

# Universal polar coding for binary input asymmetric models

Karthik Nagarjuna and Paul H. Siegel  
CMRR, University of California, San Diego  
Email: {tkarthik,psiegel}@ucsd.edu

**Abstract**—In this work, we first revisit the integrated scheme for achieving the capacity of binary input asymmetric channels proposed by Honda and Yamamoto. The synthetic channels due to polarization in this scheme are categorized as good channels, bad channels, deterministic channels and synthetic channels which are not polarized to the required level whose fraction is diminishing with respect to block length. We give a rigorous proof that argmax rule can also be used for encoding deterministic synthetic channels instead of randomized rounding. Secondly and mainly we are interested in universal polar codes for binary input asymmetric channels for achieving rates close to the capacity of compound channel. We propose two coding schemes that are a hybrid combination of the integrated scheme and universal polar coding schemes for symmetric channels proposed by Hassani and Urbanke. By exploiting the stair case structure in one of the proposed asymmetric universal polar coding schemes, neither side-channel transmission nor sharing boolean functions with high storage is required for bits corresponding to not completely polarized synthetic channels.

## I. INTRODUCTION

Arikan in [1] constructed capacity achieving codes for binary input symmetric channels. In this paper we first revisit the integrated coding scheme proposed by Honda and Yamamoto for achieving the binary input asymmetric channel capacity [6]. This scheme is later modified by authors [4] to bring down an exponential storage complexity to polynomial storage complexity with the aid of a side channel which is used to transmit at a rate which can be made arbitrarily small. If a side channel is not used, pseudo-random functions need to be shared between encoder and decoder for those indices of synthetic channels which are not completely polarized whose fraction is diminishing with respect to block length. In [4], for encoding deterministic channel indices randomized rounding rule is used. The authors in [7] mentioned that argmax rule for encoding these indices give better simulation results than randomized rounding rule. Here we give a rigorous proof that argmax rule can be used to achieve the capacity. We found later that similar result has been done in [3]. Here we include our proof just in the context of binary input asymmetric channel setting and our proof steps are not exactly same as in [3] that give same result.

In independent work [5] by Hassani and Urbanke, different code constructions are explained to achieve rates less than compound capacity using polar codes for binary input symmetric DMCs. In [8] Taranalli and Siegel computed the capacity of truncated support beta binomial model which is essentially computing the compound capacity for asymmetric

channel on a support set. This work is done in the context of modeling flash memory channel. So compound capacity achieving schemes for binary input asymmetric channels are of practical interest. Here we are interested to find universal polar codes for compound channel with a finite set of asymmetric binary input DMCs. The compound channel is a set of discrete memoryless channels,  $(\mathcal{X}, \{p_s(y|x) : s \in \mathcal{S}\}, \mathcal{Y})$  where  $y \in \mathcal{Y}$  for every state  $s$  in the set  $\mathcal{S}$ . The compound channel can be looked at as a discrete memoryless channel with the state where the state is arbitrarily selected and fixed for the transmission of entire block. The assumption is that the decoder knows about the channel state. The authors in [5] restricted their attention only on binary input symmetric discrete memoryless channels. Here we present the universal polar coding schemes for binary input asymmetric channels that are a combination of schemes presented in [5] and [6] to achieve rates less than compound capacity. In Section II we will define some notations and preliminaries. In Section III we present the integrated scheme for achieving the capacity for binary input asymmetric channels. In Section IV and V we present the two universal polar coding schemes for binary input asymmetric channels, one based on polarizing indices and another one using stair-case construction respectively.

## II. NOTATIONS AND PRELIMINARIES

In this paper we always consider the binary input asymmetric channels. We denote input alphabet by  $\mathcal{X} = \{0, 1\}$  and output alphabet by  $\mathcal{Y}$ . We express any set of random variables  $X_i, X_{i+1}, \dots, X_j$  ( $i < j$ ) by a row vector  $(X_i, X_{i+1}, \dots, X_j)$  in short hand as  $X^{i:j}$ . We denote the set  $[N] = \{1, 2, 3, \dots, N\}$ . For row vector  $U^{1:N}$  and for  $\mathcal{A} \subset [N]$ , we denote  $U^{\mathcal{A}}$  as row vector of dimension  $1 \times |\mathcal{A}|$  which has only elements of  $U^{1:N}$  whose position  $i \in \mathcal{A}$  in same order from left to right. We use w.p for "with probability".

