X^{N+1:2N} = V^{1:N} G_N$.

Decoding

Input: received vector $Y^{1:2N}$

Output: message estimate $\hat{M}$ of $2|I_1 \cap I_2| + G$ bits

1. Set $j = 1$, $x_0 = 0$ and $y_0 = 0$.

2. **for** $i = x_{j-1} + 1 : x_j - 1$ of block 1

   If $i \in \mathscr{H}_X - (I_1 \cap I_2)$, set

   $\hat{U}'_i = \hat{U}_i = f_1(i)$.

   If $i \in (I_1 \cap I_2) \cup \mathscr{L}_X$, set

   $\hat{U}'_i = \hat{U}_i = \underset{x \in \{0,1\}}{\operatorname{argmax}} P_{U_i|U^{1:i-1},Y^{1:N}}(x|\hat{U}^{1:i-1}, Y^{1:N})$.

   If $i \in R$, set

   $\hat{U}'_i = \hat{U}_i = \lambda_i^1(\hat{U}^{1:i-1})$.

   end

   for $i = y_{j-1} + 1 : y_j - 1$ of block 2

   If $i \in \mathscr{H}_X - (I_1 \cap I_2)$, set

   $\hat{V}'_i = \hat{V}_i = f_2(i)$.

   If $i \in (I_1 \cap I_2) \cup \mathscr{L}_X$, set

   $\hat{V}'_i = \hat{V}_i = \underset{x \in \{0,1\}}{\operatorname{argmax}} P_{U_i|U^{1:i-1},Y^{1:N}}(x|\hat{V}^{1:i-1}, Y^{N+1:2N})$.

   If $i \in R$, set

   $\hat{V}'_i = \hat{V}_i = \lambda_i^2(\hat{V}^{1:i-1})$.

   end

3. Set

   $\hat{U}'_{x_j} = f_1(x_j)$.

   $\hat{V}'_{y_j} = \underset{x \in \{0,1\}}{\operatorname{argmax}} P_{V'_{y_j}|U^{1:x_j-1}\hat{U}'_{x_j}V^{1:y_j-1}Y^{1:2N}}(x|\hat{U}^{1:x_j-1}U'_{x_j}\hat{V}^{1:y_j-1}Y^{1:2N})$.

   $\hat{U}_{x_j} = \hat{U}'_{x_j} + \hat{V}'_{y_j}$ and $\hat{V}_{y_j} = \hat{V}'_{y_j}$.

4. Repeat steps 2 and 3 for $j = \{2, 3, \ldots, G\}$.

5. **for** $i = x_G + 1 : N$ of block 1

   If $i \in \mathscr{H}_X - (I_1 \cap I_2)$, set

$\hat{U}'_i = \hat{U}_i = f_1(i)$.

If $i \in (I_1 \cap I_2) \cup \mathscr{L}_X$, set

$\hat{U}'_i = \hat{U}_i = \underset{x\in\{0,1\}}{\operatorname{argmax}} P_{U_i|U^{1:i-1},Y^{1:N}}(x|\hat{U}^{1:i-1}, Y^{1:N})$.

If $i \in R$, set

$\hat{U}'_i = \hat{U}_i = \lambda_i^1(\hat{U}^{1:i-1})$.

end

for $i = y_G + 1 : N$ of block 2

If $i \in \mathscr{H}_X - (I_1 \cap I_2)$, set

$\hat{V}'_i = \hat{V}_i = f_2(i)$.

If $i \in (I_1 \cap I_2) \cup \mathscr{L}_X$, set

$\hat{V}'_i = \hat{V}_i = \underset{x\in\{0,1\}}{\operatorname{argmax}} P_{U_i|U^{1:i-1},Y^{1:N}}(x|\hat{V}^{1:i-1}, Y^{N+1:2N})$.

If $i \in R$, set

$\hat{V}'_i = \hat{V}_i = \lambda_i^2(\hat{V}^{1:i-1})$.

end

6. Set $\hat{M}_1 = \hat{U}'^{I_1 \cap I_2}$ and $\hat{M}_2 = \hat{V}'^{(I_1 \cap I_2) \cup \mathscr{B}}$. Combine $\hat{M}_1$, $\hat{M}_2$ to get $\hat{M}$.

Step 2 and step 5 in the above decoding algorithm are for decoding bit-channels which are not involved in the combining procedure, whereas step 3 is for decoding the new bit-channels which are produced after combining. The decoding method clearly shows the order in which we decode the bit-channels of the hybrid block $(U'^{1:N}, V'^{1:N})$, which is governed by the bit-channel combinings between the two original blocks.

Theorem 2. *Let $P_{e,l}(\lambda_R^1, \lambda_R^2, f_1, f_2)$ denote the decoding probability of error when DMC l is selected in S for a given code in the above random code construction. For sufficiently large block length N, the average decoding probability of error $\mathbb{E}[P_{e,l}(\lambda_R^1, \lambda_R^2, f_1, f_2)] = O(2^{-N^{\beta'}})$ for each $l \in S$, where $\beta' < \beta < 0.5$.*

Proof: Refer to the Appendix.

2.7 Conclusion

We presented a universal polar coding scheme for a compound channel defined by a finite set of binary-input asymmetric DMCs with non-uniform compound capacity-achieving input distribution. The proposed scheme exploits the underlying staircase structure in the code construction to avoid the need for side-channel transmission, storage-intensive boolean functions, or common randomness for bits corresponding to not-completely polarized bit-channels. We assume a condition that the number of bit-channels that are good for all the DMCs in the compound channel is greater than the number of not-completely polarized bit-channels to propose the code construction. When the condition does not hold, we proposed a hybridized staircase scheme, in which we use a hybrid polar block, with a block length at most 2^{s-1} times the original polar block length, that satisfies the desired condition.

The staircase scheme we proposed also requires a large block length and suffers from delay properties as in the symmetric channel case [22], even after the implementation of the proposed continuous encoding and decoding. This leaves open the problem of designing codes with short block length. Another open problem is the construction of a stronger universal polar code with reduced storage complexity that achieves rate r less than compound capacity with non-uniform compound capacity-achieving distribution $p(x)$, while also achieving rate r for any DMC whose mutual information evaluated at $p(x)$ is larger than r and avoiding common randomness. Another interesting open problem is reducing the height of the staircase to shorten the overall block length for a single asymmetric channel capacity-achieving scheme. In each full-height column, if the information bit-channels are at least as many as not-completely polarized bit-channels, we can implement the code construction that achieves capacity. So the problem is determining the shortest height h such that any h consecutive bit-channels have as many information bit-channels as not-completely polarized bit-channels. If h is sub-linear in the block length, the delay at which the staircase scheme operates will just be $o(N)$, which can make the scheme a good asymmetric channel capacity-achieving scheme.

2.8 Appendix

Proof of Theorem 1:

We first prove part 1 of Theorem 1.

Step 1:

Consider any polar block in the extended staircase which lies completely in the full-height column regime. To get the distribution on $U^{1:N}$ for such a polar block, we first compute the conditional distribution $\mathbb{P}(U_i = u_i | U^{1:i-1} = u^{1:i-1}, W^{1:N})$ for each bit-channel i in the block.

If $i \in \mathscr{L}_X$, by the encoding rule

$$\mathbb{P}(U_i = u_i | U^{1:i-1} = u^{1:i-1}, W^{1:N}) = \delta_i(u_i | u^{1:i-1}).$$

If $i \in R$, by the encoding rule

$$\mathbb{P}(U_i = u_i | U^{1:i-1} = u^{1:i-1}, W^{1:N}) = P_{U_i|U^{1:i-1}}(u_i | u^{1:i-1}).$$

If $i \in \mathscr{H}_X - I'$, by Lemma 1

$$\mathbb{P}(U_i = u_i | U^{1:i-1} = u^{1:i-1}, W^{1:N}) = 0.5.$$

If $i \in I'$, we will have

$$\mathbb{P}(U_i = u_i | U^{1:i-1} = u^{1:i-1}, W^{1:N}) = 0.5.$$

Now we discuss the case $i \in I'$ in detail. H is the designated information bit corresponding to that column. $\tilde{U}_{g(i)}$ is the already encoded bit in the block corresponding to the bit-channel $g(i)$ in that column. We have $U_i = H \oplus \tilde{U}_{g(i)}$. The distribution of H is Bernoulli(0.5). Note that the random variables H and $\tilde{U}_{g(i)}$ are independent. Now,

$$\mathbb{P}(U_i = x | \tilde{U}_{g(i)} = y, W^{1:N}) = \mathbb{P}(H + \tilde{U}_{g(i)} = x | \tilde{U}_{g(i)} = y, W^{1:N})$$

$$= \mathbb{P}(H = x+y|\tilde{U}_{g(i)} = y, W^{1:N}).$$

Since H is independent of $\tilde{U}_{g(i)}$ and $W^{1:N}$, we get

$$\mathbb{P}(U_i = x|\tilde{U}_{g(i)} = y, W^{1:N}) = \mathbb{P}(H = x+y) = 0.5.$$

Therefore U_i and $\tilde{U}_{g(i)}$ are independent. Now we establish that U_i is independent of all the encoded bits of previous columns and the frozen vector $W^{1:N}$. This will imply that U_i is independent of $U^{1:i-1}$ of that block. $\overline{P}$ is the random vector denoting the encoded bits of the previous columns. Now the conditional probability

$$\begin{aligned}\mathbb{P}(U_i = u_i|\overline{P}, W^{1:N}) &= \sum_{y\in\{0,1\}} \mathbb{P}(U_i = u_i, \tilde{U}_{g(i)} = y|\overline{P}, W^{1:N}) \\ &= \sum_{y\in\{0,1\}} \mathbb{P}(\tilde{U}_{g(i)} = y|\overline{P}, W^{1:N})\mathbb{P}(U_i = u_i|\tilde{U}_{g(i)} = y, \overline{P}, W^{1:N}).\end{aligned}$$

Since U_i is independent of bits encoded in previous columns and $W^{1:N}$ given the random variable $\tilde{U}_{g(i)}$, we get

$$\mathbb{P}(U_i = u_i|\overline{P}, W^{1:N}) = \sum_{y\in\{0,1\}} \mathbb{P}(\tilde{U}_{g(i)} = y|\overline{P}, W^{1:N})\mathbb{P}(U_i = u_i|\tilde{U}_{g(i)} = y).$$

As U_i and $\tilde{U}_{g(i)}$ are independent, we get

$$\begin{aligned}\mathbb{P}(U_i = u_i|\overline{P}, W^{1:N}) &= \sum_{y\in\{0,1\}} \mathbb{P}(\tilde{U}_{g(i)} = y|\overline{P}, W^{1:N})\mathbb{P}(U_i = u_i) \\ &= \mathbb{P}(U_i = u_i) \\ &= 0.5.\end{aligned}$$

Hence the distribution of $U^{1:N}$ for a block which lies completely in the full-height column regime

becomes

$$\begin{aligned}\mathbb{P}(U^{1:N}=u^{1:N}|W^{1:N}) &= \Pi_{i\in[N]}\mathbb{P}(U_i=u_i|U^{1:i-1}=u^{1:i-1},W^{1:N}) \\ &= 2^{-|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}\delta_i(u_i|u^{1:i-1})\Pi_{i\in R}P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1}).\end{aligned}$$

This implies that

$$\mathbb{E}_{W^{1:N}}[\mathbb{P}(U^{1:N}=u^{1:N}|W^{1:N})] = 2^{-|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}\delta_i(u_i|u^{1:i-1})\Pi_{i\in R}P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1}).$$

Step 2:

Consider a polar block which lies partly in the non-full-height column regime on the right side. For U_i in a full-height column, the conditional probability rule is already derived in step 1. We now derive the conditional probablility for U_i in a non-full-height column.

By the encoding rule we have:

If $i\in\mathscr{H}_X$,

$$\mathbb{P}(U_i=u_i|U^{1:i-1}=u^{1:i-1},W^{1:N})=0.5.$$

If $i\in\mathscr{L}_X$,

$$\mathbb{P}(U_i=u_i|U^{1:i-1}=u^{1:i-1},W^{1:N})=\delta_i(u_i|u^{1:i-1}).$$

If $i\in R$,

$$\mathbb{P}(U_i=u_i|U^{1:i-1}=u^{1:i-1},W^{1:N})=P_{U_i|U^{1:i-1}W^{1:N}}(u_i|u^{1:i-1}).$$

This implies that

$$\begin{aligned}\mathbb{P}(U^{1:N}=u^{1:N}|W^{1:N}) &= \Pi_{i\in[N]}\mathbb{P}(U_i=u_i|U^{1:i-1}=u^{1:i-1},W^{1:N}) \\ &= 2^{-|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}\delta_i(u_i|u^{1:i-1})\Pi_{i\in R}P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1}).\end{aligned} \tag{2.12}$$

Hence the ensemble average distribution of $U^{1:N}$ becomes

$$\mathbb{E}_{W^{1:N}}[\mathbb{P}(U^{1:N}=u^{1:N}|W^{1:N})] = 2^{-|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}\delta_i(u_i|u^{1:i-1})\Pi_{i\in R}P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1}).$$

Step 3:

Consider a polar block which lies partly in the non-full-height column regime on the left side. Let $\tilde{H}$ be the set of bit-channels of the block that lie in the non-full-height column regime. Those bits are encoded as W_i corresponding to every index $i \in \tilde{H}$. Now,

$$\begin{aligned}\mathbb{E}_{W^{1:N}}[\Pi_{i\in\tilde{H}}\mathbb{1}(u_i=W_i)] &= \mathbb{E}_{W^{1:N}}[\mathbb{1}(\cap_{i\in\tilde{H}}(u_i=W_i))] \\ &= \mathbb{E}_{W^{1:N}}[\sum_{u_i\in\{0,1\}:i\in\tilde{H}^c}\mathbb{1}(\cap_{i\in[N]}(u_i=W_i))].\end{aligned}$$

By using the linearity of expectation, we get

$$\begin{aligned}&\mathbb{E}_{W^{1:N}}[\Pi_{i\in\tilde{H}}\mathbb{1}(u_i=W_i)] \\ &\qquad= \sum_{u_i\in\{0,1\}:i\in\tilde{H}^c}\mathbb{E}_{W^{1:N}}[\mathbb{1}(\cap_{i\in[N]}(u_i=W_i))] \\ &\qquad\overset{(a)}{=} \sum_{u_i\in\{0,1\}:i\in\tilde{H}^c}2^{-|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}\delta_i(u_i|u^{1:i-1})\Pi_{i\in R}P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1}). \\ &\qquad= 2^{-|\tilde{H}\cap\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X\cap\tilde{H}}\delta_i(u_i|u^{1:i-1})\Pi_{i\in R\cap\tilde{H}}P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1}).\end{aligned} \tag{2.13}$$

Identity (a) is true because of the fact that

$$\mathbb{E}_{W^{1:N}}[1(\cap_{i\in[N]}(u_i=W_i))] = 2^{-|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}\delta_i(u_i|u^{1:i-1})\Pi_{i\in R}P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1}),$$

obtained from random construction of $W^{1:N}$ in equation (2.4). Now the distribution of $U^{1:N}$ will become

$$\mathbb{P}(U^{1:N}=u^{1:N}|W^{1:N}) = \Pi_{i\in[N]}\mathbb{P}(U_i=u_i|U^{1:i-1}=u^{1:i-1},W^{1:N})$$

$$\stackrel{(a)}{=} \Pi_{i\in\tilde{H}} \mathbb{1}(u_i = W_i) 2^{-|\mathscr{H}_X - \tilde{H}|} \Pi_{i\in\mathscr{L}_X - \tilde{H}} \delta_i(u_i|u^{1:i-1})$$
$$\cdot \Pi_{i\in R-\tilde{H}} P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1}).$$

Identity (a) follows by substituting the conditional distribution of U_i given $U^{1:i-1}$ of a full-height column derived in Step 1.

This implies

$$\mathbb{E}_{W^{1:N}}[\mathbb{P}(U^{1:N} = u^{1:N}|W^{1:N})]$$
$$= \mathbb{E}_{W^{1:N}}[\Pi_{i\in\tilde{H}} \mathbb{1}(u_i = W_i) 2^{-|\mathscr{H}_X - \tilde{H}|} \Pi_{i\in\mathscr{L}_X - \tilde{H}} \delta_i(u_i|u^{1:i-1})$$
$$\cdot \Pi_{i\in R-\tilde{H}} P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1})].$$

By linearity of expectation, we get

$$\mathbb{E}_{W^{1:N}}[\mathbb{P}(U^{1:N} = u^{1:N}|W^{1:N})]$$
$$= \mathbb{E}_{W^{1:N}}[\Pi_{i\in\tilde{H}} \mathbb{1}(u_i = W_i)] 2^{-|\mathscr{H}_X - \tilde{H}|} \Pi_{i\in\mathscr{L}_X - \tilde{H}} \delta_i(u_i|u^{1:i-1})$$
$$\cdot \Pi_{i\in R-\tilde{H}} P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1})]$$
$$\stackrel{(a)}{=} 2^{-|\mathscr{H}_X|} \Pi_{i\in\mathscr{L}_X} \delta_i(u_i|u^{1:i-1}) \Pi_{i\in R} P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1}).$$

Identity (a) follows from equation (2.13). This concludes the proof of part 1.

We now prove part 2 of Theorem 1.

Let $\mathscr{E}$ be the error event and l be the DMC selected in S. The error occurs if and only if there is a decoding error while decoding some bit-channel in $\mathscr{L}_X \cup I_l$ of any polar block in the full-height column regime. We index each polar block in the staircase as $b = 1, 2, \ldots, Npk$. Let $\mathscr{E}_g$ be the error event with a genie-aided decoder, which has the accurate values of the past $U^{1:i-1}$ when decoding any bit-channel $i \in \mathscr{L}_X \cup I_l$ for all polar blocks. Let $\mathscr{E}_{ib}$ be the bit-channel error

event for the bit-channel i corresponding to the block b, which is defined as below:

$$\mathscr{E}_{ib} = \{(u^{1:N}, y^{1:N}) \text{ tuples of all the blocks } \tilde{b} \in [Npk] : \\ P_{U_i|U^{1:i-1},Y^{1:N}}(u_i+1|u^{1:i-1}, y^{1:N}) \geq P_{U_i|U^{1:i-1},Y^{1:N}}(u_i|u^{1:i-1}, y^{1:N}) \\ \text{holds for } (u^{1:N}, y^{1:N}) \text{ of block } b\}.$$

If the bit-channel $i \in \mathscr{L}_X \cup I_l$ lies in the full-height column of polar block b, then error event for bit-channel i for genie decoder will be the $\mathscr{E}_{ib}$. If the bit-channel $i \in \mathscr{L}_X \cup I_l$ lies in the non-full-height column of polar block b, then error event for bit-channel i for genie decoder will be null event. This means that $\mathscr{E}_g = \cup_{b\in[Npk]} \cup_{\{i:i\in\mathscr{L}_X\cup I_l \text{ and index } i \text{ of block } b \text{ lies in a full-height column}\}} \mathscr{E}_{ib}$. Note that error event $\mathscr{E}$ will imply at least one of the error events in $\{\mathscr{E}_{ib} : b \in [Npk], i \in \mathscr{L}_X \cup I_l$ such that index i of block b lies in a full-height column$\}$.

So we will have

$$\mathscr{E} \subset \mathscr{E}_g.$$

One the other hand, it is obvious that

$$\mathscr{E}_g \subset \mathscr{E}.$$

Hence we can deduce that

$$\mathscr{E} = \mathscr{E}_g \subset \cup_{b\in[Npk]} \cup_{i\in\mathscr{L}_X\cup I_l} \mathscr{E}_{ib}. \tag{2.14}$$

Let us also define the error event $\mathscr{E}_i$ of bit-channel i for a single polar block $(U^{1:N}, Y^{1:N})$, which is used in the single asymmetric channel polar code:

$$\mathscr{E}_i = \{(u^{1:N}, y^{1:N}) : P_{U_i|U^{1:i-1},Y^{1:N}}(u_i+1|u^{1:i-1}, y^{1:N}) \\ \geq P_{U_i|U^{1:i-1},Y^{1:N}}(u_i|u^{1:i-1}, y^{1:N})\}.$$

We apply union-bound to equation (2.14) followed by taking the expectation. That gives the following upper-bound for the ensemble average probability of error:

$$\mathbb{E}_{W^{1:N}}[P_{e,l}(W^{1:N})] = \mathbb{E}_{W^{1:N}}[\mathbb{P}(\mathscr{E}|W^{1:N})] \leq \sum_{b\in[Npk]} \mathbb{E}_{W^{1:N}}[\mathbb{P}(\cup_{i\in\mathscr{L}_X\cup I_l}\mathscr{E}_{ib}|W^{1:N})]. \tag{2.15}$$

For each block, let us define a measure on $(U^{1:N}, Y^{1:N})$ as follows:

$$\begin{aligned} Q_{U^{1:N},Y^{1:N}}(u^{1:N}, y^{1:N}) = 2^{-|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}\delta_i(u_i|u^{1:i-1})\Pi_{i\in R}P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1}) \\ \cdot \Pi_{i=1}^{N} p_l(y_i|x_i), \end{aligned} \tag{2.16}$$

where $x^{1:N}$ is obtained by applying the polar transform to $u^{1:N}$. Note that $P_{U^{1:N},Y^{1:N}}$ is the measure induced when $X^{1:N}$ is i.i.d. according to $p(x)$ and gets transmitted over the selected DMC l, and $Y^{1:N}$ is received. By using the results from [11] and [24], we have

$$||Q_{U^{1:N},Y^{1:N}} - P_{U^{1:N},Y^{1:N}}|| = O(2^{-N^{\beta'}}) \tag{2.17}$$

for $\beta' < \beta < 0.5$.

Let $\mathscr{E}_b = \cup_{i\in\mathscr{L}_X\cup I_l}\mathscr{E}_{ib}$. Note that

$$\mathbb{P}(\mathscr{E}_b|W^{1:N}) = \sum_{((u^{1:N},y^{1:N}) \text{ tuples of all blocks } [Npk])\in\mathscr{E}_b} \mathbb{P}(\cap_{\tilde{b}\in[Npk]}(U^{1:N} = u^{1:N}, Y^{1:N} = y^{1:N} \text{ of block } \tilde{b})|W^{1:N}).$$

From the definitions of $\mathscr{E}_i$ and $\mathscr{E}_b$, we get

$$\mathbb{P}(\mathscr{E}_b|W^{1:N}) = \sum_{((u^{1:N},y^{1:N}) \text{ of block } b)\in\cup_{i\in\mathscr{L}_X\cup I_l}\mathscr{E}_i} \sum_{((u^{1:N},y^{1:N}) \text{ tuples of blocks } [Npk]-\{b\})} \mathbb{P}(\cap_{\tilde{b}\in[Npk]}(U^{1:N} = u^{1:N}, Y^{1:N} = y^{1:N} \text{ of block } \tilde{b})|W^{1:N}).$$

By marginalizing over the $(U^{1:N}, Y^{1:N})$ tuples of blocks $[Npk] - \{b\}$, we get

$$\begin{aligned}&\mathbb{P}(\mathscr{E}_b|W^{1:N})\\&= \sum_{((u^{1:N},y^{1:N}) \text{ of block } b)\in\cup_{i\in\mathscr{L}_X\cup I_l}\mathscr{E}_i} \mathbb{P}(U^{1:N}=u^{1:N}, Y^{1:N}=y^{1:N} \text{ of block } b|W^{1:N}).\end{aligned}$$

By the chain rule of condition probability and also by the fact that

$$\mathbb{P}(Y^{1:N}=y^{1:N} \text{ of block } b|U^{1:N}=u^{1:N} \text{ of block } b, W^{1:N}) = \Pi_{i=1}^{N} p_l(y_i|x_i),$$

we will have the following:

$$\begin{aligned}&\mathbb{P}(\mathscr{E}_b|W^{1:N})\\&= \sum_{((u^{1:N},y^{1:N}) \text{ of block } b)\in\cup_{i\in\mathscr{L}_X\cup I_l}\mathscr{E}_i} \mathbb{P}(U^{1:N}=u^{1:N} \text{ of block } b|W^{1:N})\Pi_{i=1}^{N} p_l(y_i|x_i).\end{aligned} \tag{2.18}$$

In the term $\Pi_{i=1}^{N} p_l(y_i|x_i)$ in equation (2.18), notice that $x^{1:N}$ vector corresponds to block b, which means it is obtained by applying the polar transform to $u^{1:N}$ vector corresponding to block b and, $y^{1:N}$ vector corresponds to block b. By taking expectation on both sides of equation (2.18) and applying the linearity of expectation, we get

$$\begin{aligned}&\mathbb{E}_{W^{1:N}}[\mathbb{P}(\mathscr{E}_b|W^{1:N})]\\&= \sum_{((u^{1:N},y^{1:N}) \text{ of block } b)\in\cup_{i\in\mathscr{L}_X\cup I_l}\mathscr{E}_i} \mathbb{E}_{W^{1:N}}[\mathbb{P}(U^{1:N}=u^{1:N} \text{ of block } b|W^{1:N})]\Pi_{i=1}^{N} p_l(y_i|x_i).\end{aligned}$$

From equation (2.16) and part 1 of Theorem 1, we get

$$\begin{aligned}&\mathbb{E}_{W^{1:N}}[\mathbb{P}(\mathscr{E}_b|W^{1:N})]\\&= \sum_{((u^{1:N},y^{1:N}) \text{ of block } b)\in\cup_{i\in\mathscr{L}_X\cup I_l}\mathscr{E}_i} Q_{U^{1:N},Y^{1:N}}((u^{1:N},y^{1:N}) \text{ of block } b).\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}_{W^{1:N}}[\mathbb{P}(\mathscr{E}_b|W^{1:N})] &= Q_{U^{1:N},Y^{1:N}}(\cup_{i\in\mathscr{L}_X\cup I_l}\mathscr{E}_i) \\
&\leq ||Q_{U^{1:N},Y^{1:N}} - P_{U^{1:N},Y^{1:N}}|| + P_{U^{1:N},Y^{1:N}}(\cup_{i\in\mathscr{L}_X\cup I_l}\mathscr{E}_i) \\
&\overset{(a)}{\leq} ||Q_{U^{1:N},Y^{1:N}} - P_{U^{1:N},Y^{1:N}}|| + \sum_{i\in\mathscr{L}_X\cup I_l} P_{U^{1:N},Y^{1:N}}(\mathscr{E}_i) \\
&\overset{(b)}{\leq} O(2^{-N^{\beta'}}) + \sum_{i\in\mathscr{L}_X\cup I_l} Z(U_i|U^{1:i-1}Y^{1:N}) \\
&\overset{(c)}{\leq} O(2^{-N^{\beta'}}) + O(N2^{-N^{\beta}}) = O(2^{-N^{\beta'}}).
\end{aligned}$$

Identity (a) follows from the union bound. Identity (b) follows from equation (2.17) and also from the fact that $P(\mathscr{E}_i)$ is upper bounded by $Z(U_i|U^{1:i-1}Y^{1:N})$, as the decision rule for these bit-channels is the MAP decision rule under measure P [46, Proposition 2.7]. Identity (c) follows from polarization results mentioned in Section 2.2. Therefore $\mathbb{E}_{W^{1:N}}[\mathbb{P}(\cup_{i\in\mathscr{L}_X\cup I_l}\mathscr{E}_{ib}|W^{1:N})]$ becomes $O(2^{-N^{\beta'}})$ where $\beta' < \beta < 0.5$. The overall average probability of error will be $O(Npk2^{-N^{\beta'}})$ from equation (2.15). This concludes the proof of part 2.

We now prove part 3 of Theorem 1.

Encoding Complexity: Encoding complexity consists of two factors: encoding the polar block and encoding the RS codeword. Encoding the polar block takes $O(N\log_2(N))$ real operations. Hence the number of operations per bit is $O(\log_2(N))$ real operations. Encoding RS codeword can be done by computing a Fourier transform of length $|\mathscr{H}_X| - |R|$ which takes $O(|\mathscr{H}_X|\log|\mathscr{H}_X|)$ operations over the field $GF(2^p)$ [22]. Addition and multiplication over this field take p and $p^{\log_2(3)}$ binary operations, respectively. Hence there are $p^{\log_2(3)}O(|\mathscr{H}_X|\log_2(|\mathscr{H}_X|))$ binary operations. Therefore, overall RS encoding takes $O((\log_2 N)p^{log_2(3)-1})$ binary operations per bit.

Decoding Complexity: Decoding complexity consists of two factors: decoding the polar block and decoding the RS codeword. Decoding the polar blocks takes $O(N\log_2 N)$ real

operations. Hence the number of operations per bit is $O(\log_2(N))$ real operations. Erasure decoding can be done using error correction decoding algorithms [7, p. 256]. Error correction decoding of a RS codeword here can be done in $O(|\mathscr{H}_X|(\log_2(|\mathscr{H}_X|))^2 \log_2\log_2|\mathscr{H}_X|)$ operations over GF(2^p) [45, p. 216] since the block length of the RS code is $|\mathscr{H}_X| - |R|$. Addition and multiplication over this field take p and $p^{\log_2(3)}$ binary operations, respectively. So there will be $p^{\log_2(3)}O(|\mathscr{H}_X|(\log_2(|\mathscr{H}_X|))^2 \log_2\log_2|\mathscr{H}_X|)$ binary operations. Therefore overall RS erasure decoding takes $O((\log_2(N))^2 \log_2\log_2 N)p^{log_2(3)-1}$ binary operations per bit. This concludes the proof of part 3. □

Proof of Lemma 2:

The proof of Lemma 2 that we provide here follows the proof of Lemma 4 in [1]. First, we have

$$\begin{aligned} Z(X|Y) &= 2\sum_{y\in\mathscr{Y}} \sqrt{P_{X,Y}(0,y)P_{X,Y}(1,y)} \\ &= -1 + \sum_{y\in\mathscr{Y}} \Big[\sum_{x\in\mathscr{X}} \sqrt{P_{X,Y}(x,y)}\Big]^2. \end{aligned}$$

We now use the following form of Minkowsky's inequality that is true when $r < 1$ and a_{jk} is non-negative:

$$\sum_{k\in\mathscr{K}} \Big(\sum_{j\in\mathscr{J}} Q(j)a_{jk}^{\frac{1}{r}}\Big)^r \geq \Big[\sum_{j\in\mathscr{J}} Q(j)\Big(\sum_{k\in\mathscr{K}} a_{jk}\Big)^{\frac{1}{r}}\Big]^r.$$

Then with $r = 0.5$ and $a_{jk} = \sqrt{P^j_{X,Y}(x,y)}$, we get

$$\begin{aligned} Z(X|Y) &\geq -1 + \sum_{y\in\mathscr{Y}}\sum_{j\in\mathscr{J}} Q(j)\Big[\sum_{x\in\mathscr{X}} \sqrt{P^j_{X,Y}(x,y)}\Big]^2 \\ &= \sum_{j\in\mathscr{J}} Q(j)\Big(-1 + \sum_{y\in\mathscr{Y}} \Big[\sum_{x\in\mathscr{X}} \sqrt{P^j_{X,Y}(x,y)}\Big]^2\Big) \\ &= \sum_{j\in\mathscr{J}} Q(j)Z^j(X|Y). \end{aligned}$$

□

Proof of Proposition 1:

The proof of the proposition follows from the proof of [46, Lemma 2.9]. We have

$$\begin{aligned} P_{U_1,U_2,Y_1,Y_2}(u_1,u_2,y_1,y_2) &= P_{X_1,X_2,Y_1,Y_2}(u_1+u_2,u_2,y_1,y_2) \\ &= P_{X_1,Y_1}(u_1+u_2,y_1)P_{X_2,Y_2}(u_2,y_2). \end{aligned} \tag{2.19}$$

We also have

$$\begin{aligned} P_{U_1,Y_1,Y_2}(u_1,y_1,y_2) &= \sum_{u_2} P_{U_1,U_2,Y_1,Y_2}(u_1,u_2,y_1,y_2) \\ &= \sum_{u_2} P_{X_1,Y_1}(u_1+u_2,y_1)P_{X_2,Y_2}(u_2,y_2). \end{aligned} \tag{2.20}$$

Now we evaluate the upper bound for Bhattacharyya parameter $Z(U_1|Y_1,Y_2)$ as follows:

$$\begin{aligned} &Z(U_1|Y_1,Y_2) \\ &= 2\sum_{y_1y_2}\sqrt{P_{U_1,Y_1,Y_2}(0,y_1,y_2)P_{U_1,Y_1,Y_2}(1,y_1,y_2)} \\ &\overset{(a)}{=} 2\sum_{y_1y_2}\big((\sum_{u_2}P_{X_1,Y_1}(u_2,y_1)P_{X_2,Y_2}(u_2,y_2))(\sum_{v_2}P_{X_1,Y_1}(1+v_2,y_1)P_{X_2,Y_2}(v_2,y_2))\big)^{0.5} \\ &\leq 2\sum_{y_1y_2}\sum_{u_2v_2}\big(P_{X_1,Y_1}(u_2,y_1)P_{X_1,Y_1}(1+v_2,y_1)P_{X_2,Y_2}(u_2,y_2)P_{X_2,Y_2}(v_2,y_2)\big)^{0.5} \\ &= \sum_{u_2v_2}\big(2\sum_{y_1y_2}\big(P_{X_1,Y_1}(u_2,y_1)P_{X_1,Y_1}(1+v_2,y_1)P_{X_2,Y_2}(u_2,y_2)P_{X_2,Y_2}(v_2,y_2)\big)^{0.5}\big) \\ &\overset{(b)}{\leq} Z(X_1|Y_1)+Z(X_2|Y_2). \end{aligned}$$

Identity (a) follows from equation (2.20). Identity (b) follows because when $u_2 = v_2$, the term inside the outermost summation becomes $Z(X_1|Y_1)P_{X_2}(u_2)$ and when $u_2 = v_2+1$, the term inside the outermost summation becomes $Z(X_2|Y_2)P_{X_1}(u_2)$.