Let  $\mathcal{S} = \{1, 2, \dots, s\}$  be set of DMCs which are characterized by  $p_l(y|x)$  for each  $l \in \mathcal{S}$ . Let  $(X_1, Y_{l1}), (X_2, Y_{l2}), \dots, (X_N, Y_{lN})$  be IID random tuples distributed according to  $p(x)p_l(y|x)$ , where  $N = 2^n$ . Let  $G_N$  is the conventional polar transform which is 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  $\mathbb{P}(U_i = u_i | U^{1:i-1} Y_l^{1:N} = u^{1:i-1} y_l^{1:N})$  by  $P_{U_i | U^{1:i-1} Y_l^{1:N}}(u_i | u^{1:i-1} y_l^{1:N})$  for the measure induced on

this set up. If two random variables  $(X, Y_l)$  distributed as  $P_X(x)p_l(y|x)$ , bhattacharya parameter is defined as

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

From [2] we have that

$$Z(X|Y_l)^2 \leq H(X|Y_l) \quad (1)$$

$$H(X|Y_l) \leq \log(1 + Z(X|Y_l)) \leq Z(X|Y_l) \quad (2)$$

Let  $\beta < 0.5$  define the following sets,

$$\mathcal{H}_X = \{i \in [N] : Z(U_i|U^{1:(i-1)}) \geq 1 - 2^{-N^\beta}\}$$

$$\mathcal{L}_X = \{i \in [N] : Z(U_i|U^{1:(i-1)}) \leq 2^{-N^\beta}\}$$

$$\mathcal{H}_{X|Y_l} = \{i \in [N] : Z(U_i|U^{1:(i-1)}Y_l^{1:N}) \geq 1 - 2^{-N^\beta}\}$$

$$\mathcal{L}_{X|Y_l} = \{i \in [N] : Z(U_i|U^{1:(i-1)}Y_l^{1:N}) \leq 2^{-N^\beta}\}$$

We have the following from theorem 1 of [6].

$$\lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{H}_X| = H(X), \quad \lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{L}_X| = 1 - H(X)$$

$$\lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{H}_{X|Y_l}| = H(X|Y_l)$$

$$\lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{L}_{X|Y_l}| = 1 - H(X|Y_l)$$

The capacity of compound channel is defined as

$$C_c = \max_{p(x)} \min_{l \in S} I(X; Y_l)$$

where  $(X, Y_l)$  distributed as  $p(x)p_l(y|x)$ .

**Lemma 1.**  $\lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{H}_X \cap \mathcal{L}_{X|Y_l}| = I(X; Y_l)$

**Proof:** Refer to appendix

### III. POLAR CODE FOR ASYMMETRIC CHANNELS

Let the asymmetric DMC is characterized by  $p(y|x)$  and  $p(x)$  is capacity achieving input distribution. We use same notions as in section II except for  $Y$  instead of  $Y_l$  since it is single channel case.

#### Code construction and encoding:

1. Let  $I = \mathcal{H}_X \cap \mathcal{L}_{X|Y}$  where we provide information bits of message  $M$  uniformly distributed on  $\{0, 1\}^{|I|}$  i.e  $U^I = M$  as part of the encoding the message  $M$ .

2. Let  $F = \mathcal{H}_X \cap \mathcal{L}_{X|Y}^c$ . As part of code construction generate the function  $f_r : F \rightarrow \{0, 1\}$  where  $f_r(i)$  is generated independently and uniformly for each  $i \in F$ . We share this between encoder and decoder. We provide frozen bits in  $F$  i.e  $U_i = f_r(i)$  while encoding these channel indices.