The joint distribution $P_{U_1Y_1Y_2}(u_1,y_1,y_2)$ can be expressed as

$$\begin{aligned}
P_{U_1Y_1Y_2}(u_1,y_1,y_2) &= \sum_{u_2\in\mathscr{X}} P_{U_1U_2Y_1Y_2}(u_1,u_2,y_1,y_2) \\
&= P_{U_1U_2Y_1Y_2}(u_1,0,y_1,y_2) + P_{U_1U_2Y_1Y_2}(u_1,1,y_1,y_2) \\
&= P_{U_2}(0)P_{U_1Y_1Y_2|U_2}(u_1,y_1,y_2|0) + P_{U_2}(1)P_{U_1Y_1Y_2|U_2}(u_1,y_1,y_2|1).
\end{aligned}$$

Let

$$P^1_{U_1Y_1Y_2}(u_1,y_1,y_2) = P_{U_1Y_1Y_2|U_2}(u_1,y_1,y_2|0),$$

$$P^2_{U_1Y_1Y_2}(u_1,y_1,y_2) = P_{U_1Y_1Y_2|U_2}(u_1,y_1,y_2|1)$$

be the two joint distributions on random variable triplet (U_1,Y_1,Y_2). We now evaluate the Bhattacharyya parameter corresponding to the distribution $P^1_{U_1Y_1Y_2}(u_1,y_1,y_2)$.

$$\begin{aligned}
Z^1(U_1|Y_1Y_2) &= 2\sum_{y_1y_2} P^1_{Y_1Y_2}(y_1y_2)\sqrt{P^1_{U_1|Y_1Y_2}(0|y_1y_2)P^1_{U_1|Y_1Y_2}(1|y_1y_2)} \\
&= 2\sum_{y_1y_2} P_{Y_1Y_2|U_2}(y_1y_2|0)\sqrt{P_{U_1|Y_1Y_2U_2}(0|y_1y_20)P_{U_1|Y_1Y_2U_2}(1|y_1y_20)} \\
&\overset{(a)}{=} 2\sum_{y_1y_2} P_{Y_1}(y_1)P_{Y_2|U_2}(y_2|0)\sqrt{P_{X_1|Y_1Y_2U_2}(0|y_1y_20)P_{X_1|Y_1Y_2U_2}(1|y_1y_20)} \\
&\overset{(b)}{=} 2\sum_{y_1y_2} P_{Y_1}(y_1)P_{Y_2|U_2}(y_2|0)\sqrt{P_{X_1|Y_1}(0|y_1)P_{X_1|Y_1}(1|y_1)} \\
&= 2\sum_{y_1} P_{Y_1}(y_1)\sqrt{P_{X_1|Y_1}(0|y_1)P_{X_1|Y_1}(1|y_1)} \\
&= Z(X_1|Y_1).
\end{aligned}$$

Identity (a) is true because Y_1 is independent of U_2 and also because Y_2 is independent of Y_1 given U_2. Identity (b) is true because X_1 is independent of Y_2U_2 given Y_1. Similarly we can easily prove that $Z^2(U_1|Y_1Y_2) = Z(X_1|Y_1)$. Now Lemma 2 implies that $Z(U_1|Y_1Y_2) \geq Z(X_1|Y_1)$. By exchanging the roles of (X_1,Y_1) and (X_2,Y_2), we can also get $Z(U_1|Y_1Y_2) \geq Z(X_2|Y_2)$. Therefore

$Z(U_1|Y_1Y_2) \geq \max\{Z(X_2|Y_2), Z(X_1|Y_1)\}$. This concludes the proof of part 1.

Now we evaluate the Bhattacharyya parameter $Z(U_2|Y_1Y_2U_1)$ as follows:

$$
\begin{aligned}
&Z(U_2|Y_1Y_2U_1) = \\
&= 2 \sum_{y_1y_2u_1} \sqrt{P_{U_1,U_2,Y_1,Y_2}(u_1,0,y_1,y_2) P_{U_1,U_2,Y_1,Y_2}(u_1,1,y_1,y_2)} \\
&\overset{(a)}{=} 2 \sum_{y_1y_2u_1} \left[P_{X_1,Y_1}(u_1,y_1) P_{X_2,Y_2}(0,y_2) P_{X_1,Y_1}(u_1+1,y_1) P_{X_2,Y_2}(1,y_2)\right]^{0.5} \\
&= 2 \sum_{y_1y_2u_1} \left[P_{X_1,Y_1}(u_1,y_1) P_{X_1,Y_1}(u_1+1,y_1) P_{X_2,Y_2}(0,y_2) P_{X_2,Y_2}(1,y_2)\right]^{0.5} \\
&= \sum_{u_1 \in \mathscr{X}} \sum_{y_1 \in \mathscr{Y}_1} \left(2 \sum_{y_2 \in \mathscr{Y}_2} \left[P_{X_1,Y_1}(u_1,y_1) P_{X_1,Y_1}(u_1+1,y_1) P_{X_2,Y_2}(0,y_2) P_{X_2,Y_2}(1,y_2)\right]^{0.5}\right) \\
&= Z(X_1|Y_1) Z(X_2|Y_2).
\end{aligned}
$$

Identity (a) follows from (2.19). Since the Bhattacharyya parameter is always less than or equal to 1, it also follows that $Z(U_2|Y_1Y_2U_1) \leq \min\{Z(X_1|Y_1), Z(X_2|Y_2)\}$. This concludes the proof of part 2. □

Proof of Lemma 3:

$$
\begin{aligned}
||P-Q|| &= \sum_{(x_1,x_2)} \frac{1}{2} |P(x_1,x_2) - Q(x_1,x_2)| \\
&= \sum_{(x_1,x_2)} \frac{1}{2} |P_1(x_1)P_2(x_2) - Q_1(x_1)Q_2(x_2)| \\
&= \sum_{(x_1,x_2)} \frac{1}{2} |P_1(x_1)P_2(x_2) - Q_1(x_1)P_2(x_2) + Q_1(x_1)P_2(x_2) - Q_1(x_1)Q_2(x_2)|.
\end{aligned}
$$

By the triangular inequality, we now get

$$
\begin{aligned}
||P-Q|| &\leq \sum_{(x_1,x_2)} \left(\frac{1}{2}|P_1(x_1)P_2(x_2) - Q_1(x_1)P_2(x_2)| + \frac{1}{2}|Q_1(x_1)P_2(x_2) - Q_1(x_1)Q_2(x_2)|\right) \\
&= \frac{1}{2} \sum_{(x_1,x_2)} P_2(x_2)|P_1(x_1) - Q_1(x_1)| + \frac{1}{2} \sum_{(x_1,x_2)} Q_1(x_1)|P_2(x_2) - Q_1(x_2)|
\end{aligned}
$$

$$= \frac{1}{2}\sum_{x_1}|P_1(x_1) - Q_1(x_1)| + \frac{1}{2}\sum_{x_2}|P_2(x_2) - Q_1(x_2)|$$

$$= ||P_1 - Q_1|| + ||P_2 - Q_2||$$

$$= \varepsilon_1 + \varepsilon_2. \quad \square$$

Proof of Lemma 4:

$$\begin{aligned}
||Q_{X,Y} - P_{X,Y}|| &= \sum_{(x,y):P_{X,Y}(x,y)>Q_{X,Y}(x,y)} P_{X,Y}(x,y) - Q_{X,Y}(x,y) \\
&= \sum_{(x,y):P_X(x)p(y|x)>Q_X(x)p(y|x)} P_X(x)p(y|x) - Q_X(x)p(y|x) \\
&= \sum_{(x,y):P_X(x)>Q_X(x)} (P_X(x) - Q_X(x))p(y|x) \\
&= \sum_{x:P_X(x)>Q_X(x)} \sum_{y} (P_X(x) - Q_X(x))p(y|x) \\
&= \sum_{x:P_X(x)>Q_X(x)} (P_X(x) - Q_X(x)) \\
&= ||Q_X - P_X||. \quad \square
\end{aligned}$$

Proof of Lemma 5: The Bhattacharyya parameter of the new bit-channel produced with U'_{x_j} as input and $U^{1:x_j-1}V^{1:y_j-1}Y^{1:2N}$ as output will be lower bounded as follows:

$$\begin{aligned}
Z(U'_{x_j}|U^{1:x_j-1}V^{1:y_j-1}Y^{1:2N}) &\overset{(a)}{\geq} \max\{Z(U_{x_j}|U^{1:x_j-1}Y^{1:N}), Z(V_{y_j}|V^{1:y_j-1}Y^{N+1:2N})\} \\
&\overset{(b)}{\geq} 1 - 2^{-N^\beta}.
\end{aligned}$$

Identity (a) is true by Proposition 1. Identity (b) follows, as either $Z(U_{x_j}|U^{1:x_j-1}Y^{1:N})$ (if DMC 2 is selected in S) will be greater than $1-2^{-N^\beta}$ or $Z(V_{y_j}|V^{1:y_j-1}Y^{N+1:2N})$ (if DMC 1 is selected in S) will be greater than $1-2^{-N^\beta}$. This completes the proof of part 1.

The Bhattacharyya parameter of the new bit-channel produced with V'_{y_j} as input and

$U^{1:x_j-1}V^{1:y_j-1}U'_{x_j}Y^{1:2N}$ as output will be upper bounded as follows:

$$Z(V'_{y_j}|U^{1:x_j-1}V^{1:y_j-1}U'_{x_j}Y^{1:2N}) \overset{(a)}{=} Z(U_{x_j}|U^{1:x_j-1}Y^{1:N})Z(V_{y_j}|V^{1:y_j-1}Y^{N+1:2N}) \overset{(b)}{\leq} 2^{-N^\beta}.$$

Identity (a) is true by Proposition 1. Identity (b) follows, as either $Z(U_{x_j}|U^{1:x_j-1}Y^{1:N})$ (if DMC 1 is selected in S) will be less than 2^{-N^β} or $Z(V_{y_j}|U^{1:y_j-1}Y^{N+1:2N})$ (if DMC 2 is selected in S) will be less than 2^{-N^β}. This completes the proof of part 2. □

Proof of Theorem 2:

Let the linear bijective transform which maps $(U'^{1:N}\ \ V'^{1:N})$ to $(U^{1:N}\ \ V^{1:N})$ be H_{2N}. Let the word $(u^{1:N}\ \ v^{1:N})$ be obtained by applying H_{2N} to the word $(u'^{1:N}\ \ v'^{1:N})$. The probability that the word $(U'^{1:N}\ \ V'^{1:N})$= $(u'^{1:N}\ \ v'^{1:N})$ and received vector $Y^{1:2N}=y^{1:2N}$ when DMC l in S is selected will be

$$\begin{aligned} &2^{-(2|I_1\cap I_2|+|\mathscr{B}|)}\mathbb{1}[\cap_{i\in(\mathscr{H}_X-(I_1\cap I_2))}\{f_1(i)=u'_i\}]\mathbb{1}[\cap_{i\in(\mathscr{H}_X-((I_1\cap I_2)\cup\mathscr{B}))}\{f_2(i)=v'_i\}] \\ &\cdot\mathbb{1}[\cap_{i\in R}\{\lambda^1(u^{1:i-1})=u_i\}]\mathbb{1}[\cap_{i\in R}\{\lambda^2(v^{1:i-1})=v_i\}]\Pi_{i\in\mathscr{L}_X}(\delta_i(u_i|u^{1:i-1})\delta_i(v_i|v^{1:i-1})) \\ &\cdot P_{Y^{1:N}|U^{1:N}}(y^{1:N}|u^{1:N})P_{Y^{1:N}|U^{1:N}}(y^{N+1:2N}|v^{1:N}). \end{aligned} \tag{2.21}$$

Note that we used the fact that $P_{U^{1:N}Y^{1:N}}=P_{V^{1:N}Y^{N+1:2N}}$. Let $\mathscr{E}_i^b$ be the error event for the ith bit-channel of block b.

For $i\in I_1\cap I_2$, we define the error events as follows:

$$\mathscr{E}_i^1=\{(u'^{1:N},v'^{1:N},y^{1:2N}):P_{U_i|U^{1:i-1}Y^{1:N}}(u_i+1|u^{1:i-1}y^{1:N}) \geq P_{U_i|U^{1:i-1}Y^{1:N}}(u_i|u^{1:i-1}y^{1:N})\},$$

$$\mathscr{E}_i^2=\{(u'^{1:N},v'^{1:N},y^{1:2N}):P_{U_i|U^{1:i-1}Y^{1:N}}(v_i+1|v^{1:i-1}y^{1:N})$$

$$\geq P_{U_i|U^{1:i-1}Y^{1:N}}(v_i|v^{1:i-1}y^{1:N})\}.$$

For $j \in [G]$, we define the error event as follows:

$$\begin{aligned}\mathscr{E}^2_{y_j} = \{(u'^{1:N}, v'^{1:N}, y^{1:2N}) : P_{V'_{y_j}|U^{1:x_j-1}U'_{x_j}V^{1:y_j-1}Y^{1:2N}}(v_i+1|u^{1:x_j-1}u'_{x_j}v^{1:y_j-1}y^{1:2N}) \\ \geq P_{V'_{y_j}|U^{1:x_j-1}U'_{x_j}V^{1:y_j-1}Y^{1:2N}}(v_i|u^{1:x_j-1}u'_{x_j}v^{1:y_j-1}y^{1:2N})\}.\end{aligned} \tag{2.22}$$

Therefore the error event $\mathscr{E}$ becomes

$$\mathscr{E} = \{\cup_{i\in I_1\cap I_2}\mathscr{E}^1_i\} \cup \{\cup_{i\in(I_1\cap I_2)\cup\mathscr{B}}\mathscr{E}^2_i\}. \tag{2.23}$$

The probability of error for the given $f_1, f_2, \lambda^1_R, \lambda^2_R$ will be

$$\begin{aligned}P_{e,l}(\lambda^1_R, \lambda^2_R, f_1, f_2) = \sum_{(u'^{1:N}, v'^{1:N}, y^{1:2N})} & 2^{-(2|I_1\cap I_2|+|\mathscr{B}|)}\mathbb{1}[\cap_{i\in(\mathscr{H}_X-(I_1\cap I_2))}\{f_1(i) = u'_i\}] \\ & \cdot \mathbb{1}[\cap_{i\in(\mathscr{H}_X-((I_1\cap I_2)\cup\mathscr{B}))}\{f_2(i) = v'_i\}] \\ & \cdot \mathbb{1}[\cap_{i\in R}\{\lambda^1(u^{1:i-1}) = u_i\}]\mathbb{1}[\cap_{i\in R}\{\lambda^2(v^{1:i-1}) = v_i\}] \\ & \cdot \Pi_{i\in\mathscr{L}_X}(\delta_i(u_i|u^{1:i-1})\delta_i(v_i|v^{1:i-1}))\cdot \\ & \cdot P_{Y^{1:N}|U^{1:N}}(y^{1:N}|u^{1:N})P_{Y^{1:N}|U^{1:N}}(y^{N+1:2N}|v^{1:N}) \\ & \cdot \mathbb{1}[(u'^{1:N}, v'^{1:N}, y^{1:2N}) \in \mathscr{E}].\end{aligned} \tag{2.24}$$

By linearity of expectation and independence of random functions $(\lambda^1, \lambda^2, f_1, f_2)$, the ensemble

expectation of $P_{e,l}$ will become

$$\begin{aligned}\mathbb{E}[P_{e,l}(\lambda_R^1,\lambda_R^2,f_1,f_2)] = \sum_{(u'^{1:N},v'^{1:N},y^{1:2N})} & 2^{-2|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}(\delta_i(u_i|u^{1:i-1})\delta_i(v_i|v^{1:i-1})) \\ & \cdot\Pi_{i\in R}(P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1})P_{U_i|U^{1:i-1}}(v_i|v^{1:i-1})) \\ & \cdot P_{Y^{1:N}|U^{1:N}}(y^{1:N}|u^{1:N})P_{Y^{1:N}|U^{1:N}}(y^{N+1:2N}|v^{1:N}) \\ & \cdot\mathbb{1}[(u'^{1:N},v'^{1:N},y^{1:2N})\in\mathscr{E}].\end{aligned} \tag{2.25}$$

Now we define the measure Q on random variables $U'^{1:N}V'^{1:N}Y^{1:2N}$ as

$$\begin{aligned}Q_{U'^{1:N}V'^{1:N}Y^{1:2N}}(u'^{1:N},v'^{1:N},y^{1:2N}) := & 2^{-2|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}(\delta_i(u_i|u^{1:i-1})\delta_i(v_i|v^{1:i-1})) \\ & \cdot\Pi_{i\in R}(P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1})P_{U_i|U^{1:i-1}}(v_i|v^{1:i-1})) \\ & \cdot P_{Y^{1:N}|U^{1:N}}(y^{1:N}|u^{1:N})P_{Y^{1:N}|U^{1:N}}(y^{N+1:2N}|v^{1:N}).\end{aligned} \tag{2.26}$$

Note that

$$Q_{U'^{1:N}V'^{1:N}Y^{1:2N}}(u'^{1:N},v'^{1:N},y^{1:2N}) = Q_{U^{1:N}V^{1:N}Y^{1:2N}}(u^{1:N},v^{1:N},y^{1:2N}).$$

From equations (2.25) and (2.26), we have

$$Q_{U'^{1:N}V'^{1:N}Y^{1:2N}}(\mathscr{E}) = \mathbb{E}[P_{e,l}(\lambda_R^1,\lambda_R^2,f_1,f_2)]. \tag{2.27}$$

By marginalizing the distribution in equation (2.26) over the random variables $(V^{1:N},Y^{N+1:2N})$ and $(U^{1:N},Y^{1:N})$, respectively, we will have

$$\begin{aligned}Q_{U^{1:N}Y^{1:N}}(u^{1:N},y^{1:N}) = & 2^{-|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}\delta_i(u_i|u^{1:i-1})\Pi_{i\in R}P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1}) \\ & \cdot P_{Y^{1:N}|U^{1:N}}(y^{1:N}|u^{1:N}).\end{aligned} \tag{2.28}$$

$$Q_{V^{1:N}Y^{N+1:2N}}(v^{1:N},y^{N+1:2N}) = 2^{-|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}\delta_i(v_i|v^{1:i-1})\Pi_{i\in R}P_{U_i|U^{1:i-1}}(v_i|v^{1:i-1}) \cdot P_{Y^{1:N}|U^{1:N}}(y^{N+1:2N}|v^{1:N}). \tag{2.29}$$

Clearly,

$$Q_{U^{1:N}V^{1:N}Y^{1:2N}}(u^{1:N},v^{1:N},y^{1:N}) = Q_{U^{1:N}Y^{1:N}}(u^{1:N},y^{1:N})Q_{V^{1:N}Y^{N+1:2N}}(v^{1:N},y^{N+1:2N}).$$

Therefore $(U^{1:N},Y^{1:N})$ and $(V^{1:N},Y^{N+1:2N})$ are i.i.d. with respect to measure Q. Therefore, by using the fact that $Q_{U^{1:N}Y^{1:N}} = Q_{V^{1:N}Y^{N+1:2N}}$, and equation (2.11), we get

$$||Q_{U^{1:N}Y^{1:N}} - P_{U^{1:N}Y^{1:N}}|| = ||Q_{V^{1:N}Y^{N+1:2N}} - P_{V^{1:N}Y^{N+1:2N}}||. \tag{2.30}$$

We bound the probability of error as follows:

$$\begin{aligned} Q_{U'^{1:N}V'^{1:N}Y^{1:2N}}(\mathscr{E}) &\leq ||Q_{U'^{1:N}V'^{1:N}Y^{1:2N}} - P_{U'^{1:N}V'^{1:N}Y^{1:2N}}|| + P_{U'^{1:N}V'^{1:N}Y^{1:2N}}(\mathscr{E}) \\ &\leq ||Q_{U^{1:N}V^{1:N}Y^{1:2N}} - P_{U^{1:N}V^{1:N}Y^{1:2N}}|| + P_{U^{1:N}V^{1:N}Y^{1:2N}}(\mathscr{E}). \end{aligned}$$

From equation (2.23) and by using union bound, we get

$$\begin{aligned} Q_{U'^{1:N}V'^{1:N}Y^{1:2N}}(\mathscr{E}) \leq ||Q_{U^{1:N}V^{1:N}Y^{1:2N}} - P_{U^{1:N}V^{1:N}Y^{1:2N}}|| &+ \sum_{i\in I_1\cap I_2} P_{U'^{1:N}V'^{1:N}Y^{1:2N}}(\mathscr{E}_i^1) \\ &+ \sum_{i\in(I_1\cap I_2)\cup\mathscr{B}} P_{U'^{1:N}V'^{1:N}Y^{1:2N}}(\mathscr{E}_i^2). \end{aligned} \tag{2.31}$$

Now we bound each of the three terms of the summation in the right hand side of the inequality. We bound the first term of the summation as follows:

$||Q_{U^{1:N}V^{1:N}Y^{1:2N}} - P_{U^{1:N}V^{1:N}Y^{1:2N}}||$

$$\begin{aligned} &\overset{(a)}{\leq} ||Q_{V^{1:N}Y^{N+1:2N}} - P_{V^{1:N}Y^{N+1:2N}}|| + ||Q_{U^{1:N}Y^{1:N}} - P_{U^{1:N}Y^{1:N}}|| \\ &\overset{(b)}{=} 2||Q_{U^{1:N}Y^{1:N}} - P_{U^{1:N}Y^{1:N}}|| \overset{(c)}{=} O(2^{-N^{\beta'}}). \end{aligned} \tag{2.32}$$

Identity (a) follows from the fact that $(U^{1:N}, Y^{1:N})$ and $(V^{1:N}, Y^{N+1:2N})$ are independent with respect to measures Q and P, coupled with application of Lemma 3. Identity (b) follows from equation (2.30). Identity (c) follows from the fact that $||Q_{U^{1:N},Y^{1:N}} - P_{U^{1:N},Y^{1:N}}|| = O(2^{-N^{\beta'}})$ for $\beta' < \beta < 0.5$.

For $i \in I_1 \cap I_2$, $P(\mathscr{E}_i^b)$ is bounded below as in equation (60) in [24] for $b \in \{1,2\}$:

$$P(\mathscr{E}_i^b) \leq 2^{-N^{\beta}}. \tag{2.33}$$

For $i \in \mathscr{B}$, there exists a $j \in [G]$ such that $i = y_j$. For such a bit-channel i, note that $P(\mathscr{E}_i^2)$ is upper bounded by $Z(V'_{y_j}|U'^{1:x_j-1}V'^{1:y_j-1}Y^{1:2N})$ as the decision rule for these bit-channels is the MAP decision rule under measure P [46, Proposition 2.7]. Therefore, from Lemma 5, we get

$$P_{U'^{1:N},V'^{1:N}Y^{1:2N}}(\mathscr{E}_i^2) \leq Z(V'_{y_j}|U'^{1:x_j-1}V'^{1:y_j-1}Y^{1:2N}) \leq 2^{-N^{\beta}}. \tag{2.34}$$

From equations (2.31), (2.32), (2.33) and (2.34), we conclude that

$$\mathbb{E}[P_{e,l}(\lambda_R^1, \lambda_R^2, f_1, f_2)] = O(2^{-N^{\beta'}})$$

for each $l \in \{1,2\}$. □

Acknowledgement

This chapter is in part a reprint of the material in the paper: Karthik Nagarjuna Tunuguntla, Paul H. Siegel, "Universal polar coding for asymmetric channels," *2018 IEEE Information Theory Workshop (ITW)*, pp. 1-5, Guangzhou, China, November 2018. Dissertation author is the primary contributor of the paper.

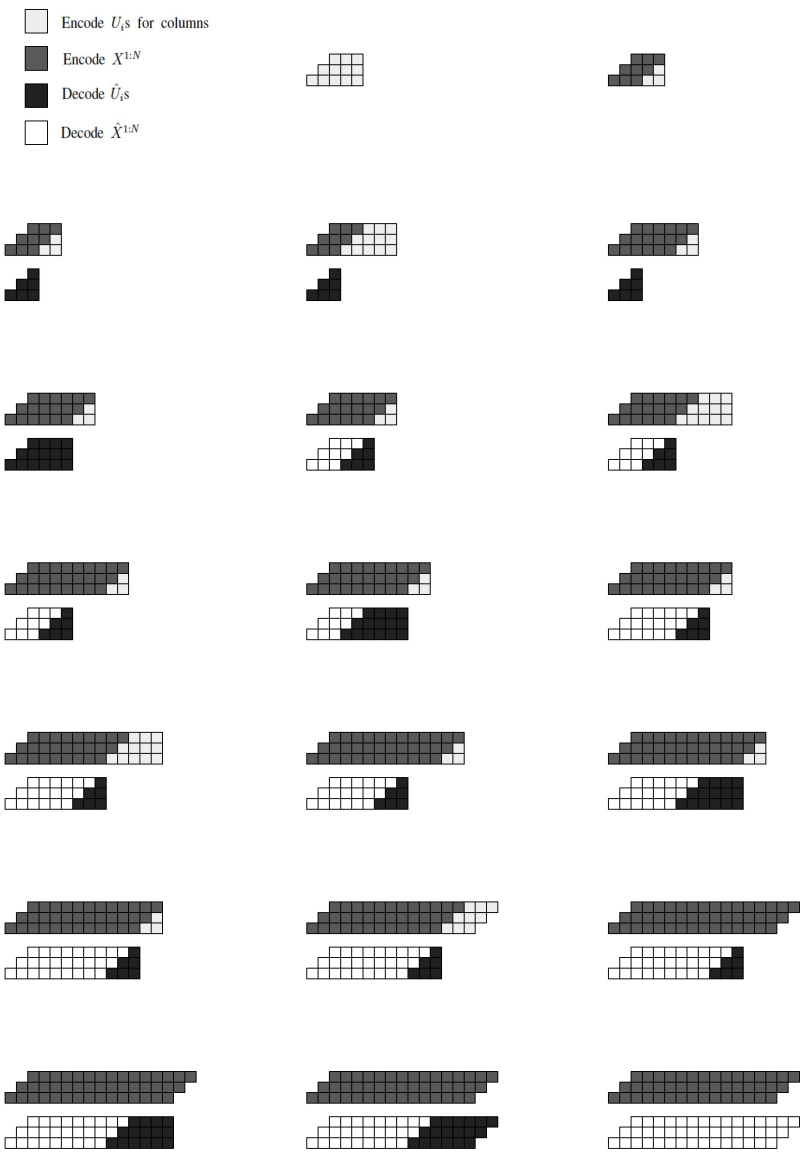

Figure 2.4. Continuous encoding and decoding of sub-staircases when $N = 3, k = 5$ and $p = 1$

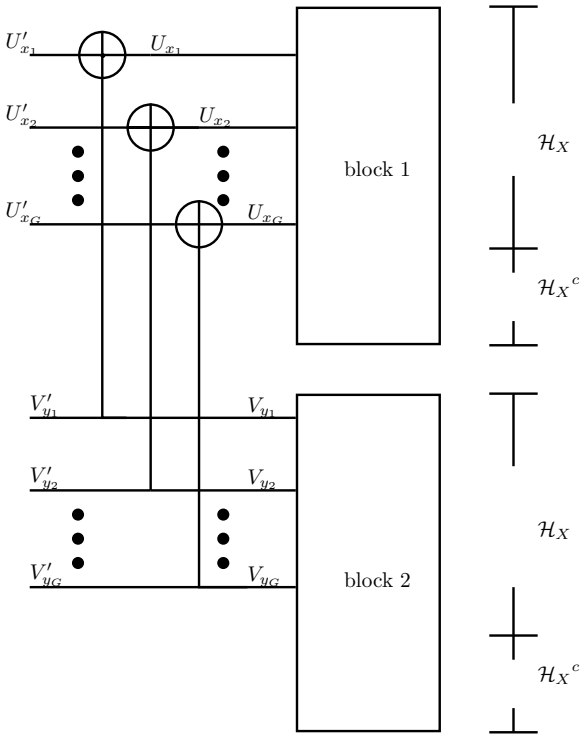

Figure 2.5. Combining two independent original polar blocks for universalization

Chapter 3

Polar Shaping Codes for Costly Noiseless and Noisy Channels

3.1 Introduction

Shaping codes encode information for use on costly channels, i.e., channels with symbol costs subject to an average cost constraint. Their conceptual origins can be traced to Shannon's classic 1948 paper [49]. Prominent applications include data transmission with a power constraint [16] and, more recently, data storage on flash memories [29] and efficient strand synthesis for DNA-based storage [28]. Codes that minimize average cost per symbol for a given rate and codes that minimize average symbol cost per source symbol (or total cost) have been investigated, as has their application to noiseless and noisy costly channels. See [29] for further references.

Arıkan [1] constructed capacity-achieving polar codes for binary input symmetric channels. Arıkan also introduced source polarization, which served as the basis for source coding for non-uniform source alphabets [3]. A capacity-achieving coding scheme based on source and channel polarization for binary input asymmetric channels was proposed by Honda and Yamamoto [24]. In this scheme, complex boolean functions are shared between encoder and decoder for non-information carrying bit-channels. The use of common randomness is proposed to avoid these complex boolean functions [24]. En Gad et al. [15] used randomized rounding for low-entropy and not-completely polarized bit-channels. In addition, a side channel was used to reliably transmit bits corresponding to not completely polarized bit-channels, whose

fraction is vanishing with respect to the block length. A proof that argmax can be used to encode low-entropy bit-channels is given by Chou and Bloch [11]. We proposed a staircase scheme [37] that avoids both common randomness and complex boolean functions to encode not-completely polarized bit-channels.

In this paper, we consider polar code design for costly memoryless channels, both noiseless and noisy. Shaping can also be viewed as a dual problem to source coding by converting information into symbols satisfying specified probabilistic properties, and we also adopt this perspective.

We first propose a polar shaping code design for a (costly) noiseless channel and a specified symbol probability distribution. The construction is an adaptation of the Honda and Yamamoto polar coding scheme for asymmetric channels [24]. The total cost of the proposed shaping code approaches the minimum possible value when the code is designed with the optimal rate and symbol distribution [29].

We then study shaping codes for costly noisy discrete memoryless channels (DMCs). This model is relevant to the design of efficient codes that combine shaping and error correction for use in a noisy transmission or storage system. We first give an upper bound on the rate that can be achieved on the DMC with a specified symbol occurrence probability distribution on codewords. Then we formulate an optimization problem whose solution gives a lower bound on the optimal total cost for the channel. (Note that the maximum rate achieved with a constraint on the average cost per code symbol has been investigated by Böcherer [8].) Finally, we show that polar codes for asymmetric channels [24] can be used to design shaping codes for costly noisy DMCs so that the total cost of the proposed code approaches the lower bound as the block length grows. The construction uses common randomness for encoding frozen bit-channels and not-completely polarized bit-channels in the code construction. Common randomness is crucial to get the desired shaping distribution on the codeword symbols.

We also show that the optimal total cost can be achieved by using random code construction methods, randomly choosing frozen bits and randomly choosing boolean functions

for not-completely polarized channels [37], [24], and thereby avoiding the need for common randomness. For such a random code construction, we show that, with high probability, there exist codes in the random ensemble whose costs approach the optimal total cost with diminishing probability of error.

We note that a scheme that combines polar-coded modulation and probabilistic amplitude shaping [9] was introduced by Prinz et al. [44], and a novel constellation shaping based on polar-coded modulation was proposed by Matsumine [31].

3.2 Preliminaries

We denote the alphabet of the costly channel by $\mathscr{X}$. We denote the output alphabet of the costly noisy DMC by $\mathscr{Y}$. We express any set of random variables $X_i, X_{i+1}, \ldots, X_j$ $(i < j)$ by a row vector $(X_i, X_{i+1}, \ldots, X_j)$ which is denoted by $X^{i:j}$. We denote the set $\{1, 2, 3, \ldots, N\}$ by $[N]$. Let $U^{1:N}$ be a row vector and let $\mathscr{A} \subset [N]$. $U^{\mathscr{A}}$ denotes the row vector consisting of elements in $U^{1:N}$ corresponding to the subset of positions $\mathscr{A}$ in the same order. Let P and Q be any two distributions on a discrete arbitrary alphabet $\mathscr{Z}$. We denote the total variation distance between the two distributions P and Q as $||P-Q||$. Therefore $||P-Q|| = \sum_{z\in\mathscr{Z}} \frac{1}{2}|P(z)-Q(z)| = \sum_{z:P(z)>Q(z)} P(z)-Q(z)$. We denote the KL-divergence between two distributions P and Q as $D(P||Q)$.

Let X be the random variable distributed as $p(x)$ over alphabet $\mathscr{X}$. In this paper, we provide polar shaping codes for binary alphabets. So we let $\mathscr{X} = \{0, 1\}$ to introduce polarization results. Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)$ be i.i.d. random tuples distributed according to $p(x)p(y|x)$ and $N = 2^n$. Let G_N be the conventional polar transformation [1], represented by a binary matrix of dimension $N \times N$. Let $U^{1:N} = X^{1:N}G_N$. We denote $\mathbb{P}(U^{1:N} = u^{1:N})$ by $P_{U^{1:N}}(u^{1:N})$ and similarly we denote $\mathbb{P}(U_i = u_i | U^{1:i-1}Y^{1:N} = u^{1:i-1}y^{1:N})$ by $P_{U_i|U^{1:i-1}Y^{1:N}}(u_i|u^{1:i-1}y^{1:N})$.