3. Let  $D = \mathcal{L}_X$ . We call them as deterministic channels. We use the following **argmax rule** for encoding these channel indices.

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

4. For  $i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c$  as part of code construction generate function  $\lambda_i : \{0, 1\}^{i-1} \rightarrow \{0, 1\}$  randomly. For every  $i$ ,  $u^{1:i-1}$  independently produce

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

Let the set of random functions be denoted as  $\lambda_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}$ . We share it between encoder and decoder. We assign  $U_i = \lambda_i(U^{1:i-1})$  while encoding. Transmit  $X^{1:N} = U^{1:N} G_N$ .

**Decoding:** ( $Y^{1:N}$  is the received vector)

**for**  $i = 1 : N$

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

2. If  $i \in \mathcal{L}_X \cup I$

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

3. If  $i \in (\mathcal{L}_X \cup \mathcal{H}_X)^c$ ,  $\hat{U}_i = \lambda_i(\hat{u}^{1:i-1})$ .

**end**

Decode  $\hat{M} = \hat{U}^I$  (Estimate of message  $M$ )

**Theorem 1.** Let  $M$  be the message chosen uniformly from  $\{0, 1\}^{|I|}$ . Then for the above coding scheme avg probability of error  $\mathbb{E}_{(\lambda_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, f_r)}[P_e(\lambda_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, f_r)] = O(2^{-N^{\beta'}})$  for any  $\beta' < \beta < \frac{1}{2}$ .

**Proof:** Refer to appendix

The code generated whose  $P_e(\lambda_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, f_r) = O(2^{-N^{\beta'}})$  is capacity achieving by Lemma 1.

$P_{U_i|U^{1:i-1}}(u|u^{1:i-1})$  which has to be computed while encoding,  $P_{U_i|U^{1:i-1}, Y^{1:N}}(u|\hat{u}^{1:i-1}, y^{1:N})$  while decoding will be done in  $O(N \log N)$  complexity using techniques in [6].

### IV. SCHEME BASED ON POLARIZING

Let  $p(x)$  be the compound capacity achieving distribution for compound channel  $S$ . The synthetic channel indices formed by polarization for channel  $l$  are categorized as  $D_l = \mathcal{L}_X$ ,  $F_l = \mathcal{H}_X \cap \mathcal{L}_{X|Y_l}^c$  and  $I_l = \mathcal{H}_X \cap \mathcal{L}_{X|Y_l}$  for  $l \in S$  except for a diminishing fraction with respect to block length as stated in the earlier section. We call indices in  $I_l$  as good channels and  $F_l$  as bad channels. Let  $\min\{|I_1|, |I_2|, \dots, |I_s|\} = L$ . Clearly,  $\lim_{N \rightarrow \infty} \frac{L}{N} = C_c$ . If the inequality  $|I_1 \cap I_2 \cap \dots \cap I_s| \leq L$  is strict, it means that some synthetic channels which are good for a channel  $l \in S$  may not be good for all channels in  $S$ . By providing information bits in  $I_1 \cap I_2 \cap \dots \cap I_s$  and randomly chosen frozen bits in  $\mathcal{H}_X - (I_1 \cap I_2 \cap \dots \cap I_s)$  and rest following the same coding scheme as in section III we can get a reliable code.

Note that  $F_l \cup I_l = \mathcal{H}_X$  for all  $l \in S$ . This implies that for any channels  $l, m \in S$ , a synthetic channel which is good for channel  $l$  and not good for channel  $m$  should be a bad channel for channel  $m$ . Now this enables us to adapt universal coding scheme for symmetric channel from [5] and apply for asymmetric case.

The following section demonstrates how to construct codes that achieve rates close to  $\frac{L}{N}$ . If  $\lim_{N \rightarrow \infty} \frac{L}{N} = \frac{I_1 \cap I_2 \cap \dots \cap I_s}{N} = C(S)$  exists as in binary input symmetric channel case, then we can define the gap to compound capacity  $C_c$  as  $\Delta = C_c - C(S)$ .

**Polar coding scheme to reduce the gap:**

Consider the two channel case i.e  $S = \{1, 2\}$ . Let  $G =$

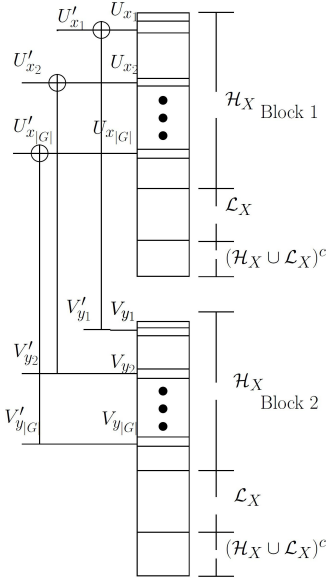


Fig. 1. Universal polarization based on polarizing channel indices

$\min\{|I_1 \cap F_2|, |I_2 \cap F_1|\}$ . Consider the sets  $\mathcal{X}$  and  $\mathcal{Y}$  to be first  $G$  indices in  $I_1 \cap F_2$  and  $I_2 \cap F_1$  respectively. Let  $\mathcal{X} = \{x_1, x_2, \dots, x_G\}$  where  $x_1 < x_2 < \dots < x_G$  and  $\mathcal{Y} = \{y_1, y_2, \dots, y_G\}$  where  $y_1 < y_2 < \dots < y_G$ . Consider two polar blocks 1 and 2 of length  $N$ . That is  $X^{1:2N}$  is IID distributed and we will have

$$U^{1:N} = X^{1:N} G_N, \quad V^{1:N} = X^{N+1:2N} G_N \quad (3)$$

for block 1 and block 2 respectively. Now we polarize by combining index  $x_j$  of block 1 with  $y_j$  of block 2 which produces two new variables for each  $j \in [G]$ .

$$U'_{x_j} = U_{x_j} + V_{y_j}, \quad V'_{y_j} = V_{y_j}$$

For remaining  $i$  which is not either  $x_j$  for block 1 or  $y_j$  for block 2,  $U'_i = U_i$  and  $V'_i = V_i$  respectively.

**Lemma 2.** For each  $j \in [G]$ , for any  $\beta' < \beta < 0.5$ , for large enough  $N$

1.  $Z(U'_{x_j} | U^{1:x_j-1} V^{1:y_j-1}) \geq 1 - 2^{-N^{\beta'}}$   
( $U'_{x_j}$  is almost random given  $U^{1:x_j-1} V^{1:y_j-1}$ )
2.  $Z(V'_{y_j} | U^{1:x_j-1} V^{1:y_j-1} U'_{x_j}) \geq 1 - 2^{-N^{\beta'}}$   
( $V'_{y_j}$  is almost random given  $U^{1:x_j-1} U'_{x_j} V^{1:y_j-1}$ )
3.  $Z(U'_{x_j} | U^{1:x_j-1} V^{1:y_j-1} Y_l^{1:2N}) \geq 1 - 2^{-N^{\beta'}}$   
( $U'_{x_j}$  is almost random given  $U^{1:x_j-1} V^{1:y_j-1} Y_l^{1:2N}$  for both  $l = 1, 2$ )
4.  $Z(V'_{y_j} | U^{1:x_j-1} V^{1:y_j-1} U'_{x_j} Y_l^{1:2N}) \leq 2^{-N^{\beta'}}$   
( $V'_{y_j}$  is almost deterministic given  $U^{1:x_j-1} U'_{x_j} V^{1:N+y_j-1} Y_l^{1:2N}$  for both  $l = 1, 2$ )

**Proof:** Refer to appendix

Hence by above lemma with this polarizing step, the indices in  $\mathcal{X}$ , become bad for both the channels 1 and 2 in block 1. The indices in  $\mathcal{Y}$ , become good for both the channels 1 and 2 in block 2. Hence we have  $|I_1 \cap I_2| + G$  good indices with

respect to  $2N$  block length which is an improvement. This is the key idea of the universalization.