For two random variables (X, Y) distributed as $p(x)p(y|x)$, the Bhattacharya parameter

is defined as

$$Z(X|Y) = 2\sum_{y} P_Y(y)\sqrt{P_{X|Y}(1|y)P_{X|Y}(0|y)}.$$

Let $\beta < 0.5$ and define the following subsets, with notation adapted from [15].

$$\begin{aligned}
\mathscr{H}_X &= \{i \in [N] : Z(U_i|U^{1:(i-1)}) \geq 1 - 2^{-N^\beta}\}. \\
\mathscr{L}_X &= \{i \in [N] : Z(U_i|U^{1:(i-1)}) \leq 2^{-N^\beta}\}. \\
\mathscr{H}_{X|Y} &= \{i \in [N] : Z(U_i|U^{1:(i-1)}Y^{1:N}) \geq 1 - 2^{-N^\beta}\}. \\
\mathscr{L}_{X|Y} &= \{i \in [N] : Z(U_i|U^{1:(i-1)}Y^{1:N}) \leq 2^{-N^\beta}\}.
\end{aligned}$$

Note that $\mathscr{L}_X \subseteq \mathscr{L}_{X|Y}$. From Theorem 1 in [24], we have the following polarization results.

$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N}|\mathscr{H}_X| = H(X), \lim_{N\to\infty} \frac{1}{N}|\mathscr{L}_X| = 1 - H(X), \\
&\lim_{N\to\infty} \frac{1}{N}|\mathscr{H}_{X|Y}| = H(X|Y), \\
&\lim_{N\to\infty} \frac{1}{N}|\mathscr{L}_{X|Y}| = 1 - H(X|Y).
\end{aligned}$$

We define several other subsets of bit-channels as follows:

$$I = \mathscr{H}_X \cap \mathscr{L}_{X|Y}, F = \mathscr{H}_X \cap \mathscr{L}^c_{X|Y}, S = (\mathscr{H}_X \cup \mathscr{L}_X)^c.$$

We refer to these as good, bad, and not completely polarized bit-channels respectively. We refer to bit-channels in $\mathscr{H}_X$ and bit-channels in $\mathscr{L}_X$ as high-entropy bit-channels and low-entropy bit-channels respectively. The size of set S is a vanishing fraction with respect to the block length as N increases due to polarization. From [24, Theorem 1],

$$\lim_{N\to\infty} \frac{|I|}{N} = I(X;Y). \tag{3.1}$$

For a codeword of length N, we define the symbol frequency function $f_j : \mathscr{X}^N \to [0,1]$

for $j \in \mathscr{X}$ as follows:

$$f_j(x^{1:N}) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(x_i = j).$$

A costly channel consists of a finite discrete alphabet $\mathscr{X}$ and a cost function $C : \mathscr{X} \to R^+$ that associates cost $C(x)$ to each symbol $x \in \mathscr{X}$. A costly noisy DMC additionally contains an alphabet $\mathscr{Y}$ for the noisy output and transition probabilities $p(y|x)$ for each $x \in \mathscr{X}$ and $y \in \mathscr{Y}$.

3.3 Polar shaping code

3.3.1 Code construction

In this section, we provide shaping code that transforms uniformly distributed message $M^{1:|\mathscr{H}_X|}$ ($|\mathscr{H}_X|$ bits) into $X^{1:N}$, whose distribution is close to the distribution induced when $X^{1:N}$ is i.i.d. according to $p(x)$ in total variation distance. We assume that the alphabet $\mathscr{X}$ is binary in our polar code construction below.

Encoding

Input: uniformly distributed message $M^{1:|\mathscr{H}_X|}$ ($|\mathscr{H}_X|$ bits)

Output: codeword $X^{1:N}$

for $i = 1 : N$, set U_i as follows.

1. For $i \in \mathscr{H}_X$, the value of U_i is given by setting

$$U^{\mathscr{H}_\mathscr{X}} = M^{1:|\mathscr{H}_X|}.$$

2. For $i \in \mathscr{L}_X$, we set U_i using the **argmax rule**

$$U_i = \text{argmax}_{x\in\{0,1\}} P_{U_i|U^{1:i-1}}(x|U^{1:i-1}).$$

3. For $i \in R$, we set U_i by randomized rounding with the

conditional distribution, $P_{U_i|U^{1:i-1}}(x|U^{1:i-1})$.

end

4. $X^{1:N} = U^{1:N} G_N$ becomes the codeword.

The decoding algorithm is as follows.

Decoding

Input: codeword $X^{1:N}$

Output: message estimate $M^{1:|\mathscr{H}_X|}$

1. We reconstruct $U^{1:N}$ by applying G_N to $X^{1:N}$.

2. Therefore $M^{1:|\mathscr{H}_X|} = U^{\mathscr{H}_{\mathscr{X}}}$.

Let Q be the measure on $X^{1:N}$ induced by the polar shaping code. Note that P is the measure on $X^{1:N}$ induced when $X^{1:N}$ is i.i.d. distributed according to $p(x)$. From the results in [24], [15], [11], it is obvious that $||P_{X^{1:N}} - Q_{X^{1:N}}|| = O(2^{-N^\beta})$.

Note that expected symbol frequency is as follows:

$$\mathbb{E}[f_j(X^{1:N})] = \frac{1}{N} \sum_{i=1}^{N} P(X_i = j).$$

We refer to the distribution given by expected symbol frequency function as symbol occurrence probability distribution for that block code. Let us call it $q_N(x)$. By using the fact that total variation distance $||P_{X^{1:N}} - Q_{X^{1:N}}|| = O(2^{-N^\beta})$, it can be easily shown that $q_N(x)$ approaches $p(x)$ as follows:

$$\begin{aligned} q_N(x) &= \frac{1}{N} \sum_{i=1}^{N} P(X_i = x) \\ &\leq \frac{1}{N} \sum_{i=1}^{N} (p(x) + ||P_{X^{1:N}} - Q_{X^{1:N}}||) \\ &= p(x) + O(2^{-N^\beta}). \end{aligned}$$

Similarly,

$$\begin{aligned} q_N(x) &= \frac{1}{N}\sum_{i=1}^{N} P(X_i = x) \\ &\geq \frac{1}{N}\sum_{i=1}^{N} (p(x) - ||P_{X^{1:N}} - Q_{X^{1:N}}||) \\ &= p(x) - O(2^{-N^{\beta}}). \end{aligned}$$

Hence $q_N(x)$ approaches $p(x)$ as N grows. Note that the polar shaping code is invertible, in contrast to polar source codes that are not strictly invertible [3]. As fraction of high-entropy bit-channels, where we provide message bits, approaches $H(X)$, we say that sequence of polar codes achieve any rate $R < H(X)$ with symbol occurrence distribution $p(x)$. The extension to the non-binary case can be done using ideas from [47].

3.3.2 Application to costly channel

Note the cost of a codeword $x^{1:N} \in \mathscr{X}^N$ per information bit

$$\bar{C}_N(x^{1:N}) = \frac{1}{R}\sum_{j\in\mathscr{X}} f_j(x^{1:N})C(j).$$

Therefore, the average cost per information bit of the shaping code will be as follows:

$$\begin{aligned} \mathbb{E}[\bar{C}_N(X^{1:N})] &= \frac{1}{R}\sum_{j\in\mathscr{X}} \mathbb{E}[f_j(X^{1:N})]C(j) \\ &= \frac{1}{R}\sum_{x\in\mathscr{X}} C(x)q_N(x). \end{aligned}$$

Note that average cost per information bit, which we refer to as **total cost**, $\mathbb{E}[\bar{C}_N(X^{1:N})]$, for the sequence of polar codes, approaches the optimal value [29, Theorem 3], by choosing R close to $H(X)$, by choosing $p(x)$ as the symbol occurrence distribution characterized in

[29, Theorem 3], which depends on the cost function. Hence, we say that the sequence of polar codes achieve optimal total cost.

3.4 Shaping for DMCs

In this section, we consider shaping codes for DMCs characterized by transition probabilities $p(y|x)$.

A $(2^{NR},N)$ code for a DMC consists of:

- message set: $\{1,2,\ldots,2^{NR}\}$,
- source of common randomness Z_N known to encoder and decoder independent of message,
- an encoder $X^{1:N} : \{1,2,\ldots,2^{NR}\} \times Z_N \to \mathscr{X}^N$ and
- a decoder at receiver $h : \mathscr{Y}^N \times Z_N \to \{1,2,\ldots,2^{NR}\}$.

R is the rate of the message. Let M be chosen uniformly from the set $\{1,2,\ldots,2^{NR}\}$. Let $Y^{1:N}$ be the received sequence. Note that we also employ common randomness in the definition of the code. Now we upper bound the rate that can be achieved on DMC with certain symbol occurrence probability distribution in following subsection.

3.4.1 Upper bound on rate under a constraint on symbol occurrence distribution

The upper bound we provide in this sub-section also applies to the case when $\mathscr{X}$ is non-binary alphabet. We refer to $q_N(x)$ and $\mathbb{E}[\bar{C}_N(X^{1:N})]$ for symbol occurrence distribution and total cost as defined in Section-3.3.

Definition: We say that R rate is achieved with symbol occurrence probability $p(x)$ iff there exists a sequence of $(2^{NR},N)$ codes for discrete memoryless channels such that $P_e^N = \mathbb{P}(h(Y^{1:N},Z_N) \neq M)$ vanishes and $q_N(x)$ approaches $p(x)$ as N goes to ∞.

Lemma 6. *If rate R is achieved with symbol occurrence probability $p(x)$, then $I(X;Y)$, mutual information evaluated at $p(x)$ for the DMC, will be an upper bound on R.*

Proof: Let us consider a sequence of $(2^{NR},N)$ codes for which $q_N(x)$ approaches $p(x)$ and probability of error diminishes. By Fano's inequality, we get

$$H(M|Y^{1:N},Z_N) = N\varepsilon_N,$$

where ε_N vanishes as N grows. Let $\tilde{X}^N$ be the random variable distributed as $q_N(x)$.

$$\begin{aligned}
NR &= H(M) \\
&= H(M) - H(M|Y^{1:N}Z_N) + H(M|Y^N Z_N) \\
&\leq H(M) - H(M|Y^{1:N}Z_N) + N\varepsilon_N \\
&\overset{(a)}{=} H(M|Z_N) - H(M|Y^N Z_N) + N\varepsilon_N \\
&= I(M;Y^{1:N}|Z_N) + N\varepsilon_N \\
&= \sum_{i=1}^{N} I(M;Y_i|Z_N Y^{1:i-1}) + N\varepsilon_N \\
&= \sum_{i=1}^{N} (H(Y_i|Y^{1:i-1}Z_N) - H(Y_i|M,Y^{1:i-1},Z_N)) + N\varepsilon_N \\
&\overset{(b)}{\leq} \sum_{i=1}^{N} (H(Y_i) - H(Y_i|M,Y^{1:i-1},X_i,Z_N)) + N\varepsilon_N \\
&\overset{(c)}{=} \sum_{i=1}^{N} (H(Y_i) - H(Y_i|X_i)) + N\varepsilon_N \\
&= \sum_{i=1}^{N} I(X_i;Y_i) + N\varepsilon_N \\
&\overset{(d)}{\leq} NI(\tilde{X}^N;Y) + N\varepsilon_N
\end{aligned}$$

where $I(\tilde{X}^N;Y)$ mutual information evaluated at distribution $q_N(x)$ for the DMC $p(y|x)$. Identity (a) follows as source of common randomness is independent of the message. Identity (b) follows

as conditioning reduces entropy. Identity (c) follows as Y_i independent of $Y^{1:i-1}, Z_N$ given X_i. Identity (d) follows as $q_N(x) = \frac{1}{N}\sum_{i=1}^{N} P(X_i = j)$ and mutual information is concave in input distribution for fixed $p(y|x)$.

As N approaches infinity, $I(\tilde{X}^N, Y)$ approaches $I(X;Y)$ since $q_N(x)$ approaches $p(x)$ and mutual information is continuous function with input distribution for fixed $p(y|x)$. Hence $R \leq I(X,Y)$. □

We use this result in the following subsection to define an optimization problem for the costly noisy DMC that provides a lower bound on the optimal total cost.

3.4.2 Lower bound on optimal total cost for costly noisy channel

Definition: We say that rate R is achieved with total cost $\tilde{C}$ iff there exists a sequence of $(2^{NR}, N)$ codes for discrete memoryless channels such that $P_e^N = \mathbb{P}(h(Y^{1:N}, Z_N) \neq M)$ vanishes and $\mathbb{E}[\bar{C}_N(X^{1:N})]$ approaches $\tilde{C}$ as N goes to ∞.

Definition: Optimal total cost C_{opt} is defined as follows:

$$C_{opt} = inf_{(R,\tilde{C})}\tilde{C},$$

where infimum is taken over $(R,\tilde{C})$ pairs such that R is achieved with total cost $\tilde{C}$.

Lemma 7. *The optimal total cost $C_{opt} = inf_{(R,p(x))}\tilde{C}$, where $\tilde{C} = \frac{1}{R}\sum_{x\in\mathscr{X}} C(x)p(x)$ and infimum is taken over $(R, p(x))$ pairs such that R is achieved with symbol occurrence distribution $p(x)$.*

Proof: We first provide the proof when $\mathscr{X}$ is binary alphabet. Without loss of generality assume that $C(0) \neq C(1)$ otherwise total cost is always equal to $\frac{C(0)}{R}$. Notice that there will be one to one correspondence between $q_N(x)$ and $\mathbb{E}[\bar{C}_N(X^{1:N})]$, which are affinely related for a given rate R. Hence $q_N(x)$ converges if and only if $\mathbb{E}[\bar{C}_N(X^{1:N})]$ converges as N goes to ∞. The limits are also affinely related by the same function as both the sequences are. Hence, R is achieved with total cost $\tilde{C}$ iff R is achieved with symbol occurrence distribution $p(x)$, where $\tilde{C} = \frac{1}{R}\sum_{x\in\mathscr{X}} C(x)p(x)$.

Therefore, $C_{opt} = inf_{(R,p(x))}\tilde{C}$, where $\tilde{C} = \frac{1}{R}\sum_{x\in\mathscr{X}} C(x)p(x)$ and infimum is taken over $(R, p(x))$ pairs such that R is achieved with symbol occurrence distribution $p(x)$.

When $\mathscr{X}$ is non-binary alphabet, total cost at rate can be same for two different distributions. So, this argument does not apply to non-binary case. We need to use the fact that every bounded sequence has convergent sub-sequence to prove that if rate R is achieved with total cost $\tilde{C}$ then there exist $p(x)$ such that rate R is achieved with symbol occurrence distribution $p(x)$ where $\tilde{C} = \frac{1}{R}\sum_{x\in\mathscr{X}} C(x)p(x)$.

If rate R is achieved with $\tilde{C}$ then there exists a sequence of $(2^{NR}, N)$ codes such that $P_e^N = \mathbb{P}(h(Y^{1:N}, Z_N) \neq M)$ vanishes and $\mathbb{E}[\bar{C}_N]$ approaches $\tilde{C}$ as N goes to ∞. The symbol occurrence distribution $q_N(x)$ may not converge. But there exists a sub-sequence of the sequence $q_N(x)$ that converges, as $q_N(x)$ is a bounded sequence. Let us index such a sub-sequence with k where block length corresponding to the kth element in the sub-sequence is N_k. Let $q_{N_k}(x)$ converges to the distribution $p(x)$. Clearly for sequence of codes $(2^{N_k R}, N_k)$, $\mathbb{E}[\bar{C}_{N_k}]$ approaches $\tilde{C}$ as k goes to ∞. As $\mathbb{E}[\bar{C}_{N_k}] = \frac{1}{R}\sum_{x\in\mathscr{X}} C(x)q_{N_k}(x)$, we will have $\tilde{C} = \frac{1}{R}\sum_{x\in\mathscr{X}} C(x)p(x)$. Therefore we have sequence of codes for which probability of error diminishes and symbol occurrence probability distribution converges to $p(x)$ such that $\tilde{C} = \frac{1}{R}\sum_{x\in\mathscr{X}} C(x)p(x)$, which means that if rate R is achieved with total cost $\tilde{C}$ then there exists a distribution $p(x)$ such that R rate is achieved with symbol occurrence distribution $p(x)$ where $\tilde{C} = \frac{1}{R}\sum_{x\in\mathscr{X}} C(x)p(x)$. On the other hand, if rate R is achieved with symbol occurrence probability distribution $p(x)$, then obviously rate R is achieved with total cost $\tilde{C} = \frac{1}{R}\sum_{x\in\mathscr{X}} C(x)p(x)$.

Therefore, $C_{opt} = inf_{(R,p(x))}\tilde{C}$, where $\tilde{C} = \frac{1}{R}\sum_{x\in\mathscr{X}} C(x)p(x)$ and infimum is taken over $(R, p(x))$ pairs such that R is achieved with symbol occurrence distribution $p(x)$. This concludes the proof of the lemma. □

As stated in the previous subsection, if rate R is achieved with symbol occurrence distribution $p(x)$, then $R < I(X;Y)$. Note that the solution for the following optimization

problem is lower bound to C_{opt}.

$$\begin{aligned} &\text{Minimize}_{(R,p(x))} \frac{1}{R} \sum_{x\in\mathscr{X}} C(x)p(x), \\ &\text{subject to } R \leq I(X;Y). \end{aligned} \tag{3.2}$$

In the next subsection, we show that polar coding technique designed for asymmetric channels can be used to achieve any rate $R < I(X;Y)$ with symbol occurrence probability $p(x)$. Therefore, the sequence of polar codes, which are designed with minimizers of the optimization problem achieve the lower bound provided by the solution of the optimization problem. This means that the solution of the optimization problem characterizes the optimal total cost of costly noisy DMCs.

We now compute the optimal total cost for a costly M-ary erasure channel. This channel has alphabet size $|\mathscr{X}| = M$, and each symbol is erased with certain erasure probability.

Theorem 3. *The optimal symbol occurrence input distribution of the shaping code that achieves optimal total cost for an M-ary erasure costly channel with erasure probability ρ is given by $p^*(x) = 2^{-\mu C(x)}$ such that $\sum_{x\in\mathscr{X}} 2^{-\mu C(x)} = 1$. We assume that the cost function $C(x)$ is non-trivial for each $x \in \mathscr{X}$. The optimal total cost is given by $C_{opt} = \frac{\sum_{x\in\mathscr{X}} p^*(x)C(x)}{(1-\rho)\sum_{x\in\mathscr{X}} p^*(x)\log_2(1/p^*(x))}$.*

Proof: Mutual information $I(X;Y)$ evaluated at the input distribution $p(x)$ for the erasure channel is given by $(1-\rho)\sum_{x\in\mathscr{X}} p(x)\log_2(1/p(x))$. By substituting the mutual information in (3.2), the optimization problem for the costly erasure channel takes the form:

$$\begin{aligned} &\text{Minimize}_{(R,p(x))} \frac{1}{R} \sum_{x\in\mathscr{X}} C(x)p(x), \\ &\text{subject to } R \leq (1-\rho) \sum_{x\in\mathscr{X}} p(x)\log_2(1/p(x)). \end{aligned}$$

For fixed rate R, finding out the symbol occurrence probability for minimum total cost will be a convex optimization problem. Using Lagrange duality, we get the optimal symbol occurrence

distribution at a fixed rate R as $\tilde{p}(x) = \frac{2^{-\mu C(x)}}{N_\mu}$, where μ is a positive constant such that

$$R = (1-\rho) \sum_{x \in \mathscr{X}} \tilde{p}(x) \log_2(1/\tilde{p}(x))$$

and $N_\mu = \sum_{x \in \mathscr{X}} 2^{-\mu C(x)}$ is normalization factor. We now minimize the function below, for minimizing total cost:

$$G(\mu) = \frac{\sum_{x \in \mathscr{X}} C(x) 2^{-\mu C(x)}}{(1-\rho)(\sum_{x \in \mathscr{X}} \mu C(x) 2^{-\mu C(x)} + N_\mu \log_2 N_\mu)}$$

when $\mu > 0$. Notice that this function is the same as function T defined in the proof of [29, Theorem 3] except for a factor $1-\rho$. The derivative $G'(\mu)$ will have negative of the sign of $\log_2 N_\mu$ as shown in the proof of [29, Theorem 3]. As we assume that $C(x) > 0$ for each $x \in \mathscr{X}$ and μ increases from 0 to ∞, N_μ is decreasing from $|\mathscr{X}|$ to 0. So G will be initially decreasing as μ increases until N_μ becomes 1 and then will be increasing. So the minimum value of G occurs when $N_\mu = 1$. Hence the symbol occurrence distribution that achieves minimum total cost is $p^*(x) = 2^{-\mu C(x)}$ for each $x \in \mathscr{X}$ where $\sum_{x \in \mathscr{X}} 2^{-\mu C(x)} = 1$. Therefore the optimal total cost $C_{opt} = \frac{\sum_{x \in \mathscr{X}} p^*(x) C(x)}{(1-\rho) \sum_{x \in \mathscr{X}} p^*(x) \log_2(1/p^*(x))}$. □

Now we provide the shaping polar code to achieve any rate $R < I(X;Y)$ with symbol occurrence probability distribution $p(x)$ over DMCs.

3.4.3 Polar shaping codes for DMCs

We assume alphabet $\mathscr{X}$ is binary in the proposed polar code construction. The polar code that we provide here transforms uniformly distributed message $M^{1:|I|}$ ($|I|$ bits) into code-word $X^{1:N}$, whose distribution is close to the distribution induced when $X^{1:N}$ is i.i.d. according to $p(x)$ in total variation distance. The code construction we propose here is actually derived from the capacity-achieving polar codes for asymmetric channels [24] by Honda and Yamamoto. We use common randomness in the code construction to get the desired shaping property. Now we

provide the encoding algorithm.

Encoding

Input: randomly chosen message $M^{1:|I|}$

Output: codeword $X^{1:N}$

for $i = 1 : N$, encode U_i as follows.

1. For $i \in I$, the value of U_i is given by setting $U^I = M^{1:|I|}$.

2. For $i \in F$, we set U_i as uniform independent

   random variable through common randomness.

3. For $i \in \mathscr{L}_X$, we encode U_i using the **argmax rule**

$$U_i = \text{argmax}_{x\in\{0,1\}} P_{U_i|U^{1:i-1}}(x|U^{1:i-1}).$$

4. For $i \in S$, we set U_i with conditional distribution

   $P_{U_i|U^{1:i-1}}(x|U^{1:i-1})$ using common randomness.

end

Transmit $X^{1:N} = U^{1:N}G_N$.

The decoding algorithm is as follows.

Decoding

Input: received vector $Y^{1:N}$

Output: message estimate $\hat{M}^{1:|I|}$

for $i = 1 : N$

1. If $i \in F$, we reconstruct $\hat{U}_i$ using common randomness,

   which is uniform independent random variable.

2. If $i \in \mathscr{L}_X \cup I$, set

$$\hat{U}_i = \text{argmax}_{x\in\{0,1\}} P_{U_i|U^{1:i-1},Y^{1:N}}(x|\hat{U}^{1:i-1}, Y^{1:N}).$$

3. If $i \in S$, we reconstruct $\hat{U}_i$ using common randomness

with conditional distribution $P_{U_i|U^{1:i-1}}(x|\hat{U}^{1:i-1})$.

end

Decode $\hat{M} = \hat{U}^I$.

Let Q be the measure on $X^{1:N}$ induced by the polar shaping code. Note that P is the measure on $X^{1:N}$ induced when it is i.i.d. distributed according to $p(x)$. From the results in [24], [15], [11], it is obvious that $||P_{X^{1:N}} - Q_{X^{1:N}}|| = O(2^{-N^{\beta'}})$ for $\beta' < \beta < 0.5$. As mentioned in Section 3.3, $q_N(x)$ approaches $p(x)$ as N grows large. The probability of decoding error is $O(2^{-N^{\beta'}})$ [24].

As fraction of information bit-channels, where we provide message bits, approaches $I(X;Y)$, the sequence of polar codes achieve rate $R < I(X;Y)$ with symbol occurrence distribution $p(x)$. Common randomness we employed in the code construction is crucial to get the desired distribution on the symbols of the codeword. If R and $p(x)$ in the polar code design are minimizers of the optimization problem proposed in the previous subsection, then sequence of polar codes clearly achieve the optimal total cost.

As proposed in [24], [37], if instead of using common randomness, we randomly generate frozen bits for bit-channels in F and use boolean functions for encoding bit-channels in S, which are shared between encoder and decoder, we will not be able to guarantee the desired shaping distribution. The ensemble average symbol occurrence distribution will have the desired shaping distribution, but we cannot guarantee the existence of a code in the random ensemble with the desired shaping distribution. Nevertheless, we now prove that the random construction generates codes whose total costs approach the optimal total cost with diminishing probability of error if we design the polar code with minimizers of the optimization problem. Code constructions

avoiding common randomness are advantageous as the practical implementation of common randomness uses pseudo-random generators which often have many limitations. They can suffer from shorter than the expected period for weak seed states.

Let $p^*(x)$ be optimal symbol occurrence distribution and R^* be the optimal rate for costly noisy DMC. Clearly, R^* is the mutual information evaluated at $p^*(x)$ for the DMC.

Now we design sequence of polar codes with optimal symbol occurrence distribution $p^*(x)$ and optimal rate R^* by random code construction method [24], as mentioned above avoiding common randomness. Let W_N denotes the random vector of frozen bits and boolean functions for bit-channels in F and R respectively. Note that the rate sequence $R_N = \frac{|I|}{N}$ approaches R^*. Clearly, the total cost for a given code will become as follows:

$$\mathbb{E}[C_N(X^{1:N})|W_N] = \sum_{x^{1:N}\in\mathscr{X}^{1:N}} 2^{-|I|}\Pi_{i\in F}\mathbb{1}\{f(i)=u_i\}$$
$$\Pi_{i\in\mathscr{L}_X}\delta_i(u_i|u^{1:i-1})\Pi_{i\in S}\mathbb{1}\{\lambda_i(u^{1:i-1})=u_i\}C_N(x^{1:N}),$$

where $x^{1:N} = u^{1:N}G_N$, $\mathbb{1}\{.\}$ is indicator function, $\delta_i(u|u^{1:i-1})$ denotes the conditional distribution induced by argmax rule for bit-channels in $\mathscr{L}_X$ as defined in [37], $f(.)$ is frozen bit function that is randomly chosen for bit-channels in F as defined in [37] and $\lambda_i(.)$ denotes the boolean functions to encode bit-channels in R as defined in [37].

Applying expectation on both sides and by the independence of frozen bits and boolean functions [24], [37], the ensemble average total cost becomes as follows:

$$\mathbb{E}_{W_N}[\mathbb{E}[C_N(X^{1:N})|W_N]]$$
$$= \sum_{x^{1:N}\in\mathscr{X}^{1:N}} 2^{-|\mathscr{H}_X|}\Pi_{i\in\mathscr{L}_X}\delta_i(u_i|u^{1:i-1})\Pi_{i\in S}P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1})C_N(x^{1:N})$$
$$= \sum_{x^{1:N}\in\mathscr{X}^{1:N}} Q(x^{1:N})C_N(x^{1:N}).$$

This implies that $\mathbb{E}_{W_N}[\mathbb{E}[C_N(X^{1:N})|W_N]] = \frac{N}{|I|}\sum_{x\in\mathscr{X}} C(x)q_N(x)$.

Therefore, as $q_N(x)$ approaches $p^*(x)$ and $\frac{|I|}{N}$ approaches R^*, ensemble average total cost $\mathbb{E}_{W_N}[\mathbb{E}[C_N(X^{1:N})|W_N]]$ approaches $\frac{1}{R^*}\sum_{x\in\mathscr{X}} C(x)p^*(x)$ which is optimal total cost C_{opt}. On the other hand, the ensemble average probability of error $\mathbb{E}_{W_N}[P_e(W_N)] = O(2^{-N^{\beta'}})$ [24] where $\beta' < \beta < 0.5$ and $P_e(W_N)$ is the probability of error of the given code. A good shaping code has total cost close to the optimal value and negligible probability of error. So we should show there

exists a sequence of codes whose total cost approaches optimal total cost and probability of error diminishes. We show the existence of such codes with high probability. We precisely state our result in Theorem 4 followed by a rigorous proof. For the sake of brevity, we denote, the total cost for a given code, $\mathbb{E}[C_N(X^{1:N})|W_N]$, by T_N.

Theorem 4. *In the above random code construction,* $\mathbb{P}\big(P_e(W_N) < N2^{-N^{\beta'}}, \tilde{b}_N\mathbb{E}_{W_N}[T_N] \leq T_N \leq \tilde{a}_N\mathbb{E}_{W_N}[T_N]\big)$ *approaches* 1 *for some* $\tilde{a}_N > 1$ *and* $\tilde{b}_N < 1$ *that converge to 1. This means that, with high probability, there exist codes in the random ensemble with total cost approaching the optimal total cost and diminishing probability of error.*

Proof: We first prove that $\mathbb{P}(b_N\mathbb{E}_{W_N}[T_N] \leq T_N \leq a_N\mathbb{E}_{W_N}[T_N])$ goes to 1 for any b_N that converges to $b < 1$ from below and a_N that converges to $a > 1$ from above. This is equivalent to proving $\mathbb{P}(T_N > a_N\mathbb{E}_{W_N}[T_N])$ converges to zero and $\mathbb{P}(T_N < b_N\mathbb{E}_{W_N}[T_N])$ converges to zero. For the sake of brevity, we denote $\mathbb{E}_{W_N}[T_N]$ as $\tilde{\mathbb{E}}[T_N]$ in the proof.

Now we prove $\mathbb{P}(T_N > a_N\tilde{\mathbb{E}}[T_N])$ converges to zero by contradiction. So we assume there is subsequence indexed by r, $\mathbb{P}(T_{N_r} > a_{N_r}\tilde{\mathbb{E}}[T_{N_r}])$, whose liminf is non-zero. Let us denote the sequence $\mathbb{P}(T_{N_r} > a_{N_r}\tilde{\mathbb{E}}[T_{N_r}])$ as p_{N_r}. Note that limsup of sequence p_{N_r} is less than 1, as p_{N_r} is upper-bounded by $\frac{1}{a_{N_r}}$, by the Markov inequality. As $a_{N_r} > 1$, we will be able to choose $0 \leq l_{N_r} < 1$ such that $\tilde{\mathbb{E}}[T_{N_r}] = p_{N_r}a_{N_r}\tilde{\mathbb{E}}[T_{N_r}] + (1-p_{N_r})l_{N_r}\tilde{\mathbb{E}}[T_{N_r}]$.

Note that $l_{N_r} = \frac{1-p_{N_r}a_{N_r}}{1-p_{N_r}}$. Therefore,

$$\begin{aligned}
\limsup_{r\to\infty} l_{N_r} &= \limsup_{r\to\infty} \frac{1-p_{N_r}a_{N_r}}{1-p_{N_r}} \\
&\overset{(a)}{=} \limsup_{r\to\infty} \frac{1-p_{N_r}(\lim_{r\to\infty} a_{N_r})}{1-p_{N_r}} \\
&\overset{(b)}{=} \limsup_{r\to\infty} \frac{1-p_{N_r}a}{1-p_{N_r}} \\
&\overset{(c)}{=} \frac{1-(\liminf_{r\to\infty} p_{N_r})a}{1-\liminf_{r\to\infty} p_{N_r}} \\
&\overset{(d)}{\leq} 1
\end{aligned}$$

Identity (a) is true as limit of a_{N_r} exist. Idenity (b) follows as limit of a_{N_r} is a. Identity (c) is true as the function $\frac{1-ax}{1-x}$ decreases as x increases when $a > 1$. Idenity (d) follows from the fact that $\frac{1-ax}{1-x} < 1$ for $a > 1$ and $0 < x < 1$.

Hence limsup of l_{N_r} less than 1. Set $l'_{N_r} = \frac{1+l_{N_r}}{2}$. Therefore, $\tilde{\mathbb{E}}[T_{N_r}] < p_{N_r} a_{N_r} \tilde{\mathbb{E}}[T_{N_r}] + (1 - p_{N_r}) l'_{N_r} \tilde{\mathbb{E}}[T_{N_r}]$ and limsup of l'_{N_r} is less than 1. Note that

$$\tilde{\mathbb{E}}[T_{N_r}] = p_{N_r} a_{N_r} \tilde{\mathbb{E}}[T_{N_r}] + (1 - p_{N_r} - q_{N_r}) l'_{N_r} \tilde{\mathbb{E}}[T_{N_r}] \tag{3.3}$$

where $q_{N_r} = (1 - p_{N_r}) \frac{1-l_{N_r}}{1+l_{N_r}}$, and liminf of q_{N_r} does not vanish as the limsups of l_{N_r} and p_{N_r} are strictly less than 1.