#### Code construction and encoding:

1. Let the message  $M$  uniformly distributed on  $\{0, 1\}^{2|I_1 \cap I_2| + |\mathcal{Y}|}$ . Partition  $M$  into  $M_1$  and  $M_2$  such that  $M_1$  takes first  $|I_1 \cap I_2|$  bits of  $M$  and  $M_2$  takes later  $|I_1 \cap I_2| + |\mathcal{Y}|$  of  $M$ . Set  $U'^{I_1 \cap I_2} = M_1$  and  $V'^{(I_1 \cap I_2) \cup \mathcal{Y}} = M_2$  as part of encoding.
2. As part of code construction, generate function  $f_{r_1} : \mathcal{H}_X - (I_1 \cap I_2) \rightarrow \{0, 1\}$  where  $f_{r_1}(i)$  is generated independently and uniformly for each  $i$  and generate function  $f_{r_2} : \mathcal{H}_X - ((I_1 \cap I_2) \cup \mathcal{Y}) \rightarrow \{0, 1\}$  where  $f_{r_2}(i)$  is generated independently and uniformly for each  $i$ . So we will have  $U'_i = f_{r_1}(i)$  for  $i \in \mathcal{H}_X - ((I_1 \cap I_2))$  and  $V'_i = f_{r_2}(i)$  for  $i \in \mathcal{H}_X - ((I_1 \cap I_2) \cup \mathcal{Y})$  for encoding these channel indices in block1 and block2 respectively.
3. Now  $U_{x_j} = U'_{x_j} + V'_{y_j}$  and  $V_{y_j} = V'_{y_j}$  for  $j \in [G]$ .  $U_i = U'_i$  and  $V_i = V'_i$  for other  $i \in \mathcal{H}_X$
4. For  $i \in \mathcal{L}_X$  for block 1 and 2, we use the following argmax rule for encoding these channel indices.

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

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

5. As part of code construction for each block  $b \in \{1, 2\}$  independently, for indices  $i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c$  we construct set of random functions  $\lambda^b_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}$ . Each function in the set is  $\lambda^b_i : \{0, 1\}^{i-1} \rightarrow \{0, 1\}$  is a random function. For every  $i$ ,  $u^{1:i-1}$  independently produce

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

We assign  $U_i = \lambda^b_i(U^{1:i-1})$  and  $V_i = \lambda^b_i(V^{1:i-1})$  for encoding these indices. For  $i \notin \mathcal{H}_X$ , take  $U'_i = U_i$  and  $V'_i = V_i$ . Transmit  $X^{1:2N} = U^{1:N} G_N$  and  $X^{N+1:2N} = V^{1:N} G_N$ . Let  $Y^{1:2N}$  is received vector.

#### Decoding when channel $l$ is selected:

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

2.

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

if  $i \in \mathcal{H}_X - (I_1 \cap I_2)$ ,  $\hat{U}'_i = \hat{U}_i = f_{r_1}(i)$

if  $i \in (I_1 \cap I_2) \cup \mathcal{L}_X$ ,

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

if  $i \in (\mathcal{H}_X \cap \mathcal{L}_X)^c$ ,  $\hat{U}'_i = \hat{U}_i = \lambda^1_i(\hat{U}^{1:N-1})$

**end**

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

if  $i \in \mathcal{H}_X - (I_1 \cap I_2)$ ,  $\hat{V}'_i = \hat{V}_i = f_{r_2}(i)$

if  $i \in (I_1 \cap I_2) \cup \mathcal{L}_X$ ,

$$\hat{V}'_i = \hat{V}_i = \mathbb{1}\{P_{V_i | V^{1:i-1}, Y_l^{1:N}}(1 | \hat{V}^{1:i-1}, Y^{N+1:2N}) \geq P_{V_i | V^{1:i-1}, Y_l^{1:N}}(0 | \hat{V}^{1:i-1}, Y^{N+1:2N})\}$$

if  $i \in (\mathcal{H}_X \cap \mathcal{L}_X)^c$ ,  $\hat{V}'_i = \hat{V}_i = \lambda^2_i(\hat{V}^{1:N-1})$

**end**

3.

$$(a) \hat{U}'_{x_j} = f_{r_1}(x_j)$$

$$(b) \hat{V}'_{y_j} =$$

$1\{P_{V'_{y_j}|U^{1:x_j-1}U'_{x_j}V^{1:y_j-1}Y^{1:2N}}(1|\hat{U}^{1:x_j-1}U'_{x_j}\hat{V}^{1:y_j-1}Y^{1:2N})$   
 $\geq P_{V'_{y_j}|U^{1:x_j-1}U'_{x_j}V^{1:y_j-1}Y^{1:2N}}(0|\hat{U}^{1:x_j-1}\hat{U}'_{x_j}\hat{V}^{1:y_j-1}Y^{1:2N})\}$   
(c)  $\hat{U}_{x_j} = \hat{U}'_{x_j} + \hat{V}'_{y_j}$  and  $\hat{V}_{y_j} = \hat{V}'_{y_j}$   
4. Repeat step 2 and 3 for  $j = \{2, 3, \dots, G\}$   
5. Set  $\hat{M}_1 = \hat{U}'_{I_1 \cap I_2}$  and  $\hat{M}_2 = \hat{V}'_{(I_1 \cap I_2) \cup \mathcal{Y}}$ . Combine  $\hat{M}_1, \hat{M}_2$  to get  $\hat{M}$ .

**Theorem 2.** The above construction for sufficiently large block length  $N$ , encoding and decoding rule achieves average probability of error  $\mathbb{E}[P_e(\lambda_1, \lambda_2, f_{r_1}, f_{r_2})] = O(2^{-N^{\beta'}})$  for any  $\beta' < \beta < 0.5$ .

**Proof:** Refer to appendix

**Remarks:**

1. For large enough  $N$ , we can recurse the above construction  $k$  times which takes  $2^k$  blocks as mentioned in [5]. In  $k$ th step we take two hybrid polar blocks produced after  $k-1$  steps and perform the above polarizing idea. After  $k$  steps, we will have  $2^k L - G$  good indices, the gap between fraction of number of good indices for both the channels with respect to blocklength and  $\frac{L}{N}$  becomes  $\frac{G}{2^k N}$  and average probability of error can be proved to  $\mathbb{E}[P_e] = 2^k O(2^{-N^{\beta'}})$  using similar ideas in the proof of theorem 2.
2. We can generalize to case when  $S = \{1, 2, \dots, s\}$  by similar arguments used in [5]. To recall, we first universalize by considering channels 1 and 2 in  $S$  like mentioned above step and produce a universalized polar block for these two channels. Next we consider the universalized polar block produced for channel 1 and 2 and perform similar universalization with channel 3. We repeat this until we finish all channels in  $S$ . The analysis for relationship between blocklength and rate is same as in [5].

## V. SCHEME BASED ON STAIR CASING BLOCKS

For large enough  $N$  blocklength we assume that,  $|I_1 \cap I_2 \cap \dots \cap I_s| \geq |(\mathcal{H}_X \cup \mathcal{L}_X)^c|$ . This property is used in the coding scheme and using this we avoid sharing pseudo random boolean functions and also avoids side channel techniques mentioned in [4].

Take  $I'$  to be subset of  $I_1 \cap I_2 \cap \dots \cap I_s$  such that  $|I'| = |(\mathcal{H}_X \cup \mathcal{L}_X)^c|$ . Let us have one to one correspondence between the set of indices  $I'$  and  $(\mathcal{H}_X \cup \mathcal{L}_X)^c$  by defining an arbitrary bijection from  $g : I' \rightarrow (\mathcal{H}_X \cup \mathcal{L}_X)^c$ . Let  $L' = \min\{|I_1|, |I_2|, \dots, |I_s|\} - |I'|$ .