Note also that

$$\tilde{\mathbb{E}}[T_{N_r}] \geq p_{N_r} a_{N_r} \tilde{\mathbb{E}}[T_{N_r}] + \mathbb{P}(l'_{N_r} \tilde{\mathbb{E}}[T_{N_r}] \leq T_{N_r} \leq a_{N_r} \tilde{\mathbb{E}}[T_{N_r}]) l'_{N_r} \tilde{\mathbb{E}}[T_{N_r}].$$

By plugging in for $\tilde{\mathbb{E}}[T_{N_r}]$ using equation (3.3), we get

$$\mathbb{P}(l'_{N_r} \tilde{\mathbb{E}}[T_{N_r}] \leq T_{N_r} \leq a_{N_r} \tilde{\mathbb{E}}[T_{N_r}]) \leq (1 - p_{N_r} - q_{N_r}).$$

This yields

$$\begin{aligned} \mathbb{P}(T_{N_r} < l'_{N_r} \tilde{\mathbb{E}}[T_{N_r}]) &= 1 - \mathbb{P}(l'_{N_r} \tilde{\mathbb{E}}[T_{N_r}] \leq T_{N_r} \leq a_{N_r} \tilde{\mathbb{E}}[T_{N_r}]) - \mathbb{P}(T_{N_r} > a_{N_r} \tilde{\mathbb{E}}[T_{N_r}]) \\ &\geq 1 - (1 - p_{N_r} - q_{N_r}) - p_{N_r} = q_{N_r}. \end{aligned}$$

Hence q_{N_r} is a lower-bound to $\mathbb{P}(T_{N_r} < l'_{N_r} \tilde{\mathbb{E}}[T_{N_r}])$. Hence $\mathbb{P}(T_{N_r} < l'_{N_r} \tilde{\mathbb{E}}[T_{N_r}])$ does not converge to zero. As $\mathbb{P}(P_e < N_r 2^{-N_r^{\beta'}})$ converges to 1, it follows that the sequence $\mathbb{P}(P_e < N_r 2^{-N_r^{\beta'}}, T_{N_r} < l'_{N_r} \tilde{\mathbb{E}}[T_{N_r}])$ does not converge to zero. As limsup of l'_{N_r} is less than 1, we can get a sequence of codes whose total costs converge to less than optimal total cost C_{opt} with diminishing probability of error. This is a contradiction. Hence $\mathbb{P}(T_N > a_N \tilde{\mathbb{E}}[T_N])$ converges to zero.

Now we prove that $\mathbb{P}(T_N < b_N\tilde{\mathbb{E}}[T_N])$ converges to zero. We again prove this by contradiction. So assume $\mathbb{P}(T_N < b_N\tilde{\mathbb{E}}[T_N])$ does not converge to zero. As $\mathbb{P}(P_e < N2^{-N^{\beta'}})$ converges to 1, the sequence $\mathbb{P}(P_e < N2^{-N^{\beta'}}, T_N < b_N\tilde{\mathbb{E}}[T_N])$ does not converge to zero. This is a contradiction as we can get a sequence of codes whose total costs converge to less than optimal total cost C_{opt} with diminishing probability of error. Hence $\mathbb{P}(T_N < b_N\tilde{\mathbb{E}}[T_N])$ converges to zero.

We conclude $\mathbb{P}(b_N\tilde{\mathbb{E}}[T_N] \leq T_N \leq a_N\tilde{\mathbb{E}}[T_N])$ goes to 1. As $\mathbb{P}(P_e < N2^{-N^{\beta'}})$ converges to 1, we will have $\mathbb{P}(P_e < N2^{-N^{\beta'}}, b_N\tilde{\mathbb{E}}[T_N] \leq T_N \leq a_N\tilde{\mathbb{E}}[T_N])$ goes to 1.

Let $\hat{a}_m > 1$ be a sequence indexed by m that converges to 1 from above. Let $\hat{b}_m < 1$ be a sequence indexed by m that converges to 1 from below. For each m, let us define a sequence $\hat{a}_{mN}$ that converges to $\hat{a}_m$ from above and also define another sequence $\hat{b}_{mN}$ that converges to $\hat{b}_m$ from below. So for each m, we have $\mathbb{P}(P_e < N2^{-N^{\beta'}}, \hat{b}_{mN}\tilde{\mathbb{E}}[T_N] \leq T_N \leq \hat{a}_{mN}\tilde{\mathbb{E}}[T_N])$ goes to 1. Consider a sequence $0 < \varepsilon_m < 1$ converging to 0 as $m \to \infty$. Notice that we can find a sequence N_m indexed by m such that $|\hat{a}_{mN_m} - \hat{a}_m| < \varepsilon_m$, $|\hat{b}_{mN_m} - \hat{b}_m| < \varepsilon_m$, and $|\mathbb{P}(P_e < N_m 2^{-N_m{}^{\beta'}}, \hat{b}_{mN_m}\tilde{\mathbb{E}}[T_{N_m}] \leq T_{N_m} \leq \hat{a}_{mN_m}\tilde{\mathbb{E}}[T_{N_m}]) - 1| < \varepsilon_m$.

This implies that $\mathbb{P}(P_e < N_m 2^{-N_m{}^{\beta'}}, \hat{b}_{mN_m}\tilde{\mathbb{E}}[T_{N_m}] \leq T_{N_m} \leq \hat{a}_{mN_m}\tilde{\mathbb{E}}[T_{N_m}])$ goes to 1, $\hat{b}_{mN_m} < 1$ and $\hat{a}_{mN_m} > 1$ converge to 1 as $m \to \infty$. Note that this essentially completes the proof of the theorem. □

The extension to non-binary case can be done using ideas in [47].

3.5 Conclusion

We presented a polar shaping code. For a costly channel, we have shown that total cost of the proposed polar shaping code approaches optimal total cost as block length grows. We looked at costly noisy discrete memoryless channels. We first gave an upper bound on the rate that can be achieved with certain symbol occurrence probability distribution over a discrete memoryless channel. We formulated an optimization problem whose solution gives optimal total cost for the costly noisy discrete memoryless channel. We showed that polar codes for asymmetric channels

by Honda and Yamamoto with the aid of common randomness can be used to get the desired shaping distribution on symbols of the codeword. To achieve the optimal total cost, we show that we can also use random construction method by randomly choosing frozen bits and randomly choosing boolean functions for not completely polarized channels [37], [24] avoiding common randomness.

Acknowledgement

This chapter is in part a reprint of the material in the paper: Karthik Nagarjuna Tunuguntla, Paul H. Siegel, "Polar shaping codes for costly noisy and noiseless channels," *2021 International Symposium on Information Theory (ISIT)*, pp. 2560-2565, Melbourne, Australia, June 2021. Dissertation author is the primary contributor of the paper.

Chapter 4

Slepian-Wolf Polar Coding with Unknown Correlation

4.1 Introduction

4.1.1 Background

Arikan [1] constructed capacity-achieving codes for binary-input symmetric channels. Korada and Urbanke [27] constructed a Slepian-Wolf polar coding scheme for two correlated sources under some assumptions. Arikan [3] proposed a polar coding method for an arbitrary discrete memoryless source with correlated side-information available at the receiver. Based on that, he also derived a Slepian-Wolf polar coding strategy for any two binary correlated random variables using a successive cancellation style decoding of the source sequences. Arikan [2] proposed a monotone chain rule to achieve all the rates of the Slepian-Wolf region without the use of a time-sharing policy.

A capacity-achieving coding scheme based on source and channel polarization for binary-input asymmetric channels was proposed by Honda and Yamamoto [24]. Hassani and Urbanke [22], [23] presented universal coding schemes to achieve rates close to the compound capacity of binary-input symmetric discrete-memoryless channels (DMCs) that are based on polar codes. The authors [37] proposed a universal polar coding scheme for the asymmetric setting that eliminates the need to store high-complexity boolean functions. The scheme uses the

elements of coding strategies in [22], [24]. Wang and Kim [55] discussed the linear code duality between channel coding and source coding when the correlated side information is available at the receiver. In this paper, we consider the variant of the Slepian-Wolf coding problem which involves a binary memoryless source and correlated side information available at the receiver, as usual, but where the conditional distribution of the side information given the source is unknown to the encoder.

4.1.2 Problem definition

Let $\mathscr{X}$ be the binary alphabet and $\mathscr{Y}$ be some arbitrary finite alphabet. A binary discrete memoryless source X_i is distributed as $P_X(x)$ with side information Y_i available at the receiver. The $(X_i, Y_i)_{i=1}^{\infty}$ sequence is an iid (identical and independently distributed) random process whose joint distribution is $P_X(x)p(y|x)$. The conditional distribution $p(y|x)$ is unknown to the encoder, but available to the decoder only. We also assume that $p(y|x)$ is known to come from a class $\mathscr{C}$ of conditional distributions of a random variable over the alphabet $\mathscr{Y}$ given a correlated random variable over $\mathscr{X}$. The class $\mathscr{C}$ is available to the encoder.

A $(2^k, N)$ code for the defined problem consists of

- an encoder $g : \mathscr{X}^N \to \{1 : 2^k\}$, and
- a decoder $h : \{1 : 2^k\} \times \mathscr{Y}^N \to \mathscr{X}^N$

where N is the block length and $\frac{k}{N}$ is called the rate of the code. Let $P_e^{(N)} = P(X^N \neq h(g(X^N), Y^N))$ be the probability of error. If there is a sequence of $(2^{NR}, N)$ codes for which the corresponding sequence of $P_e^{(N)}$ goes to zero, then the rate R is achieved. Note that classical Slepian-Wolf coding is the case when $p(y|x)$ is known to both the encoder and decoder. In that case, we know that the rate R is achieved if and only if $R > H(X|Y)$. Therefore the achievable rates of the proposed problem should be greater than $\max_{p(y|x)\in\mathscr{C}} H(X|Y)$ where (X, Y) is distributed as $P_X(x)p(y|x)$.

4.1.3 Contribution

We derive two coding strategies for the proposed setting based on the universal polar coding schemes for a compound channel [23], [22], [37]. This will establish the duality between the coding strategies for these source and channel coding settings. Our first method can achieve all rates greater than $\max_{p(y|x)\in\mathscr{C}} H(X|Y)$ for a uniformly distributed source when the class $\mathscr{C}$ contains only conditional distributions with properties of a symmetric channel. The second method can achieve all rates greater than $\max_{p(y|x)\in\mathscr{C}} H(X|Y)$ when $\mathscr{C}$ is a finite set for any arbitrary source.

In Section 4.2, we introduce some definitions and notations which will be used throughout the paper. In Section 4.3, we describe the Slepian-Wolf polar coding with correlated side information available at the receiver. In Section 4.4, we describe our first method that uses the idea of the staircase scheme for a uniform source. We also provide its applicability to a non-uniformly distributed source. In Section 4.5, we explain the second method which is based on the technique of universalization using bit-channel combining.

4.2 Preliminaries

We express any set of random variables $X_i, X_{i+1}, \ldots, X_j$ $(i < j)$ by a row vector $(X_i, X_{i+1}, \ldots, X_j)$ which is denoted by $X^{i:j}$. We denote the set $\{1,2,3,\ldots,N\}$ by $[N]$. We denote the set $\{i, i+1, \cdots, j\}$ by $[i:j]$ $(i < j)$. Let $U^{1:N}$ be a row vector and let $\mathscr{A} \subset [N]$. The row vector consisting of elements in $U^{1:N}$ corresponding to the positions in $\mathscr{A}$ is denoted by $U^{\mathscr{A}}$.

Definition 1. *A binary-input discrete memoryless channel with output alphabet $\mathscr{Y}$ with transition probabilities $p(y|x)$ for each $(x,y) \in \{0,1\} \times \mathscr{Y}$ is said to be symmetric if there exists a permutation $\pi_1 : \mathscr{Y} \to \mathscr{Y}$ such that $\pi_1 = \pi_1^{-1}$ and $p(y|x) = p(\pi_a(y)|x+a)$ for each $(x,a,y) \in \{0,1\}^2 \times \mathscr{Y}$, where $\pi_o : \mathscr{Y} \to \mathscr{Y}$ is the identity permutation.*

We denote the row vector $(\pi_{s_1}(y_1), \pi_{s_2}(y_2), ..., \pi_{s_N}(y_N))$ as $s^{1:N}.y^{1:N}$ for any $y^{1:N} \in \mathscr{Y}^N$ and $s^{1:N} \in \{0,1\}^N$, where $\pi_0 : \mathscr{Y} \to \mathscr{Y}$ is the identity permutation and $\pi_1 : \mathscr{Y} \to \mathscr{Y}$ is the

permutation corresponding to a symmetric channel.

Let G_N be the conventional polar transform [1], represented by a binary matrix of dimension $N \times N$. If $U^{1:N} = X^{1:N} G_N$, then we denote $P(U^{1:N} = u^{1:N})$ by $P_{U^{1:N}}(u^{1:N})$ and similarly we denote $P(U_i = u_i | U^{1:i-1} Y^{1:N} = u^{1:i-1} y^{1:N})$ by $P_{U_i|U^{1:i-1}Y^{1:N}}(u_i | u^{1:i-1} y^{1:N})$. We denote the subvector of $U^{1:N}$ corresponding to the bit-channel set $\mathscr{A} \subset [N]$ as $U^{\mathscr{A}}$.

Let $\mathscr{C} = \{p_1(y|x), p_2(y|x), \ldots, p_s(y|x)\}$, $s \in \mathbb{N}$. Let $(X_i, Y_i)_{i=1}^N$ be iid random tuples distributed according to $P_X(x) p_l(y|x)$, where $l \in [1:s]$ and $N = 2^n$. For the random variable pair (X,Y) distributed as $P_X(x) p_l(y|x)$, the Bhattacharyya parameter is defined as

$$Z(X|Y) = 2 \sum_y P_Y(y) \sqrt{P_{X|Y}(1|y) P_{X|Y}(0|y)}.$$

We define the following bit-channel subsets as follows, where $\beta < 0.5$.

$$\begin{aligned}
\mathscr{H}_X &= \{i \in [N] : Z(U_i | U^{1:(i-1)}) \geq 1 - 2^{-N^\beta}\}. \\
\mathscr{L}_X &= \{i \in [N] : Z(U_i | U^{1:(i-1)}) \leq 2^{-N^\beta}\}. \\
\mathscr{H}_{X|Y_l} &= \{i \in [N] : Z(U_i | U^{1:(i-1)} Y^{1:N}) \geq 1 - 2^{-N^\beta}\}. \\
\mathscr{L}_{X|Y_l} &= \{i \in [N] : Z(U_i | U^{1:(i-1)} Y^{1:N}) \leq 2^{-N^\beta}\}.
\end{aligned}$$

Note that $\mathscr{L}_X \subseteq \mathscr{L}_{X|Y_l}$, for each $l \in [1:s]$. We have the following results from Theorem 1 in [24].

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} |\mathscr{H}_X| &= H(X). \\
\lim_{N\to\infty} \frac{1}{N} |\mathscr{L}_X| &= 1 - H(X). \\
\lim_{N\to\infty} \frac{1}{N} |\mathscr{H}_{(X|Y)_l}| &= H(X|Y). \\
\lim_{N\to\infty} \frac{1}{N} |\mathscr{L}_{(X|Y)_l}| &= 1 - H(X|Y).
\end{aligned}$$

We remove the subscript l for denoting the bit-channel sets $\mathscr{L}_{(X|Y)_l}$ and $\mathscr{H}_{(X|Y)_l}$ when using (X,Y) random variable pair distributed as $P_X(x)p(y|x)$ and denote them as $\mathscr{L}_{X|Y}$ and $\mathscr{H}_{X|Y}$, respectively.

Let the $p(y|x)$s for each $(x,y) \in \mathscr{X} \times \mathscr{Y}$ be the transition probabilities of a symmetric channel. Let $(X_i,Y_i)_{i=1}^N$ be iid random variable pairs distributed according to $P_X(x)p(y|x)$ where $P_X(x)$ is distributed as Bern($\frac{1}{2}$). Let $U^{1:N} = X^{1:N}G_N$. Then the MAP (ML) decision rule for the bit-channel $i \in [N]$ in this setting will be the function $\Phi_i : \{0,1\}^{i-1} \times \mathscr{Y}^N \to \{0,1\}$ defined as follows.

$$\Phi_i(u^{1:i-1}, y^{1:N}) = \mathbb{1}\{P_{U^{1:i-1},Y^{1:N}|U_i}(\hat{u}^{1:i-1}, y^{1:N}|1) \\ \geq P_{U^{1:i-1},Y^{1:N}|U_i}(\hat{u}^{1:i-1}, y^{1:N}|0)\}.$$

Φ_i is precisely the decision rule used in the successive cancellation (SC) decoding for the bit-channel $i \in \mathscr{L}_{X|Y}$ in the polar code construction for symmetric channels. Let us denote the Bhattacharyya parameter corresponding to the bit-channel $i \in [N]$ as Z_i. Therefore $Z_i = Z(U_i|U^{1:i-1}Y^{1:N})$ in this setting.

4.3 Source coding with side-information (Slepian-Wolf polar coding)

We revisit the polar coding scheme proposed by Arikan [3] for the Slepian-Wolf setting that has the binary discrete memoryless source X_i distributed as $P_X(x)$ with correlated side information Y_i available at the receiver, $i \in [N]$. $(X_i,Y_i)_{i=1}^N$ is an iid process whose joint distribution is $P_X(x)p(y|x)$. Here we assume that $p(y|x)$ is known to both the encoder and decoder. The encoding algorithm is presented below.

Encoding

Input: $X^{1:N}$ source sequence.

Output: Compressed bit stream corresponding to the source sequence.

- Compute $U^{1:N} = X^{1:N}G_N$.
- Transmit $U^{\mathscr{L}^c_{X|Y}}$.

The decoding method is as follows.

Decoding

Input: Correlated side information $Y^{1:N}$ and $U^{\mathscr{L}^c_{X|Y}}$.

Output: Source estimate $\hat{X}^{1:N}$.

for $i = 1 : N$

1. If $i \in \mathscr{L}^c_{X|Y}$, set $\hat{U}_i = U_i$.

2. If $i \in \mathscr{L}_{X|Y}$, set

$$\hat{U}_i = \mathbb{1}\{P_{U_i|U^{1:i-1},Y^{1:N}}(1|\hat{U}^{1:i-1}, Y^{1:N}) \geq P_{U_i|U^{1:i-1},Y^{1:N}}(0|\hat{U}^{1:i-1}, Y^{1:N})\}.$$

end

Decode $\hat{X}^{1:N}$ as $\hat{U}^{1:N}G_N$.

Note that the conditional distribution $P_{U_i|U^{1:i-1},Y^{1:N}}(.|.)$ used above in the decoding algorithm is derived under the setting where $X^{1:N} = U^{1:N}G_N$ and $(X_i, Y_i)_{i=1}^N$ is iid distributed as $P_X(x)p(y|x)$. Arikan [3] proved that the probability of error for this scheme is $O(2^{N^{-\beta}})$ where $\beta < 0.5$. In our setup, however, the actual conditional distribution $p(y|x)$ is unknown to the encoder. The encoder only knows that the conditional distribution is selected from the class $\mathscr{C}$. If the encoder transmits $U^{(\cap_{p(y|x)\in\mathscr{C}}\mathscr{L}_{X|Y})^c}$, then the decoder can reliably recover the bits corresponding to bit-channels $(\cap_{p(y|x)\in\mathscr{C}}\mathscr{L}_{X|Y})$ and hence the source. The fraction of bit-channels

$(\cap_{p(y|x)\in\mathscr{C}}\mathscr{L}_{X|Y})^c$ with respect to the block length may not be going to $\max_{p(y|x)\in\mathscr{C}}H(X|Y)$ as block length grows and the fraction may always be larger than $\max_{p(y|x)\in\mathscr{C}}H(X|Y)$ by at least some positive constant. In the following sections, we provide the source coding methods that can guarantee any rate greater than $\max_{p(y|x)\in\mathscr{C}}H(X|Y)$.

4.4 Staircase scheme

In this section, we assume that the source is a binary symmetric source (uniform) and all the condition distributions in $\mathscr{C}$ are of symmetric channel type.

4.4.1 Code construction

We use polar blocks of size N, where N is sufficiently large for polarization so that we get the $H(X|Y)$ fraction of bad bit-channels ($\mathscr{H}_{X|Y}$) and $1-H(X|Y)$ fraction of good bit-channels ($\mathscr{L}_{X|Y}$) for each $p(y|x)\in\mathscr{C}$. We need an MDS code in the code construction. We use a Reed-Solomon (RS) code of block length 2^q-1 over a field GF(2^q) as an MDS code where q is $\log_2(N)$. We consider the RS code with $L=\min_{p(y|x)\in\mathscr{C}}|\mathscr{L}_{X|Y}|-1$ information bits. Let $\mathscr{M}$ be the set of codewords of the RS code.

We arrange N polar blocks of size N one above the other like a staircase which will be of height N. We extend the staircase by placing $k\in\mathbb{N}$ such staircases side-by-side. Now place q such extended staircases, one above the other. So the total number of polar blocks would be Nqk. This is illustrated in Fig. 5.5 with $N=6$, $k=3$, and $q=2$.

A staircase scheme designed for a compound channel with the class $\mathscr{C}$ of symmetric channels [22] is a binary linear code after all. Based on the linear channel code, a naive Slepian-Wolf code derived directly using the method [55] requires the computation of high dimensional systematic parity-check matrix ($q[(N-L)(1+N(k-1))+N(N-1)]\times N^2qk$) of the staircase universal channel code. We avoid the computation of such a high dimensional parity-check matrix and its use in our staircase code construction. We can also get a delay saving by continuous, sequential encoding and decoding of substaircases in our staircase implementation, similar to

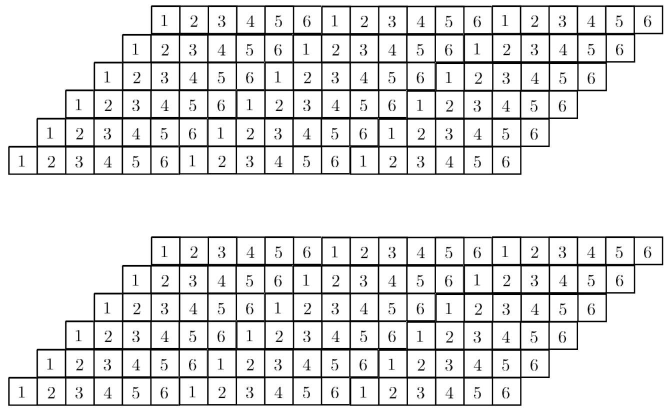

Figure 4.1. Staircase with $k = 3$, $N = 6$ and $q = 2$

universal channel coding in Chapter 2.

While encoding, we do the compression for all the polar blocks column-by-column from left to right in the staircase structure, and we follow the same order for decoding. So, we encode/decode the bit-channels of different polar blocks in parallel while encoding/decoding a column. The total number of columns is $(k+1)N-1$, and we label them with indices $1:(k+1)N-1$ from left to right. Now we describe the encoding algorithm for the compression.

Encoding

Input: $X^{1:N}$ source sequence corresponding to each polar block in all q staircases.

Output: Compressed bit stream for all the columns.

- Compute $U^{1:N} = X^{1:N} G_N$ for each polar block in all the q staircases.
- Now we start encoding the non-full-height columns on the left.
  - The U_is of the non-full-height columns on the left side are transmitted as is without any further compression for all q staircases. Note that this will not affect the rate as the fraction of these columns is diminishing as k approaches infinity.
- Next we start encoding the full-height columns $t = N \leq i \leq kN$.

- In the full-height column t, there is a polar block corresponding to each index $i \in [N]$ in all q staircases.
- In the column t, for each $i \in [N]$, the U_is of corresponding polar blocks for all q staircases will be read. Those q bits corresponding to the index $i \in [N]$ will be interpreted as a field element in GF(2^q) in its binary representation. Hence we can read N finite field elements in the column. Let us call this vector $V^{1:N}$ in GF(2^q).
- $V^{1:N-1}$ will be decomposed as an RS codeword and the error vector in a unique way using a systematic encoding method for the RS code.
- We designate the positions $1:L$ for information symbols. So, we generate the codeword in $\mathscr{M}$ corresponding to data $V^{1:L}$ by systematic encoding. The parity symbols in the systematic encoding method for the RS code can be computed by determining a remainder of a polynomial. This can be implemented using a shift register circuit with multipliers and adders [45].
- Let the encoded codeword be $V'^{1:N-1}$. Now the error vector will be $E^{1:N-1} = V^{1:N-1}$ - $V'^{1:N-1}$. Note that $E^{1:L}$ will be zero always. We set the Nth position of the error vector $E^{1:N}$, $E_N = V_N$.
- We transmit $E^{L+1:N}$ in the binary representation.
- The error vector $E^{1:N}$ can also be generated by computing the syndrome of $V^{1:N}$ using the systematic parity-check matrix. This is shown in Lemma 8. However we propose to use the shift register circuit implementation to get the systematic RS codeword [45] without explicitly computing the systematic generator or parity-check matrix.
- This decomposition is also equivalent to standard array decoding with coset leaders of the form $[0^{1:L}, x^{L+1:N-1}]$ where $x^{L+1:N-1}$ is a vector with elements of GF(2^q).

• Now we encode the non-full-height columns on the right.

– The U_is of the non-full-height columns on the right side are transmitted as is without any further compression for all q staircases. Note that this will not affect the rate as these columns are diminishing in fraction as k approaches ∞.

Lemma 8. *Let $V^{1:N-1}$ be any $N-1$ dimensional vector over GF(2^q). Let $V'^{1:N-1}$ be the Reed-Solomon codeword in $\mathscr{M}$ corresponding to the data symbol stream $V^{1:L}$ in the systematic representation. Let $E^{1:N-1}$ be the error $V^{1:N-1} - V'^{1:N-1}$. The syndrome of the word $V^{1:N-1}$ when computed with the systematic parity-check matrix becomes $E^{L+1:N-1}$.*

Proof: Let the systematic parity-check matrix be

$$H_{sys} = \begin{bmatrix} A & I \end{bmatrix}$$

where A is a $N-1-L \times L$ dimensional matrix in GF(2^q) and I is the $N-1-L \times N-1-L$ dimensional identity matrix. Then,

$$\begin{aligned} H_{sys}(V^{1:N-1})^T &= H_{sys}((V'^{1:N-1})^T + (E^{1:N-1})^T) \\ &\overset{(a)}{=} H_{sys}(E^{1:N-1})^T \\ &= \begin{bmatrix} A & I \end{bmatrix} (E^{1:N-1})^T \\ &\overset{(b)}{=} (E^{L+1:N})^T. \end{aligned}$$

We get the identity (a) because $V'^{1:N-1}$ is a codeword in $\mathscr{M}$. So its multiplication with systematic parity check matrix should be zero. Identity (b) follows because $E^{1:L}$ is a zero vector. □

Before turning to the decoding algorithm, let us define

$$U'^{1:N} := U^{1:N} - E'^{1:N}$$

for each polar block in all q staircases, where $E'^{1:N}$ is the horizontal error vector computed for each polar block in all q staircases from the vertical error vectors $E^{1:N}$ corresponding to each full-height column. So we have,

$$U^{1:N}G_N = U'^{1:N}G_N + E'^{1:N}G_N.$$

That implies,

$$X^{1:N} = U'^{1:N}G_N + S^{1:N}.$$

where $S^{1:N} = E'^{1:N}G_N$ for each polar block in all q staircases. Note that we are transmitting $N-L$ bits for each full-height column. The rate for each full-height column is $\frac{N-L}{N}$, which can be made arbitrarily close to $\max_{p(y|x)\in\mathscr{C}} H(X|Y)$ for sufficiently large N. We did not compress the bit-stream corresponding to the non-full-height columns, but their effect on the overall rate can be made arbitrarily small for a sufficiently large k because as k goes to ∞, the fraction of the number of bits in the non-full-height columns with respect to total block length goes to zero.

Lemma 9. *Let $u'^{1:N}$ and $s^{1:N}$ be any two binary vectors. The conditional distribution of permuted side information $s^{1:N}.Y^{1:N}$ given $X^{1:N} = u'^{1:N}G_N + s^{1:N}$ will be the same as the conditional distribution of the received vector given the word $u'^{1:N}G_N$ is transmitted over the symmetric channel $p(y|x)$.*

Proof:

$$\begin{aligned}
P(s^{1:N}.Y^{1:N} = y^{1:N} | X^{1:N} = u'^{1:N}G_N + s^{1:N}) & \\
&= P(Y^{1:N} = s^{1:N}.y^{1:N} | X^{1:N} = u'^{1:N}G_N + s^{1:N}) \\
&= \Pi_{i\in 1:N} p(s_i.y_i | (u'^{1:N}G_N)_i + s_i) \\
&= \Pi_{i\in 1:N} p(s_i.y_i | (u'^{1:N}G_N)_i + s_i) \\
&\overset{(a)}{=} \Pi_{i\in 1:N} p(s_i.s_i.y_i | (u'^{1:N}G_N)_i) \\
&= \Pi_{i\in 1:N} p(y_i | (u'^{1:N}G_N)_i).
\end{aligned}$$

The identity (a) follows from the symmetric channel property. $\Pi_{i\in 1:N} p(y_i|(u'^{1:N}G_N)_i)$ is precisely the conditional probability of getting the received vector $y^{1:N}$ given the word $u'^{1:N}G_N$ is transmitted over the symmetric channel $p(y|x)$. This concludes the proof. □

Fact 1: Let $u'^{1:N}$ be any binary vector. The conditional probability of error of all the bit-channels in $\mathscr{L}_{X|Y}$ when conventional decision rules [1] are used is upper bounded by 2^{-N^β} given that $u'^{1:N}G_N$ is transmitted over the symmetric channel $p(y|x)$ for any $\beta < 0.5$.

The above fact follows from Arikan's capacity-achieving polar coding construction for symmetric channels [1] where it was proved that the conditional probability of error of a bit-channel given that any particular word is transmitted over the channel remains same irrespective of the word that is transmitted. Now we start describing the decoding algorithm.

Decoding

Input: Side information $Y^{1:N}$ for each block and $E^{L+1:N}$ for each full-height column.

Output: Estimates of $X^{1:N}$ corresponding to all polar blocks.

- Using error vectors $E^{L+1:N}$ corresponding to all full-height columns, the decoder computes the horizontal error vectors $E'^{1:N}$ corresponding to all polar blocks in all q staircases.
- Now the decoder estimates U'_is of each column that corresponds to different polar blocks from left to right. Then the estimation of $U'^{1:N}$ leads to estimate $U^{1:N}$ by adding $E'^{1:N}$ for all polar blocks in all q staircases . Let the estimates be denoted as $\hat{U}'^{1:N}$, $\hat{U}^{1:N}$ and $\hat{X}^{1:N}$.
- Decoding the non-full-height columns on the left side.
  - The U_is corresponding to these columns are transmitted as is by the encoder.
  - Hence $\hat{U}'_i = U_i - E'_i = U_i$ for all these columns in all q staircases.
- To decode full-height columns from $t = N \leq i \leq kN$:
  - The decoder has knowledge of the exact $p(y|x) \in \mathscr{C}$. For the blocks corresponding to bit-channels $i \in \mathscr{L}_{X|Y}$, we use the following decision rule to decode U'_is. It recovers

those U_i's reliable due to Lemma 9 and Fact 1.

$$\hat{U}_i' = \Phi_i(\hat{U}'^{1:i-1}, S^{1:N}.Y^{1:N}).$$

- Decode $\hat{U}_i'$ as 0 for the block corresponding to the component V_N of vector $V^{1:N}$ in q staircases.
- Now we have at least L positions of the MDS codeword that are recovered. Now the erasure decoding of the MDS code recovers all $N-1$ positions of the codeword.
- Hence all $\hat{U}_i'$s corresponding to all polar blocks in the column are estimated in all q staircases. This enables the continuation of SC decoding of the polar blocks to estimate $\hat{U}_i'$s corresponding to the next column.

- Decoding non-full-height columns on the right side.
  - The U_is corresponding to these columns are transmitted as is by the encoder.
  - Hence $\hat{U}_i' = U_i - E_i' = U_i$ for all these columns in all q staircases.
- Now $\hat{U}^{1:N} = \hat{U}'^{1:N} + E'^{1:N}$ for each polar block.
- $\hat{X}^{1:N} = \hat{U}^{1:N} G_N$ for each block.

Theorem 5.

The probability of error for the above staircase scheme is $O(Nqk2^{-N^\beta})$ *for* $\beta < 0.5$.

Proof:

We decode $U'^{1:N}$ corresponding to all the polar blocks. The error occurs if and only if there is an error in decoding some good-bit-channel ($\mathscr{L}_{X|Y}$) of any polar block. If U_is for good bit-channels are recovered properly, then other U_is are recovered either by MDS erasure

decoding in a full-height column or by the knowledge of U_is at the receiver corresponding to the non-full-height columns. Let the error event be $\mathscr{E}$.

Let $\mathscr{E}_g$ be the error event with a genie aided decoder which has the accurate values of the past $U'^{1:i-1}$ when decoding any bit-channel $i \in \mathscr{L}_{X|Y}$ for all polar blocks. Let all the polar blocks in all of the q staircases be indexed as $b = 1, 2..., Nqk$. Let $\mathscr{E}_{ib}$ be the error event corresponding to an error in the ith bit-channel of block b. If bit-channel $i \in \mathscr{L}_{X|Y}$ of the polar block b lies in a full-height column, then the error event $\mathscr{E}_{ib}$ becomes as follows.

$$\begin{aligned}\mathscr{E}_{ib} = \{&(u'^{1:N}, y^{1:N}, s^{1:N})s \text{ of all blocks } [Nqk] : \\ &\Phi_i(u'^{1:i-1}, s^{1:N}.y^{1:N}) \neq u'_i \text{ holds for block } b\}.\end{aligned}$$

Note that $\mathscr{E}_{ib}$ will be the null event, if the block b has bit-channel i that lies in a non-full-height column. Clearly, $\mathscr{E}_g = \cup_{b\in\{1:Nqk\}} \cup_{i\in\mathscr{L}_{X|Y}} \mathscr{E}_{ib}$. Note that error event $\mathscr{E}$ will imply at least one of the $\mathscr{E}_{ib}$s. So we should have the following.

$$\mathscr{E} \subset \mathscr{E}_g.$$

Now the probability of error $P(\mathscr{E})$ is upper bounded as follows.

$$\begin{aligned}P(\mathscr{E}) \leq P(\mathscr{E}_g) &= P(\cup_{b\in\{1:Nqk\}} \cup_{i\in\mathscr{L}_{X|Y}} \mathscr{E}_{ib}) \\ &\overset{(a)}{\leq} \sum_{b\in\{1:Nqk\}} \sum_{i\in\mathscr{L}_{X|Y}} P(\mathscr{E}_{ib}).\end{aligned}$$

The identity (a) follows from the union bound. So, we need to bound $P(\mathscr{E}_{ib})$ for $i \in \mathscr{L}_{X|Y}$ for all polar blocks.