**Code construction:**

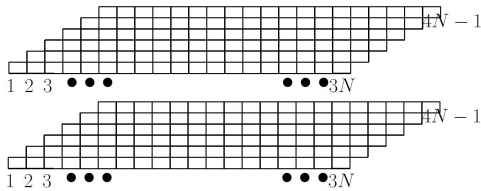


Fig. 2. Extended stair cases with  $k=3$ ,  $N=6$  and  $q=2$

Generate random frozen vector  $W^{1:N}$  such that  $W^{1:N} = u^{1:N}$  with probability  $2^{-|\mathcal{H}_X|} \prod_{i \in \mathcal{L}_X} q(u_i | u^{1:i-1}) \prod_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1})$   
 $W^{1:N}$  is shared between encoder and decoder

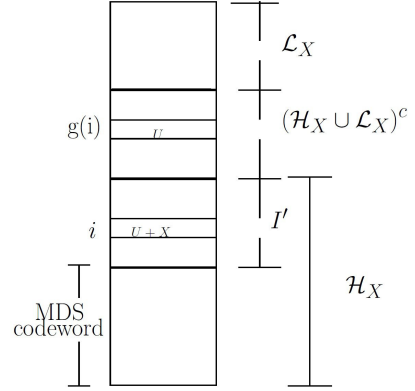


Fig. 3. Full column structure where  $X$  is parity of the information bits of the column and  $U$  is encoded bit of block with index  $g(i)$

We place polar blocks of size  $N$  large enough in a stair case manner. The height of the stair case is kept as  $N$ . We keep  $k$  such stair cases side by side shown as below. Let us assume that there is linear MDS code  $\mathcal{M}$  with block length  $|\mathcal{H}_X| - |(\mathcal{H}_X \cup \mathcal{L}_X)^c|$  over field  $GF(2^q)$  for some natural number  $q$ . Now we repeat the above stair case structure  $q$  times one above other and label each stair case from  $1 : q$ . This is shown in figure 2. We fill such polar blocks column by column from left to right of the stair case structure while encoding and we follow the same for decoding. The total no. of columns are  $(k+1)N-1$ .

**Encoding:**

**Input:**  $qL'$  information bits for each full column.

**Output:**  $U^{1:N}$  of each polar block in the staircase.

- To encode non full columns on the left from  $t = 1 : N-1$ , we assign  $U_i = W_i$  for block with channel index  $i$  in that column. Repeat this for all  $q$  stair cases.
- To encode full columns from  $t = N \leq i \leq kN$ 
  - First, encode the blocks with index  $i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c$  in column  $t$  with randomized rounding rule i.e  $U_i = d$  w.p  $P_{U_i | U^{1:i-1}}(d | U^{1:i-1})$  for  $d \in \{0, 1\}$ . Repeat this for all  $q$  stair cases.
  - Second, encode the blocks with index  $i \in \mathcal{L}_X$  in column  $t$  with argmax rule. Repeat this for all  $q$  stair cases.
  - Third, encode the blocks with indices in  $\mathcal{H}_X - I'$ 
    - \* Encode  $qL'$  information bits which is equivalent to  $L'$  symbols over  $GF(2^q)$  into codeword  $m$  (binary format) in  $\mathcal{M}$ . Replace the zero bits of  $m$  in those positions which have zeros in all codewords of  $\mathcal{M}$  with parity ( $X$ ) of  $qL'$  information bits. This is to maintain the distribution in high entropy indices.
    - \* Fill blocks whose indices in  $\mathcal{H}_X - I'$  in all  $q$  staircases with the modified codeword  $m$  in binary

format as shown in the figure 3.