Now we evaluate the conditional probability of error of bit-channel $i \in \mathscr{L}_{X|Y}$ for the block

b given the random vectors $(U'^{1:N}, Y^{1:N}, S^{1:N})$ corresponding to the block b.

$$\begin{aligned}
P(\mathscr{E}_{ib}|U'^{1:N} = u'^{1:N}, S^{1:N} = s^{1:N}) &= P(\Phi_i(U'^{1:i-1}, S^{1:N}.Y^{1:N}) \neq U'_i| \\
&\qquad U'^{1:N} = u'^{1:N}, S^{1:N} = s^{1:N}) \\
&= P(\Phi_i(u'^{1:i-1}, s^{1:N}.Y^{1:N}) \neq u'_i| \\
&\qquad U'^{1:N} = u'^{1:N}, S^{1:N} = s^{1:N}) \\
&\overset{(a)}{=} \sum_{y^{1:N}} \Pi_{i\in[1:N]} p(y_i|(u'^{1:N}G_N)_i) \\
&\qquad \mathbb{1}(\Phi_i(u'^{1:i-1}, y^{1:N}) \neq u'_i) \\
&\overset{(b)}{\leq} Z_i \\
&= 2^{-N^\beta}.
\end{aligned} \tag{4.1}$$

The identity (a) follows from Lemma 9. Identity (b) follows from Arikan's [1] symmetric channel polar coding construction where it was proved that the conditional probability of error of a bit-channel given that any particular word is transmitted over the channel is always the same irrespective of the word that is transmitted. Also, all of those conditional probabilities of errors are upper bounded by the Bhattacharyya parameter of the bit-channel. This is essentially stated as Fact 1. Now the actual probability of error of bit-channel i for the block b satisfies

$$\begin{aligned}
P(\mathscr{E}_{ib}) &= \sum_{(u'^{1:N}, s^{1:N}) \text{ of block } b} P((U'^{1:N} = u'^{1:N}, S^{1:N} = s^{1:N}) \text{ of block } b) \\
&\qquad P(\mathscr{E}_{ib}|U'^{1:N} = u'^{1:N}, S^{1:N} = s^{1:N} \text{ of block } b) \\
&\leq \sum_{(u'^{1:N}, s^{1:N}) \text{ of block } b} P((U'^{1:N} = u'^{1:N}, S^{1:N} = s^{1:N}) \text{ of block } b) 2^{-N^\beta} \\
&= 2^{-N^\beta}.
\end{aligned}$$

Therefore,

$$P(\mathscr{E}) \leq \sum_{b\in\{1:Nqk\}} \sum_{i\in\mathscr{L}_{X|Y}} P(\mathscr{E}_{ib})$$
$$\leq O(Nqk2^{-N^{\beta}}).$$

Hence the proof of Theorem 5. □

4.4.2 Coding with non-uniform source

If the conditional distributions in $\mathscr{C}$ are of symmetric type, then any rate greater than $\max_{p(y|x)\in\mathscr{C}} H(\tilde{X}|Y)$ can still be achieved using the staircase method irrespective of the source distribution $P_X(x)$. Here the random variable pair $(\tilde{X},Y)$ is distributed as $P_{\tilde{X}}(x)p(y|x)$ and $P_{\tilde{X}}(x) = 0.5$. The subtle idea is to implement the same code construction as if the source is uniformly distributed. We use the bit-channels $\mathscr{L}_{\tilde{X}|Y}$ in the code construction irrespective of the source distribution. The conditional probability of error of the bit-channel $i \in \mathscr{L}_{\tilde{X}|Y}$ is the same given any source sequence due to the symmetric channel property of the conditional distribution $p(y|x)$. Hence the average probability of error for the bit-channel $i \in \mathscr{L}_{\tilde{X}|Y}$ does not depend on the source distribution. This can be noticed from equation (4.1) in the proof of Theorem 5. Therefore the probability of error will still be $O(Nkq2^{-N^{\beta}})$ for $\beta < 0.5$.

4.4.3 Encoding and decoding complexity

The encoding complexity consists of decomposing the vector of length $N-1$ in GF(2^q) into a RS codeword and the corresponding error vector. We proposed to use the shift register circuit with adders and multipliers to get a systematic RS codeword for executing the decomposition. This takes $O(L(N-L)) = O(N^2)$ multiplications and additions in GF(2^q). Addition and multiplication over this field take q and $q^{\log_2 3}$ binary operations, respectively. Hence the bit operations sum upto $O(N^2 q^{\log_2 3})$ for each full-height column. Applying polar transform for each polar block also takes $O(N\log_2 N)$ bit-operations.

Decoding complexity consists of computing the polar transform $S^{1:N} = E'^{1:N} G_N$ for all polar blocks, SC decoding of all the polar blocks and also the erasure decoding of the RS codes of length $N-1$ over GF(2^q) for each full-height column. Applying polar transform for each polar block takes $O(N\log_2 N)$ bit-operations. The SC decoding of a polar block takes $O(N\log_2 N)$ real operations. The erasure decoding of the RS codes can be done in O($N(\log_2(N))^2$) symbol operations [22]. Addition and multiplication over this field take q and $q^{\log_2 3}$ binary operations respectively. Hence the bit operations sum to $O(N(\log_2(N))^2 q^{\log_2 3})$ for each full-height column.

4.4.4 Pros and cons

Pros:

The upside of the scheme is that it can be designed for a class $\mathscr{C}$ with infinite cardinality as well. The block length does not increase with cardinality of the class $\mathscr{C}$. On the other hand, a code designed with rate r supports any arbitrary source with any side information whose conditional distribution given source $p(y|x)$ is of symmetric channel type whenever $r > H(\tilde{X}|Y)$. Here the random variable pair $(\tilde{X},Y)$ is distributed as $P_{\tilde{X}}(x)p(y|x)$ and $P_{\tilde{X}}(x) = 0.5$.

Cons:

The downside is that it can be applied only when class $\mathscr{C}$ contains only the conditional distributions of symmetric channel type. Also, for the non-uniform source with distribution $P_X(x)$, the staircase construction does not support all the rates greater than $\max_{p(y|x)\in\mathscr{C}} H(X|Y)$ where the random variable pair (X,Y) is distributed as $P_X(x)p(y|x)$.

4.5 Scheme based on combining bit-channels

In this scheme, we assume the class $\mathscr{C}$ contains a finite number of conditional distributions. Let $|\mathscr{C}|$ be s. The bit-channel sets $\mathscr{L}_{X|Y}$ and $\mathscr{H}_{X|Y}$ may not be the same for all $p(y|x) \in \mathscr{C}$. The obvious approach is to share $U^{(\cap_{i\in\mathscr{C}}\mathscr{L}_{X|Y})^c}$, so that decoder can reliably decode the other bits corresponding to bit-channels in $(\cap_{i\in\mathscr{C}}\mathscr{L}_{X|Y})$ by SC decoding. The scheme based on bit-channel combining is a recursive procedure of combining polar blocks that increases the fraction of

bit-channels $\cap_{p(y|x)\in\mathscr{C}}\mathscr{L}_{X|Y}$ with respect to the updated polar block length. The fraction of bit-channels $\cap_{p(y|x)\in\mathscr{C}}\mathscr{L}_{X|Y}$ with respect to the updated polar block length in the recursive procedure can get arbitrarily close to $1-\max_{p(y|x)\in\mathscr{C}}H(X|Y)$ when N is sufficiently large. Hence, this gives the compression algorithm that can achieve any rate greater than $\max_{p(y|x)\in\mathscr{C}}H(X|Y)$. Hassani and Uranke [22] essentially did this for a symmetric source in the context of universal channel coding. We validate that such a recursive method can be used for a non-uniform memoryless source setting as well. So, this method is straightforward to use in this source coding setting in view of the original scheme [22] proposed in the context of universal channel coding.

We need Proposition 1 to validate this method for an arbitrary discrete memoryless source (which may be non-uniform) with the arbitrary class $\mathscr{C}$ (which may contain non-symmetric $p(y|x)$) of finite cardinality.

We now validate the method with an arbitrary memoryless source while recalling the idea of this method proposed in [22]. Let $\mathscr{C}=\{p_1(y|x),p_2(y|x),....,p_s(y|x)\}$. The first step is to increase the fraction of bit-channels $\mathscr{L}_{X|Y_1}\cap\mathscr{L}_{X|Y_2}$ with respect to the updated block length. To do this, first consider the two independent polar blocks $U^{1:N}=X^{1:N}G_N$ and $U'^{1:N}=X'^{1:N}G_N$, where $Y^{1:N}$ and $Y'^{1:N}$ are the correlated side information vectors corresponding to the two blocks, respectively. Then combine the bit-channels $\mathscr{L}_{X|Y_1}\cap\mathscr{H}_{X|Y_2}$ of the first block with bit-channels $\mathscr{L}_{X|Y_2}\cap\mathscr{H}_{X|Y_1}$ in the order. Suppose the bit-channel $i\in\mathscr{L}_{X|Y_1}\cap\mathscr{H}_{X|Y_2}$ with input U_i and output $U^{1:i-1}Y^{1:N}$ from the first polar block is combined with bit-channel $j\in\mathscr{L}_{X|Y_2}\cap\mathscr{H}_{X|Y_1}$ with input U'_j and output $U'^{1:j-1}Y'^{1:N}$ from the second polar block. One of the two new bit-channels produced by this combining has the input $U_i+U'_j$ and the output $U^{1:i-1}U'^{1:j-1}Y^{1:N}Y'^{1:N}$; the other bit-channel produced has the input U'_j and the output $U_i+U'_j,U^{1:i-1}U'^{1:j-1}Y^{1:N}Y'^{1:N}$. By Proposition 1, the second bit-channel produced by the combining has the Bhattacharyya parameter

$$Z(U'_i|U_i+U'_j,U^{1:i-1}U'^{1:j-1}Y^{1:N}Y'^{1:N})$$

$$= Z(U_i|U^{1:i-1}Y^{1:N})Z(U'_j|U^{1:j-1}Y^{1:N})) \overset{(a)}{\leq} O(2^{-N^\beta}).$$

where $\beta < 0.5$. The identity (a) is true because either the Bhattacharyya parameter $Z(U_i|U^{1:i-1}Y^{1:N}$) is 2^{-N^β} if the conditional distribution is $p_1(y|x)$ or the Bhattacharyya parameter $Z(U'_j|U'^{1:j-1}Y^{1:N}$) is 2^{-N^β} if the conditional distribution is $p_2(y|x)$. So we have $G = \min\{|\mathscr{L}_{X|Y_2} \cap \mathscr{H}_{X|Y_1}|, |\mathscr{L}_{X|Y_1} \cap \mathscr{H}_{X|Y_2}\}|$ new bit-channels that come into the category of $\mathscr{L}_{X|Y_1} \cap \mathscr{L}_{X|Y_2}$ in the updated polar block of length $2N$. We use a bold font from now on to denote the bit-channels in the updated polar block to disitnguish them from the bit-channels of the original polar block. Now the fraction of the updated bit-channels $\mathscr{L}_{\mathbf{X}|\mathbf{Y_1}} \cap \mathscr{L}_{\mathbf{X}|\mathbf{Y_2}}$ with respect to the updated block length becomes as $\frac{2(\mathscr{L}_{X|Y_1} \cap \mathscr{L}_{X|Y_2})+G}{2N}$.

The procedure can be done recursively. In stage t of the recursive procedure, we take two polar blocks obtained in stage $t-1$ and perform the same bit-channel combinings that were mentioned in the first step. After t recursions, the fraction of the updated $\mathscr{L}_{\mathbf{X}|\mathbf{Y_1}} \cap \mathscr{L}_{\mathbf{X}|\mathbf{Y_2}}$ with respect to the updated block length becomes

$$\frac{2^t(\mathscr{L}_{X|Y_1} \cap \mathscr{L}_{X|Y_2}) + (2^t - 1)G}{2^t N}.$$

This will increase and become closer to $|\mathscr{L}_{X|Y_1} \cap \mathscr{L}_{X|Y_2}| + G = \min\{|\mathscr{L}_{X|Y_1}|, |\mathscr{L}_{X|Y_2}|\}$ per block length N as t grows. Now let the bit-channels $\mathscr{L}_{\mathbf{X}|\mathbf{Y_1}} \cap \mathscr{L}_{\mathbf{X}|\mathbf{Y_2}}$ in the updated polar block be $\mathscr{L}_{\mathbf{12}}$ and repeat the same recursive procedure to increase the bit-channels $\mathscr{L}_{12} \cap \mathscr{L}_{X|Y_3}$. We continue the recursive procedure until we finish all $p(y|x) \in \mathscr{C}$. Hence by this method, one can increase the cardinality of bit-channels $\cap_{p(y|x)\in\mathscr{C}} \mathscr{L}_{\mathbf{X}|\mathbf{Y}}$ per block length N that can get arbitrarily close to $\min_{p(y|x)\in\mathscr{C}} |\mathscr{L}_{X|Y}|$. Order of bit-channels are governed by the recursive combinings done to produce the hybrid block. Other details for this method are given in [22]. The scheme supports any non-uniform source with an arbitrary class $\mathscr{C}$ of finite cardinality. But the block length can become unbounded as the cardinality of the class $\mathscr{C}$ grows, in contrast to the staircase scheme.

4.6 Conclusion

We defined the problem of source coding with side information at the receiver whose correlation is unknown to the encoder. We studied two coding strategies based on polar codes for this problem. The code designed by the staircase scheme with rate r supports any source with any side information whose conditional distribution given source $p(y|x)$ is of symmetric channel type whenever $r > H(\tilde{X}|Y)$. Here the random variable pair $(\tilde{X}, Y)$ is distributed as $P_{\tilde{X}}(x)p(y|x)$ and $P_{\tilde{X}}(x) = 0.5$. A naive Slepian-Wolf code derived using the method [55] requires the computation of a high dimensional systematic parity-check matrix ($q[(N-L)(1+N(k-1))+N(N-1)] \times N^2qk$) for the staircase universal channel code. We avoid the computation of such a high dimensional parity-check matrix and its use in the proposed staircase code construction. The second scheme is based on the technique of universalization using bit-channel combining. Using this method, we can design a code for a non-uniform source with arbitrary $\mathscr{C}$ of finite cardinality. An open problem is to find a stronger coding strategy where a code designed for an arbitrary source X at rate r can support any correlated side information Y whenever $r > H(X|Y)$.

Acknowledgement

This chapter is in part a reprint of the material in the paper: Karthik Nagarjuna Tunuguntla, Paul H. Siegel, "Slepian-Wolf polar coding with unknown correlation,"*2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 132-139, October, 2019. Dissertation author is the primary contributor of the paper.

Chapter 5

Polar Coding for Multi-level 3-Receiver Broadcast Channels

5.1 Introduction

5.1.1 Background

Arikan [1] constructed capacity-achieving polar codes for binary input symmetric channels. Since then, many coding strategies have been introduced for multi-user settings using the polarization method [3], [27], [2]. Goela, Abbe and Gastpar [17] introduced polar codes for m-user deterministic broadcast channels. They also introduced polar coding for 2-user noisy broadcast channels. They implemented superposition and Marton schemes which involve some assumptions of degradation on the channel parameters to align the polar indices. Mondelli, Hassani, Sason, and Urbanke [32] proposed schemes to remove such constraints using a polar-based chaining construction [22], [23]. Chou and Bloch [11] proposed a polar coding scheme for a broadcast channel with confidential messages. Alos and Fonollosa [39] proposed a polar coding scheme for a broadcast channel with two legitimate receivers, that receive a confidential and private message, and one eavesdropper.

In this paper, we consider the problem of achieving the rates in the capacity region for a discrete memoryless (DM) 3-receiver broadcast channel with degraded message sets [38], [14]. The second receiver is degraded with respect to the first receiver. The problem is to transmit

a public message intended for all three receivers and a private message intended for the first receiver. Our motivation to consider this problem for the broadcast channel came from a very useful practical file transfer application in a client-server network. We now describe the file transfer application in a client-server network, where our problem setting is applied, in the following sub-section.

5.1.2 Motivation with a client-server model

We consider a client server model, in which server sends data to its three clients that are computer, phone-1 and phone-2. Computer and phone-1 receive data from the server directly via internet. Phone-2 receives data from server indirectly through bluetooth connection between computer and phone-2. Computer supports both video and audio applications whereas the two phones only support audio application. Suppose that server has one audio and one video file to send its clients. Note that all three clients here are interested in receiving the audio file. Also notice that computer is the only client interested in receiving both the audio and video files. This is shown in the Fig.5.1. In some scenarios, computer may just want to receive the video file.

The internet link and bluetooth link add noise to the signal received at the clients. So this can be modelled as noisy broadcast channel with 3-receivers which we are interested to look at in this paper. The goal is to find a method where the server can send both the audio and video files to its clients reliably at all possible data rates that can be supported by the network. It amounts to finding a coding scheme that can achieve rates in the capacity region of this broadcast channel problem. This particular client-server setting with the file transfer scenario is highly applicable to household internet links, which is why we are motivated to consider this problem.

We define the coding problem for the DM multi-level 3-receiver broadcast channel with degraded message sets in the following subsection. In general, the setting of a broadcast channel with degraded message sets arises in video or music broadcasting over a wireless network at varying levels of quality [14].

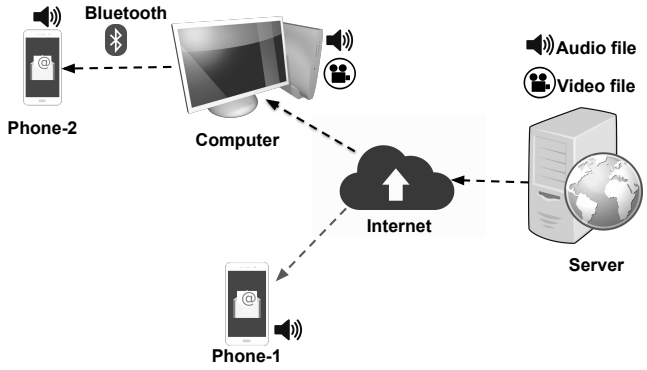

Figure 5.1. A client-server network with 3 clients

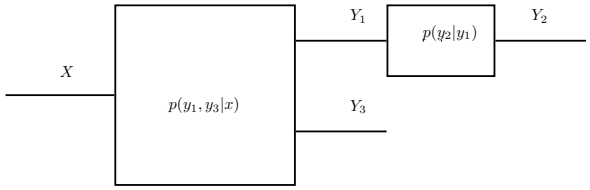

Figure 5.2. A 3-receiver broadcast channel model

5.1.3 Coding problem of DM (discrete memoryless) multi-level broadcast channel with degraded message sets

The 3-receiver multi-level broadcast channel that we consider consists of a finite input alphabet $\mathscr{X}$ and arbitrary output alphabets $\mathscr{Y}_j$ for each output at the receiver-j for $j \in \{1,2,3\}$. The conditional distribution of outputs at receiver-1 and receiver-3 given the input, i.e. $p_{Y_1,Y_3|X}(y_1,y_3|x)$, along with the conditional distribution of output at receiver-2 given the output at receiver-1, i.e. $p_{Y_2|Y_1}(y_2|y_1)$, are given for this broadcast channel setting, where X is the input, Y_j is output at receiver-j for $j = 1,2,3$, $x \in \mathscr{X}$ and $y_j \in \mathscr{Y}_j$ for each $j \in \{1,2,3\}$. These two conditional distributions define this broadcast channel with three receivers since the output at the receiver-2 is degraded with respect to output at receiver-1. The broadcast channel model is shown in the Fig. 5.2.

Now we define the coding problem to transmit a public message for all the receivers and a private message intended only for receiver-1.

A $(2^{NR_0}, 2^{NR_1}, N)$ code consists of

- a message set for public message: $\{1,2,\ldots,2^{NR_0}\}$
- a message set for private message of receiver-1: $\{1,2,\ldots,2^{NR_1}\}$
- an encoder $X^N : \{1,2,\ldots,2^{NR_0}\} \times \{1,2,\ldots,2^{NR_1}\} \to \mathscr{X}^n$,
- a decoder at receiver-1 $h_1 : \mathscr{Y}_1^N \to \{1,2,\ldots,2^{NR_0}\} \times \{1,2,\ldots,2^{NR_1}\}$,
- a decoder at receiver-2 $h_2 : \mathscr{Y}_2^N \to \{1,2,\ldots,2^{NR_0}\}$,
- a decoder at receiver-3 $h_3 : \mathscr{Y}_3^N \to \{1,2,\ldots,2^{NR_0}\}$.

where N is the block length, R_0 is the rate of the public message and R_1 is the rate of the private message. Let M_0 be the public message which is chosen uniformly from the set $\{1,2,\ldots,2^{NR_0}\}$ and M_1 be the private message of receiver-1 which is chosen uniformly from the set $\{1,2,\ldots,2^{NR_1}\}$. Let $Y_j^{1:N}$ be the output vector at receiver-j where $j \in \{1,2,3\}$. Let $P_e^{(N)} = P\big((h_1(Y_1^{1:N}) \neq (M_0,M_1)) \cup (h_2(Y_2^{1:N}) \neq M_0) \cup (h_3(Y_3^{1:N}) \neq M_0)\big)$ be the probability of error. If there is a sequence of $(2^{NR_0}, 2^{NR_1}, N)$ codes, for which the $P_e^{(N)}$ goes to zero, then the rate (R_0, R_1) is achieved. The closure of all such achievable rate pairs is the capacity region.

5.1.4 Contribution

In this paper, we use a polar coding strategy to achieve the rates in the capacity region for the multi-level 3-receiver broadcast with degraded message sets without time-sharing. This represents the first time in the literature that polar coding for 3-receiver broadcast channels without eavesdropper is considered.

Three layered polarization results are established using auxiliary random variables that characterize the capacity region. We do a suitable rate splitting of the private message of receiver-1 for the implementation of our polar coding strategy. We use a chaining construction at two levels, one of which is within first and second layers whereas the second level of chaining is

done within the second layer. The two-level chaining construction that we provide essentially translates into polar coding strategy the ideas of three layered superposition coding and more importantly, indirect-coding [38] with the rate splitting of the private message.

The two-level chaining construction is new in the context of reliable decoding at three receivers. In particular, first level of chaining is done to recover public message by all the receivers. Second level of chaining helps to recover the split of private message reliably at receiver-1 while translating indirect coding of public message for receiver-3. In contrast, note that Marton's coding [32] uses a two-level chaining construction, where first level of chaining is to align good bit-channels of the two receivers and the second level of chaining is to maintain the joint distribution of auxiliary random variables involved.

We also consider a slight variation to the problem of degraded message sets. Suppose that receiver-1 requires to decode only M_1. Then M_1 becomes private message to receiver-1 and M_0 is common private message to receiver-2 and receiver-3. We show that the capacity region does not enlarge by relaxing the decoding constraint at receiver-1. So the same polar coding strategy achieves the capacity region of the modified problem. This is an interesting observation, as we know that for any 2-receiver broadcast channel, superposition coding is not optimal in general, unless it is a problem with degraded message sets.

5.1.5 Organization

The paper is organized as follows. In Section 5.2, we introduce some notations and recall some background results. In Section 5.3, we give our chaining construction to achieve the rate pairs in the capacity region of the 3-receiver broadcast channel with degraded message sets and provide the detailed decoding error analysis. In Section 5.3, we also show that the capacity of the broadcast channel remains same, even when receiver-1 is relaxed to recover only its private message. In Section 5.4, we conclude the paper.

5.2 Preliminaries

We denote the set $\{1,2,\ldots,n\}$ as $[n]$ where $n \in \mathscr{Z}^+$. Let G_N be the conventional polar transform [1], represented by a binary matrix of dimension $N \times N$ where $N = 2^n$, $n \in \mathscr{Z}^+$.

Let X be a binary random variable. Let the random variable pair (X,Y) be distributed as $P_{X,Y}(x,y)$, then the Bhattacharya parameter is defined as

$$Z(X|Y) = 2\sum_y P_Y(y)\sqrt{P_{X|Y}(1|y)P_{X|Y}(0|y)}.$$

The following are the identities from [24, Proposition 1] which provides the relationship between entropy and Bhattacharya parameter.

$$(Z(X|Y))^2 \leq H(X|Y) \tag{5.1}$$

$$H(X|Y) \leq \log(1+Z(X|Y)) \leq Z(X|Y) \tag{5.2}$$

The **capacity region** for this multi-level 3-receiver broadcast problem [14], [38] is as follows:

$$R_0 < \min\{I(W;Y_2), I(V;Y_3)\} \tag{5.3}$$

$$R_1 < I(X;Y_1|W) \tag{5.4}$$

$$R_0 + R_1 < I(V;Y_3) + I(X;Y_1|V) \tag{5.5}$$

for some joint distribution $p(w,v)p(x|v)$ with $|\mathscr{W}| \leq |\mathscr{X}|+4$ and $|\mathscr{V}| \leq (|\mathscr{X}|+1)(|\mathscr{X}|+4)$. Here W and V are random variables over the alphabets $\mathscr{W}$ and $\mathscr{V}$, respectively, Y_j is the output at receiver-j when X is input for $j = 1,2,3$.

Let $(W_i, V_i, X_i)_{i=1}^N$ be the binary triplet random variable sequence that is i.i.d. (identical

and independently distributed) according to distribution $p(w,v)p(x|v)$. So $|\mathscr{X}| = |\mathscr{Y}| = |\mathscr{V}| = 2$. Let (W,V,X) also be binary random triplet distributed according to $p(w,v)p(x|v)$. Let $Y_j^{1:N}$ be the received vector at receiver-j when the random variable sequence $X^{1:N}$ is transmitted over the 3-receiver discrete memoryless broadcast channel and let Y_j be the output at receiver-j when X is input for $j = 1,2,3$.

Now we establish three-layered polarization results that are going to be used in the code construction.

Let $\beta < 0.5$. Let $(U_w)^{1:N} = W^{1:N}G_N$, we define the following bit-channel subsets as follows for $j = 1,2,3$.

$$\begin{aligned}
\mathscr{H}_W &= \{i \in [N] : Z((U_w)_i|(U_w)^{1:(i-1)}) \geq 1-\delta_n\}.\\
\mathscr{L}_W &= \{i \in [N] : Z((U_w)_i|(U_w)^{1:(i-1)}) \leq \delta_n\}.\\
\mathscr{H}_{W|Y_j} &= \{i \in [N] : Z((U_w)_i|(U_w)^{1:(i-1)}Y_j^{1:N}) \geq 1-\delta_n\}.\\
\mathscr{L}_{W|Y_j} &= \{i \in [N] : Z((U_w)_i|(U_w)^{1:(i-1)}Y_j^{1:N}) \leq \delta_n\}.
\end{aligned}$$

where $\delta_n = 2^{-N^\beta}$. Note that $\mathscr{L}_{W|Y_2} \subseteq \mathscr{L}_{W|Y_1}$ from Lemma 7 in [17] due to the degradation assumption on receiver-2. Then,

$$\begin{aligned}
\lim_{N\to\infty} \frac{|\mathscr{H}_W|}{N} &= H(W), \qquad \lim_{N\to\infty} \frac{|\mathscr{L}_W|}{N} = 1-H(W),\\
\lim_{N\to\infty} \frac{|\mathscr{H}_{W|Y_j}|}{N} &= H(W|Y_j), \lim_{N\to\infty} \frac{|\mathscr{L}_{W|Y_j}|}{N} = 1-H(W|Y_j).
\end{aligned}$$

Let $(U_v)^{1:N} = V^{1:N}G_N$. We now define bit-channel subsets $\mathscr{H}_{V|W}$ and $\mathscr{L}_{V|W}$ based on the Bhattacharyya parameter $Z((U_v)_i|(U_v)^{1:(i-1)}W^{1:N})$ as we did above. Similarly we define the $\mathscr{H}_{V|WY_j}$ and $\mathscr{L}_{V|WY_j}$ based on the value of Bhattacharyya parameter $Z((U_v)_i|(U_v)^{1:(i-1)}W^{1:N}Y_j^{1:N})$ for $j = 1,3$. Then,

$$\lim_{N\to\infty} \frac{|\mathscr{H}_{V|W}|}{N} = H(V|W), \lim_{N\to\infty} \frac{|\mathscr{L}_{V|W}|}{N} = 1-H(V|W),$$

$$\lim_{N\to\infty}\frac{|\mathscr{H}_{V|WY_j}|}{N} = H(V|WY_j),$$
$$\lim_{N\to\infty}\frac{|\mathscr{L}_{V|WY_j}|}{N} = 1-H(V|WY_j).$$

Let $(U_x)^{1:N} = X^{1:N}G_N$. We define the bit-channel subsets $\mathscr{H}_{X|V}$, $\mathscr{L}_{X|V}$ and also $\mathscr{H}_{X|VY_1}$, $\mathscr{L}_{X|VY_1}$ based on the values of Bhattacharyya parameters $Z((U_x)_i|(U_x)^{1:(i-1)}V^{1:N})$ and $Z((U_x)_i|(U_x)^{1:(i-1)}V^{1:N}Y_1^{1:N})$, respectively, as we did above. Then,

$$\lim_{N\to\infty}\frac{|\mathscr{H}_{X|V}|}{N} = H(X|V), \lim_{N\to\infty}\frac{|\mathscr{L}_{X|V}|}{N} = 1-H(X|V).$$
$$\lim_{N\to\infty}\frac{1}{N}|\mathscr{H}_{X|VY_1}| = H(X|VY_1),$$
$$\lim_{N\to\infty}\frac{1}{N}|\mathscr{L}_{X|VY_1}| = 1-H(X|VY_1).$$

Under this probability distribution of $(W^{1:N},V^{1:N},X^{1:N})$, we denote $\mathbb{P}((U_w)^{1:N}=(u_w)^{1:N})$ by $P_{(U_w)^{1:N}}((u_w)^{1:N})$ and similarly we denote $\mathbb{P}((U_v)_i=(u_v)_i\,|W^{1:N}(U_v)^{1:i-1}Y_1^{1:N}=w^{1:N}u_v^{1:i-1}y_1^{1:N}$) by $P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}Y_1^{1:N}}((u_v)_i|w^{1:N}(u_v)^{1:i-1}y_1^{1:N})$.

Define $I_j^w = \mathscr{L}_{W|Y_j}\cap\mathscr{H}_W$ for $j=1,2,3$, $I_j^v = \mathscr{L}_{V|WY_j}\cap\mathscr{H}_{V|W}$ for $j=1,3$, and $I_j^x = \mathscr{L}_{X|VY_j}\cap\mathscr{H}_{X|V}$ for $j=1$. Note that $\lim_{N\to\infty}\frac{|I_j^w|}{N} = I(W;Y_j)$ for $j=1,2,3$, $\lim_{N\to\infty}\frac{|I_j^v|}{N} = I(V;Y_j|W)$ for $j=1,3$, and $\lim_{N\to\infty}\frac{|I_j^x|}{N} = I(X;Y_j|V)$ for $j=1$. We refer to I_j^w $(j=1,2,3)$, I_j^v $(j=1,3)$ and I_j^x $(j=1)$ as information bit-channels of receiver-j in $(U_w)^{1:N}$, $(U_v)^{1:N}$ and $(U_x)^{1:N}$ respectively.

Define $F_j^w = \mathscr{H}_W - I_j^w$ $(j=1,2,3)$, $F_j^v = \mathscr{H}_{V|W} - I_j^v$ $(j=13)$ and $F_j^x = \mathscr{H}_{X|V} - I_j^x$ $(j=1)$. We refer to F_j^w $(j=1,2,3)$, F_j^v $(j=1,3)$ and F_j^x $(j=1)$ as frozen bit-channels of receiver-j in $(U_w)^{1:N}$, $(U_v)^{1:N}$ and $(U_x)^{1:N}$ respectively.

Define $R^w = (\mathscr{H}_W\cup\mathscr{L}_W)^c$, $R^v = (\mathscr{H}_{V|W}\cup\mathscr{L}_{V|W})^c$ and $R^x = (\mathscr{H}_{X|V}\cup\mathscr{L}_{X|V})^c$. We refer to R^w, R^v and R^x as not-completely polarized bit-channels in $(U_w)^{1:N}$, $(U_v)^{1:N}$ and $(U_x)^{1:N}$ respectively.

We denote the subvector of $U^{1:N}$ corresponding to the bit-channel set $\mathscr{A} \subset [N]$ by $U^{\mathscr{A}}$. Let P and Q be any two distributions on a discrete arbitrary alphabet $\mathscr{Z}$. We denote the total variation distance between the two distributions P and Q as $||P-Q||$. Therefore $||P-Q|| = \sum_{z\in\mathscr{Z}} \frac{1}{2}|P(z)-Q(z)| = \sum_{z:P(z)>Q(z)} P(z)-Q(z)$. We denote the KL-divergence between two distributions P and Q as $D(P||Q)$.

5.3 Polar coding for the DM multi-level 3-receiver broadcast channel

In this section, we are going to discuss the polar coding scheme for achieving the capacity region of the DM multi-level 3-receiver broadcast channel with degraded message sets. To achieve the capacity region, we need to achieve the rate pairs that satisfy equations (5.3), (5.4) and (5.5) for all joint distributions on random variables over the alphabets of the required size mentioned in the definition of the capacity region. We consider the case when $|\mathscr{X}| = |\mathscr{V}| = |\mathscr{W}| = 2$ to describe the polar coding scheme. The fundamental idea of the polar coding strategy which we present is applicable even when the alphabets $\mathscr{X}, \mathscr{W}$ or $\mathscr{V}$ are of higher size. In [47], [46], polarization for the alphabets of higher size is discussed.

5.3.1 Typical set coding

Before we go into our polar coding construction, we briefly discuss the achievability of the rate pairs in the capacity region using random coding approach by typical sets [14, p. 200]. Three layered superposition coding with a rate splitting of the private message and indirect coding of public message at receiver-3 are used in the scheme. Let N be the block length. Let $R_1 = R_{11} + R_{12}$ be the rate split of the private message. We first generate 2^{NR_0}-w^N sequences, whose components are i.i.d. according to the distribution $p(w)$, independently for the public message. Then we use superposition coding to generate $2^{NR_{11}}$-v^N sequences, whose components are independent according to conditional distribution $p(v|w)$ given each w^N sequence, independently for the part of private message. We again use superposition coding to generate

$2^{NR_{12}}$-x^N sequences, whose components are independent according to conditional distribution $p(x|v)$ given each v^N sequence, independently for the other part of private message. For each public message and private message pair, their corresponding x^N sequence gets transmitted as a codeword. Receiver-1 recovers the unique public message and private message pair whose (w^N, v^N, x^N) is jointly typical with received sequence at the receiver. Receiver-2 recovers the unique public message whose w^N is jointly typical with received sequence at the receiver. Instead of recovering public message like how receiver-2 does, receiver-3 recovers the unique public message whose w^N sequence and at-least one of its v^N sequence in second layer is jointly typical with received sequence at the receiver, which is referred to as indirect decoding method. If R_0, R_1, R_{11}, R_{12} satisfy the following:

$$R_0 < I(W;Y_2) \tag{5.6}$$

$$R_{12} < I(X;Y_1|V) \tag{5.7}$$

$$R_{11} + R_{12} < I(X;Y_1|W) \tag{5.8}$$

$$R_0 + R_{11} + R_{12} < I(X;Y_1) \tag{5.9}$$

$$R_0 + R_{11} < I(V;Y_3), \tag{5.10}$$

then reliable recovery of the intended messages at each of the receivers is ensured. After eliminating variables R_{11} and R_{12} by Fourier-Motzkin procedure [14] by substituting $R_1 = R_{11} + R_{12}$, we get the region described by equations (5.3), (5.4) and (5.5) that defines the capacity region.

The intuition behind the rate splitting is that if we want to achieve a private message rate satisfying $R_1 > I(X;Y_1|V)$ and $R_1 < I(X;Y_1|W)$, then we rate split R_1 into R_{11} and R_{12} such that $R_{12} < I(X;Y_1|V)$. As we recover public message indirectly using v^N sequences at receiver-3, the sum of public message rate R_0 and R_{11} should be less than $I(V;Y_3)$. So, if we make R_{11} small while rate splitting, then it can be noticed that the public message rate can be improved, provided

the reliability constraint at receiver-2, $R_0 < I(W;Y_2)$, is loose.

5.3.2 Rate splitting of the private message for polar coding

Notice that a point in the region satisfied by equations (5.6), (5.7), (5.8), (5.9) and (5.10) does not always satisfy the constraint $R_{11} < I(V;Y_1|W)$. We impose the new additional constraint $R_{11} < I(V;Y_1|W)$ for the rate split in the implementation of our polar coding strategy through following lemma.

Lemma 10. *For any rate pair* (R_0,R_1) *that satisfies equations* (5.3) (5.4) *and* (5.5) *and for a particular joint distribution* $p(w,v)p(x|v)$ *on* (W,V,X)*, there exist rates* R_{11} *and* R_{12} *such that* $R_1 = R_{11} + R_{12}$ *(rate split of* R_1*) and following three identities hold.*

$$R_{11} < I(V;Y_1|W)$$
$$R_{12} < I(X;Y_1|V)$$
$$R_0 + R_{11} < I(V;Y_3)$$

Proof:

It is easy to find the split for R_1 such that the first two identities hold since $I(V;Y_1|W) + I(X;Y_1|V) = I(X;Y_1|W)$ ($W \to V \to X \to Y_1$ is chain). Let R'_{11} and R'_{12} be such a rate split for R_1. Suppose that the third identity does not hold for the split $R_1 = R'_{11} + R'_{12}$. That means $R_0 + R'_{11} \geq I(V;Y_3)$. Say that $R_0 + R'_{11} = I(V;Y_3) + \delta$ for some $\delta \geq 0$. On the other hand we have $R_0 + R'_{11} + R'_{12} < I(V;Y_3) + I(X;Y_1|V)$. So we should have $R'_{12} < I(X;Y_1|V) - \delta$. Say that $R'_{12} = I(X;Y_1|V) - \delta_1$. Clearly $\delta_1 > \delta$.

Note that $R'_{11} > \delta$, since $R_0 < I(V;Y_3)$. Choose $R_{11} = R'_{11} - \delta^+$ and $R_{12} = R'_{12} + \delta^+$ where $\min\{R'_{11}, \delta_1\} > \delta^+ > \delta$. Clearly, R_{11} and R_{12} is a split of R_1 that satisfies the required three identities. Hence the claim of the lemma is shown. □

In our polar coding strategy, the private message bits for receiver-1 are given in bits I_1^v and I_1^x that are corresponding to V^N vectors and X^N vectors, which are involved in the chaining construction we provide, respectively. The rate split in Lemma 10 allows us to associate the

private message bits encoded in I_1^v and I_1^x to split rates of the private message R_{11} and R_{12}, respectively. We also involve the bits corresponding to R_{11}, which are private message bits encoded in I_1^v, in the chaining procedure to translate the indirect coding method at receiver-3 into polar coding. We also use the degradation condition of receiver-2 in our code construction. Now we provide our code construction in the following subsection.

5.3.3 Code construction

We give a polar coding strategy for each of the following possible cases for the rate pair (R_0, R_1).

- $R_0 \geq I(W;Y_3)$
- $R_0 < I(W;Y_3)$

We consider k polar blocks of size N large enough so that the polarization happens. We propose a chaining construction with these k polar blocks for the rate pair (R_0, R_1) by using the rate split given by the Lemma 10.

While encoding each polar block, we first construct $(U_w)^{1:N}$ and compute $W^{1:N} = (U_w)^{1:N} G_N$. We next construct $V^{1:N} = (U_v)^{1:N} G_N$ given $W^{1:N}$ and apply polar transform to obtain $V^{1:N}$. Lastly, we construct $(U_x)^{1:N}$ given $V^{1:N}$ and apply polar transform to obtain $X^{1:N}$ (codeword). This encoding method ensures that the average distribution of $(W_i, V_i, X_i)_{i=1}^N$ is close in total variation distance to the distribution which is induced when $(W_i, V_i, X_i)_{i=1}^N$ is i.i.d. according to $p(w)p(v|w)p(x|v)$. The total variation distance becomes $O(2^{-N^{\beta'}})$ where $\beta' < \beta < 0.5$.

We first give the construction for the case where $R_0 \geq I(W;Y_3)$. This is the case where we translate the indirect coding into polar coding strategy. We assume $NR_{11} > |I_1^v \cap I_3^v|$ to demonstrate the code construction. The construction we give under this assumption gives the general idea of the chaining construction which can easily be extended to the case where this assumption does not hold.

Note that public message bits have to be recovered at all the receivers. If we give NR_0 public message bits in I_2^w, receiver-2 and receiver-1 (due to degradation condition) can recover these bits. But receiver-3 may not be able to decode in that case. On the other hand we can recover these bits at receiver-3, if we place these bits into I_3^w and remaining $NR_0 - |I_3^w|$ bits in I_3^v, as $NR_0 > |I_3^w|$. In this case, In this case, receiver-1 and receiver-2 may not be able to decode. We do a chaining, to resolve the alignment of the bit-channel set in I_2^w with bit-channels sets in I_3^w and I_3^v to allocate the public message bits for reliable recovery at all the receivers.

Since we are assigning a portion of public message bits in $(U_v)^{1:N}$ vectors for receiver-3, we need to recover $(U_v)^{1:N}$ vectors at receiver-3. But we also use $(U_v)^{1:N}$ vectors for encoding private message bits corresponding to the rate R_{11}. If we give these private message bits in I_1^v, receiver-3 cannot recover these bits, which blocks receiver-3 from recovering $(U_v)^{1:N}$ vectors for decoding the portion of intended public message bits. Here is where we need to do a second level of chaining for aligning bit-channel set in I_3^v with bit-channel set in I_1^v where we provide private message bits corresponding to R_{11}. This summarizes the main idea behind the construction that translates indirect coding at receiver-3.

Fig. 5.3 shows how we fill $(U_w)^{1:N}$, $(U_v)^{1:N}$ and $(U_x)^{1:N}$ vectors when $k = 3$ allocating public and private message bits. The links between vectors in Fig. 5.3 indicate the copying of bits between bit-channel sets of successive blocks. Now we provide detailed steps in encoding and decoding methods in the two-level chaining construction for this case, $R_0 \geq I(W;Y_3)$.

Encoding:

- Encoding $(k-1)NR_0 + |I_3^w \cap I_2^w|$ bits of the public message, first level of chaining:
  - We first place $|I_3^w \cap I_2^w|$ bits in $(U_w)^{I_3^w \cap I_2^w}$ for all the blocks $t = 1 : k$. Note that NR_0 is the sum of $|I_3^w \cap I_2^w| + |I_3^w \cap F_2^w| + (NR_0 - |I_3^w|)$.
  - We place $|I_3^w \cap F_2^w|$ bits in $(U_w)^{I_3^w \cap F_2^w}$ and $NR_0 - |I_3^w|$ in $(U_v)^{I_{31}^v}$ for the blocks $t = 1 : k-1$ where I_3^v is partitioned as disjoint union $I_{31}^v \cup I_{32}^v$, $|I_{31}^v| = NR_0 - |I_3^w|$. Note that $NR_0 + NR_{11} < |I_3^w| + |F_1^v \cap I_3^v| + |I_1^v \cap I_3^v|$ due to Lemma 10. As we assumed the case

where $NR_{11} > |I_1^v \cap I_3^v|$, we can select I_{31} such that $I_{31} \subset I_3^v \cap F_1^v$.

- We copy bits in $(U_w)^{I_3^w \cap F_2^w}$ and $(U_v)^{I_{31}^v}$ of block t to $(U_w)^{B_{w1}}$ of block $t+1$ for $t = 1 : k-1$ where $I_2^w \cap F_3^w$ is partitioned as disjoint union $B_{w1} \cup B_{w2}$ and $|B_{w1}| = NR_0 - |I_3^w \cap I_2^w|$.

- Encoding $(k-1)NR_{11} + |I_1^v \cap I_3^v|$ bits of the private message for receiver 1, second level of chaining:

  - We first place $|I_3^v \cap I_2^v|$ private message bits in $(U_v)^{I_3^v \cap I_2^v}$ for all the blocks $t = 1 : k$.
  - We place $NR_{11} - |I_3^v \cap I_2^v|$ bits in $(U_v)^{I_{321}^v}$ for the blocks $t = 1 : k-1$ where I_{32}^v is partitioned as disjoint union $I_{321}^v \cup I_{322}^v \cup (I_1^v \cap I_3^v)$, and $|I_{321}^v| = NR_{11} - |I_1^v \cap I_3^v|$. Note that $NR_{11} < \min\{|I_{32}^v|, |I_1^v|\}$ due to Lemma 10.
  - We copy the bits in $(U_w)^{I_{321}^v}$ of block t to $(U_v)^{I_{11}^v}$ of block $t+1$ for $t = 1 : k-1$, where $(I_1^v \cap F_3^v)$ is partitioned as the disjoint union $I_{11}^v \cup I_{12}^v$ and $|I_{11}^v| = |I_{321}^v|$.

- Encoding kNR_{12} bits of the private message for receiver-1: We place NR_{12} bits in $(U_x)^{I_1^x}$ for all these blocks $t = 1 : k$. Note that $NR_{12} < |I_1^x|$ due to Lemma 10. We do not involve this portion of the private message bits in the chaining.

- We place randomly chosen frozen bits with i.i.d. uniform distribution in $(U_w)^{B_{w1}}$, $(U_v)^{I_{11}^v}$ for the block $t = 1$. We place randomly chosen frozen bits with i.i.d. uniform distribution in $(U_w)^{(I_3^w \cap F_2^w)}$, $(U_v)^{I_{31}^v}$, $(U_v)^{I_{321}^v}$ for the block $t = k$. We place randomly chosen bits with i.i.d. uniform distribution in the remaining positions of $(U_w)^{\mathscr{H}_W}$, $(U_v)^{\mathscr{H}_{V|W}}$ and $(U_x)^{\mathscr{H}_{X|V}}$, which are not filled by private or public message bits, in all the k blocks. We share these remaining bits that are in $(U_w)^{F_j}$, $(U_v)^{F_j}$ and $(U_x)^{F_j}$ of each block with the receiver-j for $j = 1, 2, 3$, in all the k blocks.

- We have constructed $(U_w)^{\mathscr{H}_W}$, $(U_v)^{\mathscr{H}_{V|W}}, (U_x)^{\mathscr{H}_{X|V}}$ for all the k blocks. Now we encode other positions in $(U_w)^{1:N}$, $(U_v)^{1:N}, (U_x)^{1:N}$ as we do for single asymmetric channel

case [24], [37] for all the blocks $t = 1 : k$.

- We use the following decision rule for encoding $(U_w)^{\mathscr{L}_W}$.

$$(U_w)_i = \text{argmax}_{x\in\{0,1\}} P_{(U_w)_i|(U_w)^{1:i-1}}(x|(U_w)^{1:i-1}).$$

For $i \in \mathscr{L}_W$, the induced conditional distribution $\delta_i^w((u_w)_i|(u_w)^{1:i-1})$ on $(U_w)_i$ given $(U_w)^{1:i-1}$ satisfies $\delta_i^w((u_w)_i|(u_w)^{1:i-1}) = 1$ and $\delta_i((u_w)_i + 1|(u_w)^{1:i-1}) = 0$ where

$$(u_w)_i = \text{argmax}_{x\in\{0,1\}} P_{(U_w)_i|(U_w)^{1:i-1}}(x|(u_w)^{1:i-1}).$$

- We use either randomly chosen boolean functions, which are shared with all the receivers, or common randomness [24], [37] for encoding bit-channels in $(U_w)^{R_w}$ to maintain the conditional distribution $P_{(U_w)_i|(U_w)^{1:i-1}}$ on an average over the random ensemble. We now compute $W^{1:N} = (U_w)^{1:N} G_N$.

- We use the decision rule below for encoding $(U_v)^{\mathscr{L}_{V|W}}$.

$$(U_v)_i = \text{argmax}_{x\in\{0,1\}} P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}}(x|W^{1:N}(U_v)^{1:i-1}).$$

For $i \in \mathscr{L}_{V|W}$, the induced conditional distribution $\delta_i^v((u_v)_i|w^{1:N}(u_v)^{1:i-1})$ on $(U_v)_i$ given $W^{1:N}(U_v)^{1:i-1}$ satisfies $\delta_i^v((u_v)_i|w^{1:N}(u_v)^{1:i-1}) = 1$ and $\delta_i^v((u_v)_i + 1|w^{1:N}(u_v)^{1:i-1}) = 0$ where

$$(u_v)_i = \text{argmax}_{x\in\{0,1\}} P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}}(x|w^{1:N}(u_v)^{1:i-1}).$$

- We use either randomly choosen boolean functions, which are shared with all the receivers, or common randomness for encoding bit-channels in $(U_v)^{R_v}$ to maintain the conditional distribution $P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}}$. Now we compute $V^{1:N} = (U_v)^{1:N} G_N$.

- We use the decision rule below for encoding $(U_x)^{\mathscr{L}_{X|V}}$.

$$(U_x)_i = \text{argmax}_{x\in\{0,1\}} P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}}(x|V^{1:N}(U_x)^{1:i-1}).$$

  For $i \in \mathscr{L}_{X|V}$, the induced conditional distribution $\delta_i^x((u_x)_i|v^{1:N}(u_x)^{1:i-1})$ on $(U_x)_i$ given $V^{1:N}(U_x)^{1:i-1}$ satisfies $\delta_i^x((u_x)_i|v^{1:N}(u_x)^{1:i-1}) = 1$ and $\delta_i^x((u_x)_i+1|v^{1:N}(u_x)^{1:i-1}) = 0$ where

$$(u_x)_i = \text{argmax}_{x\in\{0,1\}} P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}}(x|v^{1:N}(u_x)^{1:i-1}).$$

- We use either randomly chosen boolean functions, which are shared with all the receivers, or common randomness for encoding bit-channels in $(U_x)^{R_x}$ to maintain the conditional distribution $P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}}$. Now we compute $X^{1:N} = (U_x)^{1:N}G_N$.
- We transmit $X^{1:N}$ for all k blocks.

Rate of scheme: We encoded $(k-1)\cdot NR_0 + |I_2^w \cap I_3^w|$ public message bits for k blocks. We encoded $(k-1)\cdot NR_{11} + k\cdot NR_{12} + |I_1^v \cap I_3^v|$ private message bits for k blocks. Hence the we achieve the rate pair $(\frac{(k-1)\cdot NR_0+|I_2^w\cap I_3^w|}{k\cdot N}, \frac{(k-1)\cdot NR_{11}+k\cdot NR_{12}+|I_1^v\cap I_3^v|}{k\cdot N})$, which approaches the pair (R_0, R_1) as k goes infinity. Now we provide the decoding method for the case $R_0 \geq I(W;Y_3)$.

Decoding, using $Y_j^{1:N}$ at receiver-j for all k blocks:

- The following steps $1)-4)$ give the decoding procedure at receiver-3. We decode both $(U_w)^{1:N}$s and $(U_v)^{1:N}$s of all the blocks to recover the public message bits at receiver-3.

  1. Set $t = 1$. We decode $(U_w)^{1:N}$ and $(U_v)^{1:N}$ by successive cancellation for the block t. NR_0 bits $(U_w)^{I_3^w}$ and $(U_v)^{I_{31}^v}$ of the public message will be recovered in this step. NR_{11} bits in $(U_v)^{I_{32}^v}$ of the private message of receiver-1 will also be recovered.
  2. We decode $(U_w)^{1:N}$ followed by $(U_v)^{1:N}$ by successive cancellation for the block $t+1$. The bits $(U_w)^{I_3^w\cap F_2^v}$, $(U_v)^{I_{31}^v}$ and $(U_v)^{I_{321}^v}$ recovered for block t give bits in $(U_w)^{B_{w1}}$ and $(U_v)^{I_{11}^v}$ during the successive cancellation decoding of block $t+1$. NR_0 bits $(U_w)^{I_3^w}$

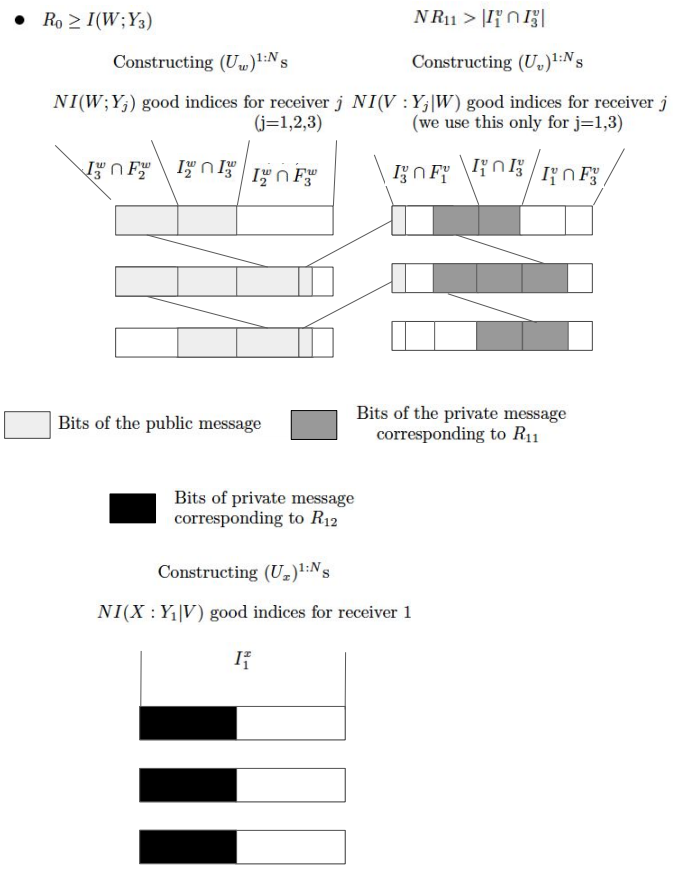

Figure 5.3. Private and public message bits allocation in $(U_w)^{1:N}$, $(U_v)^{1:N}$ and $(U_x)^{1:N}$ vectors when $k = 3$

and $(U_v)^{I_{31}^v}$ of the public message will be recovered in this step. NR_{11} bits in $(U_v)^{I_{32}^v}$ of the private message of receiver-1 will also be recovered. Increase t by 1.

3. Repeat step (2) until $t = k-1$.

4. We decode $(U_w)^{1:N}$ and $(U_v)^{1:N}$ in successive cancellation style for block k. The bits $(U_w)^{I_3^w \cap F_2^v}$, $(U_v)^{I_{31}^v}$ and $(U_v)^{I_{321}^v}$ and recovered for block $k-1$ give bits in $(U_w)^{B_{w1}}$ and $(U_v)^{I_{11}^v}$ during the successive cancellation decoding of block k. The bits $(U_w)^{I_3^w \cap I_2^w}$ of the public message will be recovered in this step. The bits $(U_v)^{I_3^v \cap I_2^v}$ of the private message of receiver-1 will also be recovered in this step.

- The following steps (1)-(4) give the decoding procedure at receiver-1. The content in

parentheses-$\big(\big)$ is ignored when decoding at receiver-2. We decode all $(U_w)^{1:N}$s, $(U_v)^{1:N}$s and $(U_x)^{1:N}$s of all the blocks to recover the public message bits and private message bits at receiver-1. We only decode $(U_w)^{1:N}$s of all the blocks to recover public message bits at receiver-2.

1. Set $t = k$. We decode $(U_w)^{1:N}$ $\big($,$(U_v)^{1:N}$ and $(U_x)^{1:N}\big)$ by successive cancellation for block t. NR_0 bits $(U_w)^{I_3^w \cap I_2^w}$ and $(U_w)^{B_{w1}}$ of the public message will be recovered for block t. $\big(NR_{11}$ bits in $(U_v)^{I_{11}^v \cup (I_1^v \cap I_3^v)}$ of the private message of receiver-1 will also be recovered. NR_{12} bits in $(U_x)^{I_1^x}$ of the private message of receiver-1 will also be recovered.$\big)$
2. We decode $(U_w)^{1:N}\big($,$(U_v)^{1:N}$ and $(U_x)^{1:N}\big)$ by successive cancellation for block $t-1$. The bits $(U_w)^{B_{w1}}$, $\big((U_v)^{I_{11}^v}\big)$ recovered for block t give bits in $(U_w)^{I_3^w \cap F_2^w}$ $\big($, $(U_v)^{I_{31}^v}$ and $(U_v)^{I_{321}^v}\big)$ during the successive cancellation decoding of block $t-1$. NR_0 bits $(U_w)^{I_3^w \cap I_2^w}$ and $(U_w)^{B_{w1}}$ of the public message will be recovered.$\big(NR_{11}$ bits in $(U_v)^{I_{11}^v \cup (I_1^v \cap I_3^v)}$ of the private message of receiver-1 will also be recovered. NR_{12} bits in $(U_x)^{I_1^x}$ of the private message of receiver-1 will also be recovered.$\big)$ Decrease t by 1.
3. Repeat step (2) until $t = 2$.
4. We decode $(U_w)^{1:N}$ $\big($,$(U_v)^{1:N}$ and $(U_x)^{1:N}\big)$ by successive cancellation for block 1. The bits $(U_w)^{B_{w1}}$ $\big($, $(U_v)^{I_{11}^v}\big)$ recovered for block 2 give bits in $(U_w)^{I_3^w \cap F_2^w}$ $\big($, $(U_v)^{I_{31}^v}$ and $(U_v)^{I_{321}^v}\big)$ during the successive cancellation decoding of block 1. The bits $(U_w)^{I_3^w \cap I_2^w}$ of the public message will be recovered. $\big($The bits $(U_v)^{I_1^v \cap I_3^v}$ of the private message of receiver-1 will also be recovered. NR_{12} bits in $(U_x)^{I_1^x}$ of the private message of receiver-1 will also be recovered.$\big)$

- During the successive cancellation decoding, we recover the needed bits in $(U_w)^{\mathscr{L}_W}$, $(U_x)^{\mathscr{L}_{V|W}}$ and $(U_v)^{\mathscr{L}_{X|V}}$ at each receiver by an appropriate decision/arg-max rule.

- We use the following decision rule for decoding $(U_w)^{\mathscr{L}_W}$ and $(U_w)^{I_j^w}$ at receiver-$j = 1,2,3$.

$$(U_w)_i = \text{argmax}_{x\in\{0,1\}} P_{(U_w)_i|(U_w)^{1:i-1}Y_j^{1:N}}(x|(U_w)^{1:i-1}Y_j^{1:N}).$$

We use the following decision rule for decoding $(U_v)^{\mathscr{L}_{V|W}}$ and $(U_w)^{I_j^v}$ at reciever-$j = 1,3$.

$$(U_v)_i = \text{argmax}_{x\in\{0,1\}} P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}Y_j^{1:N}}(x|W^{1:N}(U_v)^{1:i-1}Y_j^{1:N}).$$

We use the following decision rule below for decoding $(U_x)^{\mathscr{L}_{X|V}}$ and $(U_w)^{I_1^x}$ at reciever-1.

$$(U_x)_i = \text{argmax}_{x\in\{0,1\}} P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}Y_1^{1:N}}(x|V^{1:N}(U_x)^{1:i-1}Y_1^{1:N}).$$

- The remaining bits could be either the bits in frozen positions which are available at the corresponding receiver or the bits in $(U_w)^{R_w}$, $(U_v)^{R_v}$ and $(U_x)^{R_x}$ for which we use shared boolean functions/common randomness to decode.

We assumed that $NR_{11} > |I_1^v \cap I_3^v|$. Suppose if that does not hold, then we do not have to perform chaining at the second level. The private message bits corresponding to the rate R_{11} will fit into $I_1^v \cap I_3^v$ and hence can be recovered by receiver-3 and receiver-1. Allocation of the private message bits in $(U_x)^{1:N}$ corresponding to the rate R_{12} will still be the same as in construction for the previously assumed condition. Fig. 5.4 shows the allocation of private and public message bits in $(U_w)^{1:N}$ and $(U_v)^{1:N}$ in the chaining procedure for $k = 3$ when $NR_{11} \leq |I_1^v \cap I_3^v|$. The other details of the construction can easily be extended from the construction under the assumption $NR_{11} > |I_1^v \cap I_3^v|$. Fig. 5.4 shows a case where the public message bits in I_3^v fit into $I_3^v \cap F_1^v$. Notice that the same chaining procedure still applies, as shown in Fig. 5.4 even when these public message bits overflow into $I_3^v \cap I_1^v$.

Now we look at the other case where $R_0 < I(W;Y_3)$. Assume $NR_0 > |I_2^w \cap I_3^w|$ where there will be non-trivial chaining construction. In this case, note that NR_0 public message bits

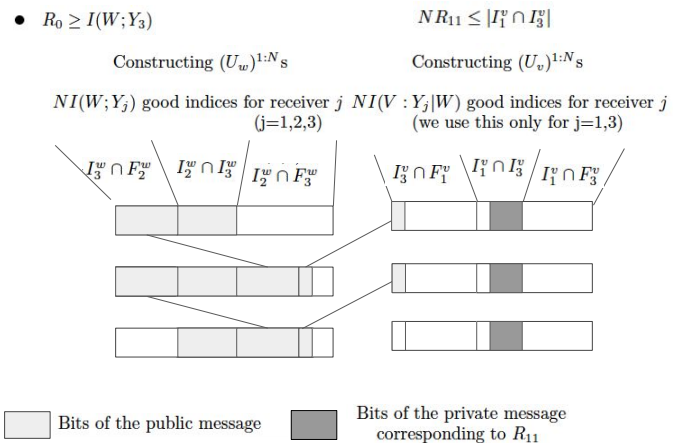

Figure 5.4. Private and public message bits allocation in $(U_w)^{1:N}$ and $(U_v)^{1:N}$ vectors when $k = 3$

totally fit into $|I_3^w|$. We perform chaining within the layer $(U_w)^{1:N}$ itself and resolve the alignment of bit-channel sets I_2^w and I_3^w so that these public message bits can be reliably decoded at all the receivers. Since we do not require to fill the public message bits in I_3^v, receiver-3 can ignore decoding the $(U_v)^{1:N}$ vectors. Hence there will be no need of chaining at the second level that aligns bit-channel sets in I_1^v and I_3^v for private message bits corresponding to the rate R_{11}. It is just enough to provide private message bits in I_1^v. The other details of the code construction can easily be extended from earlier case. Fig. 5.5 shows the case under the assumption $NR_0 > |I_2^w \cap I_3^w|$ when $k = 3$.

Suppose if $NR_0 \leq |I_2^w \cap I_3^w|$, we can fill NR_0 public message bits in $I_2^w \cap I_3^w$ so that they can be recovered at all the receivers. We can the fill NR_{11} and NR_{12} private message bits in I_1^v of $(U_v)^{1:N}$ and I_1^x of $(U_x)^{1:N}$ so that they can be reliably decoded at receiver-1. Hence chaining is not needed when $NR_0 \leq |I_2^w \cap I_3^w|$.

5.3.4 Probability of error analysis

Let $\mathbb{C}$ denotes the random vector which contains randomly chosen frozen bits in the code construction. The random variable $\mathbb{C}$ also contains randomly chosen boolean functions for not-completely polarized bit-channels in case we do not employ common randomness in the code

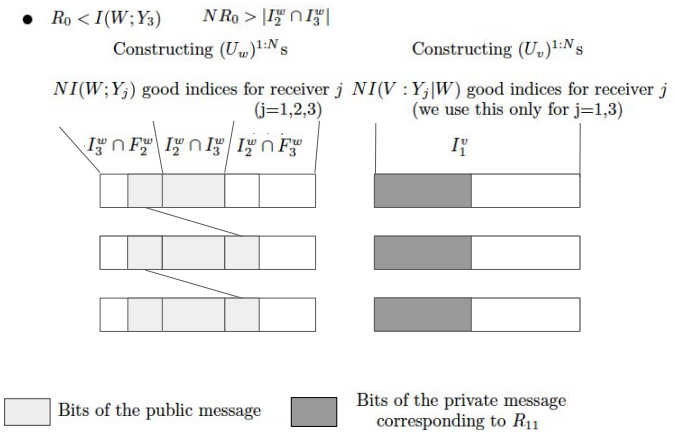

Figure 5.5. Private and public message bits allocation in $(U_w)^{1:N}$ and $(U_v)^{1:N}$ vectors when $k = 3$

construction, of all the blocks as its components. We give the analysis for the code construction that uses common randomness for not-completely polarized bit-channels. The spirit of the analysis will remain the same for the case where we use randomly chosen boolean functions for encoding not completely polarized bit-channels as in [24]. The analysis of probability of error that we provide is done in three steps. First step is deriving the average distribution of each block which is close to the distribution induced when $(W^{1:N}, V^{1:N}, X^{1:N})$ is i.i.d. according to $p(w)p(v|w)p(x|v)$ in total variation distance through Lemma 11. Secondly, we write error event at each receiver as a union of error events we define for each of these blocks. Notice that blocks involved in the chaining are statistically dependent due to the chaining construction we did. We use linearity of expectation and union bound to get an upper bound on average probability of error at a receiver, which is sum of average probability of errors of each of these blocks at that receiver. Finally, we use the fact that the total variation distance between average distribution of the block in the code construction and distribution when $(W^{1:N}, V^{1:N}, X^{1:N})$ is i.i.d. according to $p(w)p(v|w)p(x|v)$ are close and polarization results to get bound on the average probability of each block at that receiver. Theorem 6 provides a detailed analysis of the probability of decoding error for the chaining construction. We now give Lemma 11 used in proof of Theorem 6.

Lemma 11. *Let $Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}$ be the measure on $(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}$ as follows:*

$$
\begin{aligned}
&Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}(u_w^{1:N}u_v^{1:N}u_x^{1:N}) \\
&= \big(2^{-|\mathscr{H}_W|}\Pi_{i\in\mathscr{L}_W}\delta_i^w((u_w)_i|(u_w)^{1:i-1})\Pi_{i\in R^w}P_{(U_w)_i|(U_w)^{1:i-1}}((u_w)_i|(u_w)^{1:i-1})\big)\cdot \\
&\quad \big(2^{-|\mathscr{H}_{V|W}|}\Pi_{i\in\mathscr{L}_{V|W}}\delta_i^v((u_v)_i|w^{1:N}(u_v)^{1:i-1}) \\
&\qquad\qquad \Pi_{i\in R^v}P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}}((u_v)_i|w^{1:N}(u_v)^{1:i-1})\big)\cdot \\
&\quad \big(2^{-|\mathscr{H}_{X|V}|}\Pi_{i\in\mathscr{L}_{X|V}}\delta_i^x((u_x)_i|v^{1:N}(u_x)^{1:i-1}) \\
&\qquad\qquad \Pi_{i\in R^x}P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}}((u_x)_i|v^{1:N}(u_x)^{1:i-1})\big).
\end{aligned}
$$

Let $P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}$ be the measure induced when $(W^{1:N},V^{1:N},X^{1:N})$ are i.i.d. according to $p(w)p(v|w)p(x|v)$. The total variation distance, $||P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}} - Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}|| = O(2^{-N^{\beta'}})$, where $\beta' < \beta < 0.5$, $w^N = (u_w)^{1:N}G_N$, $v^N = (u_v)^{1:N}G_N$ and $x^N = (u_x)^{1:N}G_N$.

Proof:

We use short hand notation $Q((u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:N})$ and $P((u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:N})$ for $Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}((u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:N})$ and $P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}((u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:N})$, respectively. The proof is inspired from Lemma 1 in [24]. From equation (56) in [24], we have the following identity:

$$B_1^n - A_1^n = \sum_{i=1}^{n}(B_i - A_i)A_1^{i-1}B_{i+1}^n \tag{5.11}$$

where A_j^k and B_j^k denotes the product $\prod_{i=j}^k A_i$ and $\prod_{i=j}^k B_i$ respectively. We are going to apply this for $n = 3N$ length vector, which is $(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}$.

$$
\begin{aligned}
&2||Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}} - P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}|| \\
&\qquad = \sum_{(u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:N}} \\
&\qquad\quad |\big(\sum_{i=1}^{N} Q((u_w)_i|(u_w)^{1:i-1}) - P((u_w)_i|(u_w)^{1:i-1})\big) \\
&\qquad\qquad \Pi_{m=1}^{i-1}P((u_w)_m|(u_w)^{1:m-1}) \\
&\qquad\qquad \Pi_{m=i+1}^{N}Q((u_w)_m|(u_w)^{1:m-1})
\end{aligned}
$$

$$
\begin{aligned}
&\Pi_{p=1}^{N} Q((u_v)_p|w^{1:N}(u_v)^{1:p-1}) \\
&\Pi_{q=1}^{N} Q((u_x)_q|w^{1:N}v^{1:N}(u_x)^{1:q-1}) \\
&+\sum_{k=1}^{N}(Q((u_v)_k|w^{1:N}(u_v)^{1:k-1}) - P((u_v)_k|w^{1:N}(u_v)^{1:k-1})) \\
&\quad \Pi_{m=1}^{N} P((u_w)_m|(u_w)^{1:m-1}) \\
&\quad \Pi_{p=1}^{k-1} P((u_v)_p|w^{1:N}(u_v)^{1:p-1}) \\
&\quad \Pi_{p=k+1}^{N} Q((u_v)_p|w^{1:N}(u_v)^{1:p-1}) \\
&\quad \Pi_{q=1}^{N} Q((u_x)_q|w^{1:N}v^{1:N}(u_x)^{1:q-1}) \\
&+\sum_{l=1}^{N}(Q((u_x)_l|w^{1:N}v^{1:N}(u_x)^{1:l-1}) - \\
&\qquad\qquad P((u_x)_k|w^{1:N}v^{1:N}(u_x)^{1:k-1})) \\
&\quad \Pi_{m=1}^{N} P((u_w)_m|(u_w)^{1:m-1}) \\
&\quad \Pi_{p=1}^{N} P((u_v)_p|w^{1:N}(u_v)^{1:p-1}) \\
&\quad \Pi_{q=1}^{l-1} P((u_x)_q|w^{1:N}v^{1:N}(u_x)^{1:q-1})) \\
&\quad \Pi_{q=l+1}^{N} Q((u_x)_q|w^{1:N}v^{1:N}(u_x)^{1:q-1}))|
\end{aligned}
$$

This implies that

$$
\begin{aligned}
&2||Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}} - P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}|| \\
&\qquad \leq \sum_{(u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:N}} \\
&\qquad \Big(\sum_{i=1}^{N} |Q((u_w)_i|(u_w)^{1:i-1}) - P((u_w)_i|(u_w)^{1:i-1}))| \\
&\qquad\quad P((u_w)^{1:i-1})Q((u_w)^{i+1:N}(u_v)^{1:N}(u_x)^{1:N}|(u_w)^{1:i}) \\
&\qquad + \sum_{k=1}^{N} |(Q((u_v)_k|w^{1:N}(u_v)^{1:k-1}) - \\
&\qquad\qquad\qquad P((u_v)_k|w^{1:N}(u_v)^{1:k-1}))|
\end{aligned}
$$

$$
\begin{aligned}
&P((u_w)^{1:N}(u_v)^{1:k-1}) \\
&\quad Q((u_v)^{k+1:N}(u_x)^{1:N}|(w)^{1:N}(u_v)^{1:k}) \\
&+\sum_{l=1}^{N}|(Q((u_x)_l|v^{1:N}(u_x)^{1:l-1})- \\
&\qquad\qquad P((u_x)_k|v^{1:N}(u_x)^{1:l-1}))| \\
&\quad P((u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:l-1}) \\
&\quad Q((u_x)^{l+1:N}|(w)^{1:N}(v)^{1:N}(u_x)^{1:l})\big)
\end{aligned}
$$

This implies that

$$
\begin{aligned}
&2||Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}-P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}|| \\
&\qquad\leq\big(\sum_{(u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:N}} \\
&\quad\sum_{i=1}^{N}|P((u_w)_i|(u_w)^{1:i-1})-Q((u_w)_i|(u_w)^{1:i-1}))| \\
&\qquad P((u_w)^{1:i-1})Q((u_w)^{i+1:N}(u_v)^{1:N}(u_x)^{1:N}|(u_w)^{1:i})\big) \\
&\quad+\big(\sum_{(u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:N}} \\
&\quad\sum_{k=1}^{N}|(P((u_v)_k|w^{1:N}(u_v)^{1:k-1})-Q((u_v)_k|w^{1:N}(u_v)^{1:k-1}))| \\
&\quad P((u_w)^{1:N}(u_v)^{1:k-1})Q((u_v)^{k+1:N}(u_x)^{1:N}|(w)^{1:N}(u_v)^{1:k})\big) \\
&\quad+\big(\sum_{(u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:N}} \\
&\quad\sum_{l=1}^{N}|(P((u_x)_l|v^{1:N}(u_x)^{1:l-1})-Q((u_x)_k|v^{1:N}(u_x)^{1:k-1}))| \\
&\quad P((u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:l-1})Q((u_x)^{l+1:N}|(w)^{1:N}(v)^{1:N}(u_x)^{l})\big)
\end{aligned}
$$

This implies that

$$
\begin{aligned}
2||Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}} - P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}|| & \\
&\leq \Big(\sum_{i=1}^{N} \sum_{(u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:N}} \\
&\quad |P((u_w)_i|(u_w)^{1:i-1}) - Q((u_w)_i|(u_w)^{1:i-1}))| \\
&\quad\quad P((u_w)^{1:i-1}) Q((u_w)^{i+1:N}(u_v)^{1:N}(u_x)^{1:N}|(u_w)^{1:i}) \Big) \\
&\quad + \Big(\sum_{k=1}^{N} \sum_{(u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:N}} \\
&\quad |(P((u_v)_k|w^{1:N}(u_v)^{1:k-1}) - Q((u_v)_k|w^{1:N}(u_v)^{1:k-1}))| \\
&\quad P((u_w)^{1:N}(u_v)^{1:k-1}) Q((u_v)^{k+1:N}(u_x)^{1:N}|(w)^{1:N}(u_v)^{1:k}) \Big) \\
&\quad + \Big(\sum_{l=1}^{N} \sum_{(u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:N}} \\
&\quad |(P((u_x)_l|v^{1:N}(u_x)^{1:l-1}) - Q((u_x)_k|v^{1:N}(u_x)^{1:k-1}))| \\
&\quad P((u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:l-1}) Q((u_x)^{l+1:N}|(w)^{1:N}(v)^{1:N}(u_x)^{l}) \Big) \\
&\leq \Big(\sum_{i=1}^{N} \sum_{(u_w)^{1:i}} P((u_w)^{1:i-1}) \\
&\quad\quad |P((u_w)_i|(u_w)^{1:i-1}) - Q((u_w)_i|(u_w)^{1:i-1}))| \Big) \\
&\quad + \Big(\sum_{k=1}^{N} \sum_{(u_w)^{1:N}(u_v)^{1:k}} P((u_w)^{1:N}(u_v)^{1:k-1}) \\
&\quad\quad |(P((u_v)_k|w^{1:N}(u_v)^{1:k-1}) - Q((u_v)_k|w^{1:N}(u_v)^{1:k-1}))| \Big) \\
&\quad + \Big(\sum_{l=1}^{N} \sum_{(u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:l}} P((u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:l-1}) \\
&\quad\quad |(P((u_x)_l|v^{1:N}(u_x)^{1:l-1}) - Q((u_x)_k|v^{1:N}(u_x)^{1:k-1}))| \Big)
\end{aligned}
\tag{5.12}
$$

Now we consider the individual sum terms in the above bound. Let us first bound the term, $\sum_{i=1}^{N} \sum_{(u_w)^{1:i}} P((u_w)^{1:i-1}) |P((u_w)_i|(u_w)^{1:i-1}) - Q((u_w)_i|(u_w)^{1:i-1}))|$.

If $i \in \mathscr{H}_W$, then

$$\begin{aligned}
&\sum_{(u_w)^{1:i}} P((u_w)^{1:i-1})|P((u_w)_i|(u_w)^{1:i-1}) - Q((u_w)_i|(u_w)^{1:i-1}))| \\
&\quad = \sum_{(u_w)^{1:i-1}} 2P((u_w)^{1:i-1})||P_{(U_w)_i|(U_w)^{1:i-1}=(u_w)^{1:i-1}} - Q_{(U_w)_i|(U_w)^{1:i-1}=(u_w)^{1:i-1}}|| \\
&\quad \overset{(a)}{\leq} \sum_{(u_w)^{1:i-1}} P((u_w)^{1:i-1})\sqrt{(2\ln 2)} \\
&\qquad\qquad \Big(D(P_{(U_w)_i|(U_w)^{1:i-1}=(u_w)^{1:i-1}}||Q_{(U_w)_i|(U_w)^{1:i-1}=(u_w)^{1:i-1}})\Big)^{0.5} \\
&\quad \overset{(b)}{\leq} \sqrt{(2\ln 2)}\Big(\sum_{(u_w)^{1:i-1}} P((u_w)^{1:i-1}) \\
&\qquad\qquad D(P_{(U_w)_i|(U_w)^{1:i-1}=(u_w)^{1:i-1}}||Q_{(U_w)_i|(U_w)^{1:i-1}=(u_w)^{1:i-1}})\Big)^{0.5} \\
&\quad \overset{(c)}{\leq} \sqrt{(2\ln 2)(1 - H((U_w)_i|(U_w)^{1:i-1}))} \\
&\quad \overset{(d)}{\leq} \sqrt{(2\ln 2)(1 - (Z((U_w)_i|(U_w)^{1:i-1}))^2)} \\
&\quad \overset{(e)}{\leq} \sqrt{(4\ln 2)(2^{-n^\beta})} \\
&\quad = O(2^{-n^{\beta'}})
\end{aligned}$$

where $\beta' < \beta$.

(a) follows by pinsker inequality, (b) follows by jensen's inequality, (c) follows due to the fact that $Q((u_w)_i|(u_w)^{1:i-1}) = 0.5$ and by the formula of conditional entropy, (d) follows from equation (5.1), and (e) follows from polarization results mentioned in Section 5.2.

If $i \in \mathscr{L}_W$, let

$$p_{(u_w)^{1:i-1}} = \max\{P(0|(u_w)^{1:i-1}), P(1|(u_w)^{1:i-1})\}$$

Then,

$$\sum_{(u_w)^{1:i}} P((u_w)^{1:i-1})|P((u_w)_i|(u_w)^{1:i-1}) - Q((u_w)_i|(u_w)^{1:i-1}))|$$

$$
\begin{aligned}
&= \sum_{(u_w)^{1:i-1}} 2P((u_w)^{1:i-1})||P_{(U_w)_i|(U_w)^{1:i-1}=(u_w)^{1:i-1}} - Q_{(U_w)_i|(U_w)^{1:i-1}=(u_w)^{1:i-1}}|| \\
&\overset{(a)}{\leq} \sum_{(u_w)^{1:i-1}} P((u_w)^{1:i-1})\sqrt{(2\ln 2)} \\
&\qquad \Big(D(Q_{(U_w)_i|(U_w)^{1:i-1}=(u_w)^{1:i-1}}||P_{(U_w)_i|(U_w)^{1:i-1}=(u_w)^{1:i-1}})\Big)^{0.5} \\
&\overset{(b)}{\leq} \sqrt{(2\ln 2)}\Big(\sum_{(u_w)^{1:i-1}} P((u_w)^{1:i-1}) \\
&\qquad D(Q_{(U_w)_i|(U_w)^{1:i-1}=(u_w)^{1:i-1}}||P_{(U_w)_i|(U_w)^{1:i-1}=(u_w)^{1:i-1}})\Big)^{0.5} \\
&\overset{(c)}{=} \sqrt{(2\ln 2)\sum_{(u_w)^{1:i-1}} P((u_w)^{1:i-1})(-\log(p_{(u_w)^{1:i-1}}))} \\
&\overset{(d)}{\leq} \Big((2\ln 2)\sum_{(u_w)^{1:i-1}} P((u_w)^{1:i-1}) \\
&\qquad (H((U_w)_i|(U_w)^{1:i-1} = (u_w)^{1:i-1}))\Big)^{0.5} \\
&= \sqrt{(2\ln 2)(H((U_w)_i|(U_w)^{1:i-1}))} \\
&\overset{(e)}{\leq} \sqrt{(2\ln 2)(Z((U_w)_i|(U_w)^{1:i-1}))} \\
&\overset{(f)}{\leq} \sqrt{(2\ln 2)2^{-n^\beta}} = O(2^{-n^{\beta'}})
\end{aligned}
$$

(a) follows by pinsker inequality, (b) follows by jensen's inequality for concave functions. (c) follows from $Q((u_w)_i|(u_w)^{1:i-1}) = 1$ when $(u_w)_i = \text{argmax}_{x\in\{0,1\}}\ \{P(x|(u_w)^{1:i-1})\}$. (d) is true since $\log(\frac{p_{(u_w)^{1:i-1}}}{1-p_{(u_w)^{1:i-1}}}) > 0$. (e) follows from equation (5.2), (f) follows from polarization results mentioned in Section 5.2.

Hence

$$
\sum_{i=1}^{N}\sum_{(u_w)^{1:i}} P((u_w)^{1:i-1})|P((u_w)_i|(u_w)^{1:i-1}) - Q((u_w)_i|(u_w)^{1:i-1}))| = O(2^{-N^{\beta'}}). \tag{5.13}
$$

Let us first bound the term,

$\sum_{k=1}^{N}\sum_{(u_w)^{1:N}(u_v)^{1:k}} P((u_v)^{1:k-1})|P((u_v)_k|w^{1:N}(u_v)^{1:k-1}) - Q((u_v)_k|w^{1:N}(u_w)^{1:k-1}))|$.

If $i \in \mathscr{H}_{V|W}$, then

$$\begin{aligned}
&\sum_{(u_w)^{1:N}(u_v)^{1:k}} P((u_w)^{1:N}(u_v)^{1:k-1})|P((u_v)_k|w^{1:N}(u_v)^{1:k-1}) - Q((u_v)_k|w^{1:N}(u_v)^{1:k-1}))| \\
&\qquad = \sum_{w^{1:N}(u_v)^{1:k-1}} 2P(w^{1:N}(u_v)^{1:i-1})||P_{(U_v)_k|W^{1:N}(U_v)^{1:k-1}=w^{1:N}(u_v)^{1:k-1}} - \\
&\qquad\qquad\qquad\qquad Q_{(U_v)_k|W^{1:N}(U_v)^{1:k-1}=w^{1:N}(u_v)^{1:k-1}}|| \\
&\qquad \overset{(a)}{\leq} \sum_{w^{1:N}(u_v)^{1:k-1}} P(w^{1:N}(u_v)^{1:i-1})\sqrt{(2\ln 2)} \\
&\qquad\quad \Big(D(P_{(U_v)_k|W^{1:N}(U_v)^{1:k-1}=w^{1:N}(u_v)^{1:k-1}}||Q_{(U_v)_k|W^{1:N}(U_v)^{1:k-1}=w^{1:N}(u_v)^{1:k-1}}\Big)^{0.5} \\
&\qquad \overset{(b)}{\leq} \sqrt{(2\ln 2)}\Big(\sum_{w^{1:N}(u_v)^{1:k-1}} P(w^{1:N}(u_v)^{1:i-1}) \\
&\qquad\quad \Big(D(P_{(U_v)_k|W^{1:N}(U_v)^{1:k-1}=w^{1:N}(u_v)^{1:k-1}}||Q_{(U_v)_k|W^{1:N}(U_v)^{1:k-1}=w^{1:N}(u_v)^{1:k-1}}\Big)^{0.5} \\
&\qquad \overset{(c)}{\leq} \sqrt{(2\ln 2)(1-H((U_v)_k|W^{1:N}(U_v)^{1:k-1}))} \\
&\qquad \overset{(d)}{\leq} \sqrt{(2\ln 2)(1-(Z((U_v)_k|W^{1:N}(U_v)^{1:k-1}))^2)} \\
&\qquad \overset{(e)}{\leq} \sqrt{(4\ln 2)(2^{-n^\beta})} \\
&\qquad = O(2^{-n^{\beta'}})
\end{aligned}$$

where $\beta' < \beta$.

(a) follows by pinsker inequality, (b) follows by jensen's inequality, (c) follows due to the fact that $Q((u_v)_k|w^{1:N}(u_v)^{1:k-1}) = 0.5$ and by the formula of conditional entropy, (d) follows from equation (5.1) and (e) follows from polarization results mentioned in Section 5.2.

Let $p_{w^{1:N}(u_v)^{1:k-1}} = \max\{P(0|w^{1:N}(u_v)^{1:k-1}), P(1|w^{1:N}(u_v)^{1:k-1})\}$.

If $i \in \mathscr{L}_{V|W}$, then,

$$\sum_{(u_w)^{1:N}(u_v)^{1:k}} P((u_w)^{1:N}(u_v)^{1:k-1})|P((u_v)_k|w^{1:N}(u_v)^{1:k-1}) - Q((u_v)_k|w^{1:N}(u_v)^{1:k-1}))|$$

$$
\begin{aligned}
&= \sum_{w^{1:N}(u_v)^{1:k-1}} 2P(w^{1:N}(u_v)^{1:i-1}) || P_{(U_v)_k|W^{1:N}(U_v)^{1:k-1}=w^{1:N}(u_v)^{1:k-1}} - \\
&\qquad\qquad Q_{(U_v)_k|W^{1:N}(U_v)^{1:k-1}=w^{1:N}(u_v)^{1:k-1}} || \\
&\overset{(a)}{\leq} \sum_{w^{1:N}(u_v)^{1:k-1}} P(w^{1:N}(u_v)^{1:i-1}) \sqrt{(2\ln 2)} \\
&\quad \big(D(Q_{(U_v)_k|W^{1:N}(U_v)^{1:k-1}=w^{1:N}(u_v)^{1:k-1}} || P_{(U_v)_k|W^{1:N}(U_v)^{1:k-1}=w^{1:N}(u_v)^{1:k-1}}\big)^{0.5} \\
&\overset{(b)}{\leq} \sqrt{(2\ln 2)} \big(\sum_{w^{1:N}(u_v)^{1:k-1}} P(w^{1:N}(u_v)^{1:i-1}) \\
&\quad \big(D(Q_{(U_v)_k|W^{1:N}(U_v)^{1:k-1}=w^{1:N}(u_v)^{1:k-1}} || P_{(U_v)_k|W^{1:N}(U_v)^{1:k-1}=w^{1:N}(u_v)^{1:k-1}}\big)^{0.5} \\
&\overset{(c)}{=} \sqrt{(2\ln 2)} \quad \big(\sum_{(u_w)^{1:N}(u_v)^{1:k}} P((u_w)^{1:N}(u_v)^{1:k-1})(-\log(p_{w^{1:N}(u_v)^{1:k-1}}))\big)^{0.5} \\
&\overset{(d)}{\leq} \sqrt{(2\ln 2)} \big(\sum_{(u_w)^{1:N}(u_v)^{1:k}} P((u_w)^{1:N}(u_v)^{1:k-1}) \\
&\qquad\qquad (H((U_v)_i|W^{1:N}(U_v)^{1:i-1} = w^{1:N}(u_v)^{1:i-1}))\big)^{0.5} \\
&= \sqrt{(2\ln 2)(H((U_v)_i|W^{1:N}(U_v)^{1:i-1}))} \\
&\overset{(e)}{\leq} \sqrt{(2\ln 2)(Z((U_v)_i|W^{1:N}(U_v)^{1:i-1}))} \\
&\overset{(f)}{\leq} \sqrt{(2\ln 2)2^{-n^\beta}} = O(2^{-n^{\beta'}})
\end{aligned}
$$

(a) follows by pinsker inequality, (b) follows by jensen's inequality for concave functions. (c) follows from $Q((u_w)_i|(u_w)^{1:i-1}) = 1$ when $(u_v)_k = \text{argmax}_{x\in\{0,1\}} P(x|w^{1:N}(u_v)^{1:k-1})$. (d) is true since $\log(\frac{p_{w^{1:N}(u_v)^{1:k-1}}}{1-p_{w^{1:N}(u_v)^{1:k-1}}}) > 0$. (e) follows from equation (5.2), (f) follows from equation (5.2), (f) follows from polarization results mentioned in Section 5.2. Hence

$$
\begin{aligned}
&\sum_{k=1}^{N} \sum_{(u_w)^{1:N}(u_v)^{1:k}} P((u_v)^{1:k-1}) |P((u_v)_k|w^{1:N}(u_v)^{1:k-1}) - Q((u_v)_k|w^{1:N}(u_w)^{1:k-1}))| \\
&\qquad\qquad = O(2^{-N^{\beta'}})
\end{aligned} \tag{5.14}
$$

By using the same approach as we just used to derive equation (5.14), we will also get

$$\sum_{l=1}^{N} \sum_{(u_w)^{1:N}(u_v)^{1:N}(u_x)^{1:l}} P((u_x)^{1:l-1})|P((u_x)_l|v^{1:N}(u_x)^{1:l-1}) - Q((u_x)_l|v^{1:N}(u_x)^{1:l-1}))| = O(2^{-N^{\beta'}}) \tag{5.15}$$

From equations (5.12), (5.13), (5.14) and (5.15), we get

$$||P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}} - Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}|| = O(2^{-N^{\beta'}}).$$

Hence proof of the lemma. □

Now we provide Theorem 6 that gives a detailed analysis of the probability of decoding error in the chaining construction.

Theorem 6.

1. For every polar block encoded in the chaining construction, we have

$$\begin{aligned}
&\mathbb{E}_{\mathbb{C}}[\mathbb{P}(U_w^{1:N} = u_w^{1:N}, U_v^{1:N} = u_v^{1:N}, U_x^{1:N} = u_x^{1:N}|\mathbb{C})] \\
&= \left(2^{-|\mathscr{H}_W|}\Pi_{i\in\mathscr{L}_W}\delta_i^w((u_w)_i|(u_w)^{1:i-1})\Pi_{i\in R^w}P_{(U_w)_i|(U_w)^{1:i-1}}((u_w)_i|(u_w)^{1:i-1})\right)\cdot \\
&\quad \left(2^{-|\mathscr{H}_{V|W}|}\Pi_{i\in\mathscr{L}_{V|W}}\delta_i^v((u_v)_i|w^{1:N}(u_v)^{1:i-1})\right. \\
&\qquad\qquad \left.\Pi_{i\in R^v}P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}}((u_v)_i|w^{1:N}(u_v)^{1:i-1})\right)\cdot \\
&\quad \left(2^{-|\mathscr{H}_{X|V}|}\Pi_{i\in\mathscr{L}_{X|V}}\delta_i^x((u_x)_i|v^{1:N}(u_x)^{1:i-1})\right. \\
&\qquad\qquad \left.\Pi_{i\in R^x}P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}}((u_x)_i|v^{1:N}(u_x)^{1:i-1})\right).
\end{aligned}$$

where $w^{1:N} = (u_w)^{1:N}G_N$, $v^{1:N} = (u_v)^{1:N}G_N$ *and* $x^{1:N} = (u_x)^{1:N}G_N$.

2. Let $P_e(\mathbb{C})$ *be the probability of error for a given code in the proposed random chaining construction above with k blocks. The average probability of error for the random code construction,* $\mathbb{E}_{\mathbb{C}}[P_e(\mathbb{C})] = O(k2^{-N^{\beta'}})$ *for* $\beta' < \beta < 0.5$.

Proof:

1.

Let us consider a polar block in the random chaining construction. We now compute the ensemble

average distribution of such a block. We first evaluate $\mathbb{P}(U_w^{1:N} = u_w^{1:N}|\mathbb{C})$ for that block. Remember that in the code construction, we give the private and public message bits in a portion of $U^{\mathscr{H}_W}$ and we put randomly chosen frozen bits with i.i.d. uniform distribution in the remaining portion of it. Let I_w be index set where we put private/public message bits in $U^{\mathscr{H}_W}$ in that block. Let the randomly chosen frozen bit function be $f_w : \mathscr{H}_W - I_w \to \{0,1\}$. By encoding method, we get,

$$\begin{aligned}\mathbb{P}(U_w^{1:N} = u_w^{1:N}|\mathbb{C}) &= \Pi_{i\in[N]}\mathbb{P}((U_w)_i = (u_w)_i|\mathbb{C}, (U_w)^{1:i-1} = (u_w)^{1:i-1}) \\ &= 2^{-|I_w|}\Pi_{i\in\mathscr{H}_W - I_w}\mathbb{1}\{f_w(i) = w_i\}\Pi_{i\in\mathscr{L}_W}\delta_i^w((u_w)_i|(u_w)^{1:i-1}) \\ &\qquad \Pi_{i\in R^w}P_{(U_w)_i|(U_w)^{1:i-1}}\big((u_w)_i|(u_w)^{1:i-1}\big)\big).\end{aligned}$$

By taking expectation on both sides, by independence of frozen bits and by the linearity of expectation, we get the following:

$$\begin{aligned}\mathbb{E}_{\mathbb{C}}[\mathbb{P}(U_w^{1:N} = u_w^{1:N}|\mathbb{C})] &= 2^{-|I_w|}\Pi_{i\in\mathscr{H}_W - I_w}\mathbb{E}_{\mathbb{C}}[\mathbb{1}\{f_w(i) = w_i\}]\Pi_{i\in\mathscr{L}_W}\delta_i^w((u_w)_i|(u_w)^{1:i-1}) \\ &\qquad \Pi_{i\in R^w}P_{(U_w)_i|(U_w)^{1:i-1}}\big((u_w)_i|(u_w)^{1:i-1}\big)\big).\end{aligned}$$

This implies that

$$\begin{aligned}\mathbb{E}_{\mathbb{C}}[\mathbb{P}(U_w^{1:N} = u_w^{1:N}|\mathbb{C})] &= 2^{-|\mathscr{H}_W|}\Pi_{i\in\mathscr{L}_W}\delta_i^w((u_w)_i|(u_w)^{1:i-1}) \\ &\qquad \Pi_{i\in R^w}P_{(U_w)_i|(U_w)^{1:i-1}}\big((u_w)_i|(u_w)^{1:i-1}\big)\big).\end{aligned}$$

Similarly, we give the private and public message bits in a portion of $U^{\mathscr{H}_{V|W}}$ and we give randomly chosen frozen bits with i.i.d. uniform distribution in the remaining portion of it. Let I_v be index set where we put private/public message bits in $\mathscr{H}_{V|W}$ of the block we considered. Let the randomly chosen frozen bit function be $f_v : \mathscr{H}_{V|W} - I_v \to \{0,1\}$. By encoding rule, we get

$$\begin{aligned}\mathbb{P}(U_v^{1:N} = u_v^{1:N}|\mathbb{C}, W^{1:N} = w^{1:N}) &= 2^{-|I_v|}\Pi_{i\in\mathscr{H}_{V|W} - I_v}\mathbb{1}\{f_v(i) = v_i\} \\ &\qquad \Pi_{i\in\mathscr{L}_{V|W}}\delta_i^v((u_v)_i|w^{1:N}(u_v)^{1:i-1}) \\ &\qquad \Pi_{i\in R^v}P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}}\big((u_v)_i|w^{1:N}(u_v)^{1:i-1}\big)\big).\end{aligned}$$

By taking expectation on both sides, by the independence of frozen bits and by the linearity of expectation, we get the following:

$$\begin{aligned}\mathbb{E}_{\mathbb{C}}[\mathbb{P}(U_v^{1:N}=u_v^{1:N}|\mathbb{C},W^{1:N}=w^{1:N})] = 2^{-|I_v|}\Pi_{i\in\mathscr{H}_{V|W}-I_v}\mathbb{E}_{\mathbb{C}}[\mathbb{1}\{f_v(i)=v_i\}]\\ \Pi_{i\in\mathscr{L}_{V|W}}\delta_i^v((u_v)_i|w^{1:N}(u_v)^{1:i-1})\\ \Pi_{i\in R^v}P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}}((u_v)_i|w^{1:N}(u_v)^{1:i-1})\big).\end{aligned}$$

This implies that

$$\begin{aligned}\mathbb{E}_{\mathbb{C}}[\mathbb{P}(U_v^{1:N}=u_v^{1:N}|\mathbb{C},W^{1:N}=w^{1:N})] = 2^{-|\mathscr{H}_{V|W}|}\Pi_{i\in\mathscr{L}_{V|W}}\delta_i^v((u_v)_i|(u_v)^{1:i-1}w^{1:N})\\ \Pi_{i\in R^v}P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}}((u_v)_i|w^{1:N}(u_v)^{1:i-1}).\end{aligned}$$

Similarly, we give the private and public message bits in a portion of $U^{\mathscr{H}_{X|V}}$ and we give randomly chosen frozen bits with i.i.d. uniform distribution in the remaining portion. Let I_x be index set where we put private/public message bits in $\mathscr{H}_{X|V}$ of the block we considered. Let the randomly chosen frozen bit function be $f_x : \mathscr{H}_{X|V} - I_x \to \{0,1\}$. By encoding rule, we get

$$\begin{aligned}\mathbb{P}(U_x^{1:N}=u_x^{1:N}|\mathbb{C},V^{1:N}=v^{1:N}) = 2^{-|I_x|}\Pi_{i\in\mathscr{H}_{X|V}-I_x}\mathbb{1}\{f_x(i)=x_i\}\\ \Pi_{i\in\mathscr{L}_{X|V}}\delta_i^v((u_x)_i|v^{1:N}(u_x)^{1:i-1})\\ \Pi_{i\in R^x}P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}}((u_x)_i|v^{1:N}(u_x)^{1:i-1})\big).\end{aligned}$$

By taking expectation on both sides, by the independence of frozen bits and by the linearity of expectation, we get the following:

$$\begin{aligned}\mathbb{E}_{\mathbb{C}}[\mathbb{P}(U_x^{1:N}=u_x^{1:N}|\mathbb{C},V^{1:N}=v^{1:N})] = 2^{-|I_x|}\Pi_{i\in\mathscr{H}_{X|V}-I_x}\mathbb{E}_{\mathbb{C}}[\mathbb{1}\{f_x(i)=x_i\}]\\ \Pi_{i\in\mathscr{L}_{X|V}}\delta_i^x((u_x)_i|v^{1:N}(u_x)^{1:i-1})\\ \Pi_{i\in R^x}P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}}((u_x)_i|v^{1:N}(u_x)^{1:i-1})\big).\end{aligned}$$

This implies that

$$\mathbb{E}_{\mathbb{C}}[\mathbb{P}(U_x^{1:N} = u_x^{1:N}|\mathbb{C}, V^{1:N} = v^{1:N})]$$
$$= 2^{-|\mathscr{H}_{X|V}|}\Pi_{i\in\mathscr{L}_{X|V}}\delta_i^x((u_x)_i|v^{1:N}(u_x)^{1:i-1})\Pi_{i\in R^x}P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}}((u_x)_i|v^{1:N}(u_x)^{1:i-1}).$$

By the chain-rule of conditional probability, we get

$$\mathbb{P}(U_w^{1:N} = u_w^{1:N}, U_v^{1:N} = u_v^{1:N}, U_x^{1:N} = u_x^{1:N}|\mathbb{C})$$
$$= \mathbb{P}(U_w^{1:N} = u_w^{1:N}|\mathbb{C}) \cdot \mathbb{P}(U_v^{1:N} = u_v^{1:N}|\mathbb{C}, W^{1:N} = w^{1:N}) \cdot$$
$$\mathbb{P}(U_x^{1:N} = u_x^{1:N}|\mathbb{C}, V^{1:N} = v^{1:N}).$$

By taking expectations on the both the sides and by using the fact that the frozen bit functions f_w, f_v and f_x are independent, we get the following:

$$\mathbb{E}_{\mathbb{C}}[\mathbb{P}(U_w^{1:N} = u_w^{1:N}, U_v^{1:N} = u_v^{1:N}, U_x^{1:N} = u_x^{1:N}|\mathbb{C})]$$
$$= \mathbb{E}_{\mathbb{C}}[\mathbb{P}(U_w^{1:N} = u_w^{1:N}|\mathbb{C})] \cdot \mathbb{E}_{\mathbb{C}}[\mathbb{P}(U_v^{1:N} = u_v^{1:N}|\mathbb{C}, W^{1:N} = w^{1:N})] \cdot$$
$$\mathbb{E}_{\mathbb{C}}[\mathbb{P}(U_x^{1:N} = u_x^{1:N}|\mathbb{C}, V^{1:N} = v^{1:N})].$$

After substituting each of the three product terms on the right hand side, we finish the proof of part 1.

2.

Let $\mathscr{E}$ be the error event. Notice that the error occurs if and only if there is an error while decoding bit-channels $\mathscr{L}_W \cup I_j^w$ in $(U_w)^{1:N}$ for $j = 1,2,3$ or $\mathscr{L}_{V|W} \cup I_j^v$ in $(U_v)^{1:N}$ for $j = 1,3$ or $\mathscr{L}_{X|V} \cup I_1^x$ in $(U_x)^{1:N}$ in any of the blocks involved in the chaining construction. Let us index the blocks in chaining construction as $b = 1,2,\ldots,k$.

The error event of bit-channel i of block b for receivers $j = 1,2$ or 3 in the first layer will be as follows:

$$\mathscr{E}_{ij}^{wb} = \{(w^{1:N}, v^{1:N}, x^{1:N}, y_j^{1:N})\text{s of all the blocks } \tilde{b} \in [k] \ :$$
$$P_{(U_w)_i|(U_w)^{1:i-1}Y_j^{1:N}}((u_w)_i + 1|(u_w)^{1:i-1}y_j^{1:N})$$
$$\geq P_{(U_w)_i|(U_w)^{1:i-1}Y_j^{1:N}}((u_w)_i|(u_w)^{1:i-1}y_j^{1:N})$$
$$\text{holds for } (u_w^{1:N}, y_j^{1:N}) \text{ of block } b\}.$$

When there is only a single block, the error event of bit-channel i for receivers $j = 1, 2$ or 3 in the first layer will be as follows:

$$\begin{aligned}\mathscr{E}_{ij}^{w} = \{&(w^{1:N}, v^{1:N}, x^{1:N}, y_j^{1:N}) : \\ &P_{(U_w)_i|(U_w)^{1:i-1}Y_j^{1:N}}((u_w)_i + 1|(u_w)^{1:i-1}y_j^{1:N}) \\ &\geq P_{(U_w)_i|(U_w)^{1:i-1}Y_j^{1:N}}((u_w)_i|(u_w)^{1:i-1}y_j^{1:N})\}.\end{aligned}$$

The error event of bit-channel i of block b for receivers $j = 1$ or 3 in the second layer will be as follows:

$$\begin{aligned}\mathscr{E}_{ij}^{vb} = \{&(w^{1:N}, v^{1:N}, x^{1:N}, y_j^{1:N})\text{s of all the blocks } \tilde{b} \in [k] \ : \\ &P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}Y_j^{1:N}}((u_v)_i + 1|w^{1:N}(u_w)^{1:i-1}y_j^{1:N}) \\ &\geq P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}Y_j^{1:N}}((u_v)_i|w^{1:N}(u_w)^{1:i-1}y_j^{1:N}) \\ &\qquad \text{holds for } (w^{1:N}, u_v^{1:N}, y_j^{1:N}) \text{ of block } b\}.\end{aligned}$$

When there is only a single block, the error event of bit-channel i for receivers $j = 1$ or 3 in the second layer will be as follows:

$$\begin{aligned}\mathscr{E}_{ij}^{v} = \{&(w^{1:N}, v^{1:N}, x^{1:N}, y_j^{1:N}) : \\ &P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}Y_j^{1:N}}((u_v)_i + 1|w^{1:N}(u_w)^{1:i-1}y_j^{1:N}) \\ &\geq P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}Y_j^{1:N}}((u_v)_i|w^{1:N}(u_w)^{1:i-1}y_j^{1:N})\}.\end{aligned}$$

The error event of bit-channel i of block b for receiver $j = 1$ in the third layer will be as follows:

$$\begin{aligned}\mathscr{E}_{ij}^{xb} = \{&(w^{1:N}, v^{1:N}, x^{1:N}, y_j^{1:N})\text{s of all the blocks } \tilde{b} \in [k] \ : \\ &P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}Y_j^{1:N}}((u_x)_i + 1|v^{1:N}(u_x)^{1:i-1}y_j^{1:N})\end{aligned}$$

$$\geq P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}Y_j^{1:N}}((u_x)_i|v^{1:N}(u_x)^{1:i-1}y_j^{1:N})$$
$$\text{holds for } (v^{1:N},u_x^{1:N},y_j^{1:N}) \text{ of block } b\}.$$

When there is only a single block, the error event of bit-channel i for receiver $j=1$ in the third layer will be as follows:

$$\begin{aligned}\mathscr{E}_{ij}^{x} = \{&(w^{1:N},v^{1:N},x^{1:N},y_j^{1:N}) : \\ &P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}Y_j^{1:N}}((u_x)_i+1|v^{1:N}(u_x)^{1:i-1}y_j^{1:N}) \\ &\geq P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}Y_j^{1:N}}((u_x)_i|v^{1:N}(u_x)^{1:i-1}y_j^{1:N})\}.\end{aligned}$$

We define $\mathscr{E}_j^{wb} = \cup_{i\in I_j^w \cup \mathscr{L}_W}\mathscr{E}_{ij}^{wb}$ for $j=1,2,3$, $\mathscr{E}_j^{vb} = \cup_{i\in I_j^v\cup\mathscr{L}_{V|W}}\mathscr{E}_{ij}^{vb}$ for $j=1,3$ and $\mathscr{E}_j^{xb} = \cup_{i\in I_j^x\cup\mathscr{L}_{X|V}}\mathscr{E}_{ij}^{xb}$ for $j=1$.

We define $\mathscr{E}_j^{w} = \cup_{i\in I_j^w \cup \mathscr{L}_W}\mathscr{E}_{ij}^{w}$ for $j=1,2,3$, $\mathscr{E}_j^{v} = \cup_{i\in I_j^v\cup\mathscr{L}_{V|W}}\mathscr{E}_{ij}^{v}$ for $j=1,3$ and $\mathscr{E}_j^{x} = \cup_{i\in I_j^x\cup\mathscr{L}_{X|V}}\mathscr{E}_{ij}^{x}$ for $j=1$.

We define $\mathscr{E}_1^b = \mathscr{E}_1^{wb}\cup\mathscr{E}_1^{vb}\cup\mathscr{E}_1^{xb}$, $\mathscr{E}_2^b = \mathscr{E}_2^{wb}$ and $\mathscr{E}_3^b = \mathscr{E}_3^{wb}\cup\mathscr{E}_3^{vb}$ for each block b.

We define $\mathscr{E}_{1s} = \mathscr{E}_1^{w}\cup\mathscr{E}_1^{v}\cup\mathscr{E}_1^{x}$, $\mathscr{E}_{2s} = \mathscr{E}_2^{w}$ and $\mathscr{E}_{3s} = \mathscr{E}_3^{w}\cup\mathscr{E}_3^{v}$ for each block b.

We define $\mathscr{E}_j = \cup_{b=1}^k \mathscr{E}_j^b$, which will be error event for receiver-j, where $j=1,2,3$.

Therefore the overall error event $\mathscr{E} = \cup_{j=1}^3\mathscr{E}_j$. By union bound, we the following identity:

$\mathbb{P}(\mathscr{E}|\mathbb{C}) \leq \sum_{j=1}^3\sum_{b=1}^k \mathbb{P}(\mathscr{E}_j^b|\mathbb{C})$.

By taking expectation on both the sides and also by applying linearity of expectation, we get

$$\mathbb{E}_{\mathbb{C}}[\mathbb{P}(\mathscr{E}|\mathbb{C})] \leq \sum_{j=1}^{3}\sum_{b=1}^{k}\mathbb{E}_{\mathbb{C}}[\mathbb{P}(\mathscr{E}_j^b|\mathbb{C})]. \tag{5.16}$$

Let $Q_{((U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N})}$ be the measure on $((U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N})$ as follows:

$$\begin{aligned}&Q_{((U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}(u_w^{1:N},u_v^{1:N},u_x^{1:N}) \\ &= Q_{(U_w)^{1:N}}(u_w^{1:N})Q_{(U_v)^{1:N}|W^{1:N}=w^{1:N}}(u_w^{1:N})Q_{(U_x)^{1:N}|V^{1:N}=v^{1:N}}(u_v^{1:N}) \\ &= \left(2^{-|\mathscr{H}_W|}\Pi_{i\in\mathscr{L}_W}\delta_i^w((u_w)_i|(u_w)^{1:i-1})\Pi_{i\in R^w}P_{(U_w)_i|(U_w)^{1:i-1}}((u_w)_i|(u_w)^{1:i-1})\right).\end{aligned}$$

$$\Big(2^{-|\mathscr{H}_{V|W}|}\Pi_{i\in\mathscr{L}_{V|W}}\delta_i^v((u_v)_i|w^{1:N}(u_v)^{1:i-1})$$
$$\Pi_{i\in R^v}P_{(U_v)_i|W^{1:N}(U_v)^{1:i-1}}((u_v)_i|w^{1:N}(u_v)^{1:i-1})\Big)\cdot$$
$$\Big(2^{-|\mathscr{H}_{X|V}|}\Pi_{i\in\mathscr{L}_{X|V}}\delta_i^x((u_x)_i|v^{1:N}(u_x)^{1:i-1})$$
$$\Pi_{i\in R^x}P_{(U_x)_i|V^{1:N}(U_x)^{1:i-1}}((u_x)_i|v^{1:N}(u_x)^{1:i-1})\Big).$$

Note that $P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}$ is the measure induced when $(W^{1:N},V^{1:N},X^{1:N})$ is i.i.d. according to the distribution $p(w)p(v|w)p(x|v)$.

From Lemma 11, we have

$||P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}-Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}||=O(2^{-N^{\beta'}})$, where $\beta'<\beta$.

$\mathbb{P}(\mathscr{E}_j^b|\mathbb{C})$

$=\sum_{((u_w)^{1:N},(u_v)^{1:N},(u_x)^{1:N},y_j^{1:N})\text{s of all blocks }[k])\in\mathscr{E}_j^b}$
$\mathbb{P}(\cap_{\tilde{b}\in[k]}(U_w^{1:N}=u_w^{1:N},U_v^{1:N}=u_v^{1:N},U_x^{1:N}=u_x^{1:N},Y_j^{1:N}=y_j^{1:N}\text{ of block }\tilde{b})|\mathbb{C})$.

From the definitions of $\mathscr{E}_j^b$ and $\mathscr{E}_{js}$, we get

$\mathbb{P}(\mathscr{E}_j^b|\mathbb{C})$

$=\sum_{(((u_w)^{1:N},(u_v)^{1:N},(u_x)^{1:N},y_j^{1:N})\text{ of block }b)\in\mathscr{E}_{js}}$
$\sum_{(((u_w)^{1:N},(u_v)^{1:N},(u_x)^{1:N},y_j^{1:N})\text{s of blocks }[k]-\{b\})}$
$\mathbb{P}(\cap_{\tilde{b}\in[k]}(U_w^{1:N}=u_w^{1:N},U_v^{1:N}=u_v^{1:N},U_x^{1:N}=u_x^{1:N},Y_j^{1:N}=y_j^{1:N}\text{ of block }\tilde{b})|\mathbb{C})$.

By marginalizing over

$(U_w^{1:N},U_v^{1:N},U_x^{1:N},Y_j^{1:N})$s of blocks $[k]-\{b\}$, we now get

$\mathbb{P}(\mathscr{E}_j^b|\mathbb{C})=\sum_{(((u_w)^{1:N},(u_v)^{1:N},(u_x)^{1:N},y_j^{1:N})\text{ of block }b)\in\mathscr{E}_{js}}$
$\mathbb{P}((U_w^{1:N}=u_w^{1:N},U_v^{1:N}=u_v^{1:N},U_x^{1:N}=u_x^{1:N},Y_j^{1:N}=y_j^{1:N}\text{ of block }b)|\mathbb{C})$.

By chain rule of condition probability and also by the fact that

$\mathbb{P}(Y_j^{1:N}=y_j^{1:N}\text{ of block }b|X^{1:N}=x^{1:N}\text{ of block }b,\,\mathbb{C})=\Pi_{i=1}^N p(y_{ji}|x_i)$,

we will have the following:

$\mathbb{P}(\mathscr{E}_j^b|\mathbb{C})=\sum_{(((u_w)^{1:N},(u_v)^{1:N},(u_x)^{1:N},y_j^{1:N})\text{ of block }b)\in\mathscr{E}_{js}}$
$\mathbb{P}((U_w^{1:N}=u_w^{1:N},U_v^{1:N}=u_v^{1:N},U_x^{1:N}=u_x^{1:N}\text{ of block }b|\mathbb{C})\Pi_{i=1}^N p(y_{ji}|x_i)$.

In the term $\Pi_{i=1}^N p_l(y_{ji}|x_i)$ here, notice that $x^{1:N}$ vector is corresponding to block b, which means it is obtained by applying polar transform to $(u_x)^{1:N}$ vector corresponding to block b and also

$y_j^{1:N}$ vector is corresponding to block b.

By taking expectation on both the sides and by the linearity of expectation, we get the following:

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{C}}[\mathbb{P}(\mathscr{E}_j^b|\mathbb{C})] \\
&= \sum_{((u_w)^{1:N},(u_v)^{1:N},(u_x)^{1:N},y_j^{1:N}) \text{ of block } b) \in \mathscr{E}_{js}} \\
&\qquad \mathbb{E}_{\mathbb{C}}[\mathbb{P}(U_w^{1:N}=u_w^{1:N},U_v^{1:N}=u_v^{1:N},U_x^{1:N}=u_x^{1:N} \text{ of block } b|\mathbb{C})]\Pi_{i=1}^N p(y_{ji}|x_i) \\
&\overset{(a)}{=} \sum_{((u_w)^{1:N},(u_v)^{1:N},(u_x)^{1:N},y_j^{1:N}) \text{ of block } b) \in \mathscr{E}_{js}} \\
&\qquad\qquad Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}(u_w^{1:N}u_v^{1:N}u_x^{1:N})\Pi_{i=1}^N p(y_{ji}|x_i) \\
&= Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_j^{1:N}}(\mathscr{E}_{js}) \\
&\leq ||P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_j^{1:N}} - Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_j^{1:N}}|| \\
&\qquad\qquad\qquad\qquad + P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_j^{1:N}}(\mathscr{E}_{js}) \\
&\overset{(b)}{=} ||P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}} - Q_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}|| + P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_j^{1:N}}(\mathscr{E}_{js}) \\
&= O(2^{-N^{\beta'}}) + P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_j^{1:N}}(\mathscr{E}_{js}).
\end{aligned}
$$

Identity (a) follows from part 1. Identity (b) follows from Lemma 4.

For receiver-1, that is $j=1$, we get

$$
\begin{aligned}
\mathbb{E}_C[\mathbb{P}(\mathscr{E}_1^b)|\mathbb{C}] &= O(2^{-N^{\beta'}}) + P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_1^{1:N}}(\mathscr{E}_{1s}) \\
&\overset{(a)}{\leq} O(2^{-N^{\beta'}}) + P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}}(\mathscr{E}_1^w) + P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_1^{1:N}}(\mathscr{E}_1^v) \\
&\qquad + P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_1^{1:N}}(\mathscr{E}_1^x) \\
&\overset{(b)}{\leq} O(2^{-N^{\beta'}}) + \sum_{i\in\mathscr{L}_W\cup I_1^w} P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_1^{1:N}}(\mathscr{E}_{i1}^w) \\
&\qquad + \sum_{i\in\mathscr{L}_{V|W}\cup I_1^v} P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_1^{1:N}}(\mathscr{E}_{i1}^v) \\
&\qquad + \sum_{i\in\mathscr{L}_{X|V}\cup I_1^x} P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_1^{1:N}}(\mathscr{E}_{i1}^x) \\
&\leq O(2^{-N^{\beta'}}) + \sum_{i\in\mathscr{L}_W\cup I_1^w} Z((U_w)_i|(U_w)^{1:i-1}Y_1^{1:N}) \\
&\qquad + \sum_{i\in\mathscr{L}_{V|W}\cup I_1^v} Z((U_v)_i|W^{1:N}(U_v)^{1:i-1}Y_1^{1:N}) \\
&\qquad + \sum_{i\in\mathscr{L}_{X|V}\cup I_1^x} Z((U_x)_i|V^{1:N}(U_x)^{1:i-1}Y_1^{1:N}) \\
&\leq O(2^{-N^{\beta'}}) + N2^{-N^{\beta}} + N2^{-N^{\beta}} + N2^{-N^{\beta}} \\
&\leq O(2^{-N^{\beta'}}).
\end{aligned}
$$

Identity (a) follows from the definition of $\mathscr{E}_{1s}$ and union bound. Identity (b) follows from the

definition of $\mathscr{E}_1^w$, $\mathscr{E}_1^v$, $\mathscr{E}_1^x$ and union bound.

For receiver-2, that is $j=2$, we get

$$\begin{aligned}
\mathbb{E}_C[\mathbb{P}(\mathscr{E}_2^b)|\mathbb{C}] &= O(2^{-N^{\beta'}}) + P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_2^{1:N}}(\mathscr{E}_{2s}) \\
&\overset{(a)}{\leq} O(2^{-N^{\beta'}}) + P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_2^{1:N}}(\mathscr{E}_2^w) \\
&\overset{(b)}{\leq} O(2^{-N^{\beta'}}) + \textstyle\sum_{i\in\mathscr{L}_W\cup I_2^w} P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_2^{1:N}}(\mathscr{E}_{i2}^w) \\
&\leq O(2^{-N^{\beta'}}) + \textstyle\sum_{i\in\mathscr{L}_W\cup I_2^w} Z((U_w)_i|(U_w)^{1:i-1}Y_2^{1:N}) \\
&\leq O(2^{-N^{\beta'}}) + N2^{-N^{\beta}} \\
&= O(2^{-N^{\beta'}}).
\end{aligned}$$

Identity (a) follows from the definition of $\mathscr{E}_{2s}$. Identity (b) follows from the definition of $\mathscr{E}_2^w$ and union bound.

For receiver-3, that is $j=3$, we get

$$\begin{aligned}
\mathbb{E}_C[\mathbb{P}(\mathscr{E}_3^b)|\mathbb{C}] &= O(2^{-N^{\beta'}}) + P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_3^{1:N}}(\mathscr{E}_{3s}) \\
&\overset{(a)}{\leq} O(2^{-N^{\beta'}}) + P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_3^{1:N}}(\mathscr{E}_3^w) \\
&\qquad + P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_3^{1:N}}(\mathscr{E}_3^v) \\
&\overset{(b)}{\leq} O(2^{-N^{\beta'}}) + \textstyle\sum_{i\in\mathscr{L}_W\cup I_3^w} P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_3^{1:N}}(\mathscr{E}_{i3}^w) \\
&\qquad + \textstyle\sum_{i\in\mathscr{L}_{V|W}\cup I_3^v} P_{(U_w)^{1:N}(U_v)^{1:N}(U_x)^{1:N}Y_3^{1:N}}(\mathscr{E}_{i3}^v) \\
&\leq O(2^{-N^{\beta'}}) + \textstyle\sum_{i\in\mathscr{L}_W\cup I_3^w} Z((U_w)_i|(U_w)^{1:i-1}Y_3^{1:N}) \\
&\qquad + \textstyle\sum_{i\in\mathscr{L}_{V|W}\cup I_3^v} Z((U_v)_i|W^{1:N}(U_v)^{1:i-1}Y_3^{1:N}) \\
&\leq O(2^{-N^{\beta'}}) + N2^{-N^{\beta}} + N2^{-N^{\beta}} \\
&\leq O(2^{-N^{\beta'}}).
\end{aligned}$$

Identity (a) follows from the definition of $\mathscr{E}_{3s}$ and union bound. Identity (b) follows from the definition of $\mathscr{E}_3^w$, $\mathscr{E}_3^v$ and union bound.

From equation (5.16), the overall average probability of error will become $O(k2^{-N^{\beta'}})$. This concludes the proof of part 2. Hence the proof of Theorem 6. □

Both encoding and decoding complexities will become $O(N\log N)$ per block [24].

We have given the code-construction for the case where $|\mathscr{X}|=|\mathscr{V}|=|\mathscr{W}|=2$. If any of

these alphabets have arbitrary sizes, we can adapt multi-level polar code construction technique. Let $|\mathscr{X}| = \Pi_{j=1}^{m} p_j$ $|\mathscr{V}| = \Pi_{j=1}^{l} q_j$ $|\mathscr{W}| = \Pi_{j=1}^{k} r_j$ where $\{r_j\}$, $\{q_j\}$ and $\{p_j\}$ are prime factors of $\mathscr{W}$, $\mathscr{V}$ and $\mathscr{X}$, respectively. Then random variables W, V and X can be represented by random vectors $(W_1, \ldots, W_k)$, $(V_1, \ldots, V_l)$ and $(X_1, \ldots, X_m)$ where W_j, V_j and X_j are supported over the set $\{0, 1, \ldots, r_j - 1\}$, $\{0, 1, \ldots q_j - 1\}$ and $\{0, 1, \ldots p_j - 1\}$, respectively. By chain-rule of entropy, we get $H(W,V,X) = \Sigma_{j=1}^{k} H(W_j|W^{1:j-1}) + \Sigma_{j=1}^{l} H(V_j|WV^{1:j-1}) + \Sigma_{j=1}^{m} H(X_j|WVX^{1:j-1})$. We can use the polarization for prime alphabets for each term in the above identity and derive a polar code construction technique with an appropriate successive cancellation decoder [47], [46] for larger alphabets. The key ideas in the analysis of the probability of error we provided for the binary case still apply to the coding method for larger alphabets and can be extended.

5.3.5 Extension: receiver-1 requires only M_1

For a $(2^{NR_0}, 2^{NR_1}, N)$ code of a setting with degraded message sets, the converse proof of the capacity region just uses the fact that $H(M_1|Y_1^{1:N})$, $H(M_0|Y_2^{1:N})$ and $H(M_0|Y_3^{1:N})$ are $o(N)$ [38]. We do not have to use the stronger fact that $H(M_1, M_0|Y_1^{1:N})$ is $o(N)$ to complete the converse proof. This means that the same proof becomes the converse proof of the capacity region for the problem when receiver-1 is relaxed to recover only M_1. Hence, the capacity region does not enlarge and remains the same. So the same polar coding method can be used to achieve all rate pairs inside the capacity region.

5.4 Conclusion

We considered the problem of achieving the rates in the capacity region of a discrete memoryless multi-level 3-receiver broadcast channel with degraded message sets through polar coding. The problem is to transmit a public message to all the receivers and a private message intended for receiver-1. Our motivation for this problem is due to a file transfer application in a client-server network that has three clients, where this setting can be applied. We give a new two-level chaining construction to achieve all the points in the capacity region without

time-sharing. We also gave a detailed analysis of the probability of decoding error for constructed coding scheme. We showed that the capacity of the broadcast channel does not enlarge, even when receiver-1 is required to recover only its private message. Hence, we can use the same polar coding strategy to achieve the capacity under this setting.

Acknowledgement

This chapter is in part a reprint of the material in the paper: Karthik Nagarjuna Tunuguntla, Paul H. Siegel, "Polar coding for multi-level 3-receiver broadcast channels," *2020 Information Theory Workshop (ITW)*, pp. 1-5, Riva Del Garda, Italy, April 2021. Dissertation author is the primary contributor of the paper.

Bibliography

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

[2] E. Arikan, "Polar coding for the Slepian-Wolf problem based on monotone chain rules," *Proc. IEEE Int. Symp. Inf. Theory,* Cambridge, MA, 2012, pp. 566-570.

[3] E. Arikan, "Source polarization," *Proc. IEEE Int. Symp. Inf. Theory,* 2010, pp. 899-903.

[4] E. Arikan and E. Telatar, "On the rate of channel polarization,"*Proc. IEEE Int. Symp. Inf. Theory,* 2009, pp. 1493-1495.

[5] E. Abbe and E. Telatar, "Polar Codes for the *m*-User Multiple Access Channel," *IEEE Trans. Inf. Theory,* vol. 58, no. 8, pp. 5437-5448, Aug. 2012.

[6] A. Bhatt, N. Ghaddar and L. Wang, "Polar coding for multiple descriptions using monotone chain rules," *Proc. 55rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2017, pp. 565-571.

[7] R.E. Blahut, *Theory and Practice of Error Control Codes*. Addison-Wesley, Reading, Massachusetts, 1983.

[8] G. Böcherer, "Capacity-achieving probabilistic shaping for noisy and noiseless channels," *Ph.D. dissertation, RWTH Aachen University, 2012* [Online]. Available: http://www.georg-boecherer. de/capacityAchievingShaping.pdf

[9] G. Böcherer, F. Steiner and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commmun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.

[10] R. A. Chou and M. R. Bloch, "Polar coding for the broadcast channel with confidential messages," *Proc. IEEE Information Theory Workshop (ITW)*, Jerusalem, 2015, pp. 1–5.

[11] R. A. Chou and M. R. Bloch "Using deterministic decisions for low-entropy bits in the encoding and decoding of polar codes," *Proc. 53rd Annu. Allerton Conf. on Commun., Control, and Computing (Allerton 2015),* Monticello, IL, Sep. 2015, pp. 1380–1385.

[12] R. A. Chou and A. Yener, "Polar coding for the multiple access wiretap channel via rate-splitting and cooperative jamming,"*Proc. IEEE Int. Symp. Inf. Theory,* 2016, pp. 983-987.

[13] H. S. Cronie and S. B. Korada, "Lossless source coding with polar codes,"*Proc. IEEE Int. Symp. Inf. Theory,* 2010, pp. 904-908.

[14] A. El Gamal and Y. H. Kim, *Network Information Theory*, Cambridge, UK: Cambridge University Press, 2011.

[15] E. En Gad, Y. Li, J. Kliewer, M. Langberg, A. A. Jiang and J. Bruck, "Asymmetric error correction and flash-memory rewriting using polar codes," *IEEE Trans. Inf. Theory,* vol. 62, no. 7, pp. 4024–4038, Jul. 2016.

[16] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. Hoboken, NJ, USA: Wiley, 2002.

[17] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for broadcast channels," *IEEE Trans. Inf. Theory,* vol. 61, no. 2, pp. 758–782, Feb. 2015.

[18] D. Goldin and D. Burshtein, "Improved Bounds on the Finite Length Scaling of Polar Codes," *IEEE Trans. Inf. Theory,* vol. 60, no. 11, pp. 6966-6978, Nov. 2014.

[19] V. Guruswami and P. Xia, "Polar Codes: Speed of Polarization and Polynomial Gap to Capacity," *2013 IEEE 54th Annual Symposium on Foundations of Computer Science,* 2013, pp. 310-319.

[20] S. H. Hassani, K. Alishahi and R. L. Urbanke, "Finite-Length Scaling for Polar Codes," *IEEE Trans. Inf. Theory,* vol. 60, no. 10, pp. 5875-5898, Oct. 2014.

[21] S. H. Hassani, R. Mori, T. Tanaka and R. L. Urbanke, "Rate-Dependent Analysis of the Asymptotic Behavior of Channel Polarization," *IEEE Trans. Inf. Theory,* vol. 59, no. 4, pp. 2267-2276, April 2013.

[22] S. H. Hassani and R. L. Urbanke, *Universal polar codes, CoRR (2013),*abs/1307.7223.

[23] S. H. Hassani and R. L. Urbanke, "Universal polar codes," *Proc. IEEE Int. Symp. Inf. Theory,* Honolulu, HI, Jul. 2014, pp. 1451–1455.

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

[25] S. B. Korada, E. Şaşoğlu and R. Urbanke, "Polar Codes: Characterization of Exponent, Bounds," *IEEE Trans. Inf. Theory,* vol. 56, no. 12, pp. 6253-6264, Dec. 2010.

[26] S. B. Korada and R. Urbanke, "Polar Codes are Optimal for Lossy Source Coding," *IEEE Trans. Inf. Theory,* vol. 56, no. 4, pp. 1751-1768, April 2010.

[27] S. B. Korada and R. Urbanke, "Polar codes for Slepian-Wolf, Wyner-Ziv, and Gelfand-Pinsker," *Proc. IEEE Information Theory Workshop (ITW)*, Cairo, 2010, pp. 1-5.

[28] A. Lenz, Y. Liu, C. Rashtchian, P. H. Siegel, A. Wachter-Zeh and E. Yaakobi, "Coding for efficient DNA synthesis," *IEEE Int. Symp. Inf. Theory (ISIT),* Los Angeles, CA, Jun. 21-26, 2020, pp. 2885–2890.

[29] Y. Liu, P. Huang, A. W. Bergman and P. H. Siegel, "Rate-constrained shaping codes for structured sources," *IEEE Trans. Inf. Theory,* vol. 66, no. 8, pp. 5261-5281, Aug. 2020.

[30] H. Mahdavifar and A. Vardy, "Achieving the Secrecy Capacity of Wiretap Channels Using Polar Codes," *IEEE Trans. Inf. Theory,* vol. 57, no. 10, pp. 6428-6443, Oct. 2011.

[31] T. Matsumine, T. Koike-Akino, D. S. Millar, K. Kojima and K. Parsons, "Polar coded modulation for joint channel coding and probabilistic shaping," *Proc. Optical Fiber Commun. Conf. (OFC)*, San Diego, CA, Mar. 3-7, 2019, pp. 1–3.

[32] M. Mondelli, S. H. Hassani, R. Urbanke and I. Sason, "Achieving Marton's region for broadcast channels using polar codes," *IEEE Trans. Inf. Theory,* vol. 61, no. 2, pp. 783–800, Feb. 2015.

[33] M. Mondelli, R. Urbanke and S. H. Hassani, "How to achieve the capacity of asymmetric channels," *Proc. 52th Annu. Allerton Conf. on Commun., Control, and Computing (Allerton 2014),* Monticello, IL, Oct. 2014, pp. 789–796.

[34] R. Mori and T. Tanaka, "Channel polarization on q-ary discrete memoryless channels by arbitrary kernels," *Proc. IEEE Int. Symp. Inf. Theory,* 2010, pp. 894-898.

[35] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," *Proc. IEEE Int. Symp. Inf. Theory,* 2009, pp. 1496-1500.

[36] R. Mori and T. Tanaka, "Source and Channel Polarization Over Finite Fields and Reed–Solomon Matrices," *IEEE Trans. Inf. Theory,* vol. 60, no. 5, pp. 2720-2736, May 2014.

[37] K. Nagarjuna and P. H. Siegel, "Universal polar coding for asymmetric channels," *Proc. IEEE Inf. Theory Workshop (ITW)*, Guangzhou, China, Nov. 25-29, 2018, pp. 1–5.

[38] C. Nair and A. El Gamal, "The capacity region of a class of three-receiver broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory,* vol. 55, no. 10, pp. 4479-4493, Oct. 2009.

[39] J. d. Olmo Alòs and J. R. Fonollosa, "Polar coding for common message only wiretap broadcast channel," *Proc. IEEE Int. Symp. Inf. Theory*, Paris, France, 2019, pp. 1762–1766.

[40] S. Önay, "Successive cancellation decoding of polar codes for the two-user binary-input MAC," *Proc. IEEE Int. Symp. Inf. Theory*, 2013, pp. 1122-1126.

[41] R. Pedarsani, S. H. Hassani, I. Tal and E. Telatar, "On the construction of polar codes," *Proc. IEEE Int. Symp. Inf. Theory*, 2011, pp. 11-15.

[42] H. D. Pfister and R. L. Urbanke, "Near-Optimal Finite-Length Scaling for Polar Codes Over Large Alphabets," *IEEE Trans. Inf. Theory,* vol. 65, no. 9, pp. 5643-5655, Sept. 2019.

[43] N. Presman, O. Shapira and S. Litsyn, "Polar codes with mixed kernels," *Proc. IEEE Int. Symp. Inf. Theory*, 2011, pp. 6-10.

[44] T. Prinz, P. Yuan, G. Böcherer, F. Steiner, O. İşcan, R. Böhnke and W. Xu, "Polar coded probabilistic amplitude shaping for short packets," *Proc. IEEE 18th Int. Workshop on Sign. Proc. Advances in Wireless Commun. (SPAWC)*, Sapporo, Japan, Jul. 3-6, 2017, pp 1–5.

[45] R. Roth, *Introduction to Coding Theory*. Cambridge University Press, 2006.

[46] E. Şaşoğlu, "Polarization and polar codes," *Found. Trends Commun. Inf. Theory,* vol. 8, no. 4, pp. 259–381, Oct. 2012.

[47] E. Şaşoğlu, E. Telatar and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," *Proc. IEEE Inf. Theory Workshop (ITW),* Taormina, Italy, Oct. 11-16, 2009, pp. 144–148.

[48] E. Şaşoğlu and L. Wang, "Universal polarization," *IEEE Trans. Inf. Theory,* vol. 62, no. 6, pp. 2937-2946, June 2016.

[49] C. E. Shannon, "A mathematical theory of communication, Part I, Part II," *Bell Syst. Tech. J.,* vol. 27, pp. 379–423, 1948.

[50] I. Tal and A. Vardy, "How to Construct Polar Codes," *IEEE Trans. Inf. Theory,* vol. 62, no. 10, pp. 6562-6582, Oct 2013.

[51] T. Tanaka and R. Mori, "Refined rate of channel polarization,"*Proc. IEEE Int. Symp. Inf. Theory,* 2010, pp. 889-893.

[52] V. Taranalli, H. Uchikawa and P. H. Siegel, "On the capacity of the beta-binomial channel model for multi-level cell flash memories," *Proc. IEEE J. Select. Areas Commun.* vol. 34, no. 9, pp. 2312–2324, Sep. 2016.

[53] L. Wang, "Polar coding for relay channels,"*Proc. IEEE Int. Symp. Inf. Theory,* 2015, pp. 1532-1536.

[54] L. Wang, "Polar coding for interference networks,"*Proc. IEEE Int. Symp. Inf. Theory,* 2014, pp. 311-315.

[55] L. Wang and Y. Kim, "Linear code duality between channel coding and Slepian-Wolf coding," *Proc. 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, 2015, pp. 147-152.