- Fourth, we encode the blocks with index  $i$  in  $I'$  by assigning  $U_i = X \oplus U'_{g(i)}$  where  $U'_{g(i)}$  is already encoded bit (first step) in the same column corresponding to the block with index  $g(i) \in (\mathcal{H}_X \cup \mathcal{L}_X)^c$ . We do the same for all  $q$  staircases. This is to maintain the distribution of the high entropy indices and also ensures the independence of previously encoded bits in the block and key step for the construction to work.
  - To encode non full columns  $t = kN + 1 : (k + 1)N - 1$  on the right, we generate everything randomly to satisfy the distribution of polarblock.
    - For blocks with index  $i \in \mathcal{H}_X$  generate  $U_i$  independently and uniformly
    - For blocks with index  $i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c$ . Generate  $U_i = d$  w.p  $P_{U_i|U^{1:i-1}}(d|U^{1:i-1})$   $d \in \{0, 1\}$
    - For blocks with index  $i \in \mathcal{L}_X$  use argmax rule.
- Transmit  $X^{1:N} = U^{1:N}G_N$  for each polar block.

#### Decoding:

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

**Output:** Estimates of encoded information bits.

- To decode non full columns on the left from  $t = 1 : N - 1$ , we estimate  $\hat{U}_i = W_i$  for the block with channel index  $i$  in that column. Repeat this for all  $q$  stair cases.
- To decode full columns from  $t = N \leq i \leq kN$ 
  - First, decode the blocks with index  $i \in \mathcal{L}_X \cup I'$  in column using standard SC decoding step. It is possible since these indices are either good for all channels or deterministic.
  - Second, decode the blocks with index in  $\mathcal{H}_X - I'$ 
    - \* Decode the  $L'$  symbols from the good indices based on the channel that is selected from all the  $q$  staircases. Let partially recovered codeword be  $c$
    - \* Replace by zero in those positions which have zeros in all codewords of  $\mathcal{M}$  in partially recovered codeword  $c$ . Now modified  $c$  is codeword of  $\mathcal{M}$  with erasures.
    - \* The codeword  $\hat{m}$  can be recovered from  $c$  since it is MDS codeword. This estimates  $qL'$  information bits corresponding to the column. Let  $\hat{X}$  be parity of decoded information bits of the column.
    - \* Replace the zero bits of  $\hat{m}$  in those positions which have zeros in all codewords of  $\mathcal{M}$  with  $\hat{X}$ . Modified  $\hat{m}$  estimates  $\hat{U}_i$  for these indices.
  - Last, decode blocks with index  $i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c$  by estimating  $\hat{U}_i = \hat{X} \oplus \hat{U}'_{g^{-1}(i)}$  where  $\hat{U}'_{g^{-1}(i)}$  is already decoded bit corresponding to block of index  $g^{-1}(i) \in I'$  of the same column.
- Ignore and do not decode non full columns  $t = kN + 1 : (k + 1)N - 1$  on the right.

Now we get an upper bound on  $q$  to upper bound on

block length. If we consider RS code for linear MDS code over  $GF(2^q)$ , block length should divide  $2^q - 1$ . If  $(|\mathcal{H}_X| - |(\mathcal{H}_X \cup \mathcal{L}_X)^c|)$  is odd, by euler theorem,  $q$  can take value  $\phi(|\mathcal{H}_X| - |(\mathcal{H}_X \cup \mathcal{L}_X)^c|)$  where  $\phi$  is Euler's number.  $\phi(|\mathcal{H}_X| - |(\mathcal{H}_X \cup \mathcal{L}_X)^c|) \leq |\mathcal{H}_X| - |(\mathcal{H}_X \cup \mathcal{L}_X)^c| \leq N$ . We can consider  $(|\mathcal{H}_X| - |(\mathcal{H}_X \cup \mathcal{L}_X)^c|) - 1$  as block length of RS code if  $(|\mathcal{H}_X| - |(\mathcal{H}_X \cup \mathcal{L}_X)^c|)$  is even and we can code with parity of information bits in the other position.

**Theorem 3.** 1. For every block in the stair case

$$\mathbb{E}_{W^{1:N}}[\mathbb{P}(U^{1:N} = u^{1:N}, Y^{1:N} = y^{1:N} | W^{1:N})] \\ = 2^{-(|\mathcal{H}_X|)} \prod_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} P_{U_i|U^{1:i-1}}(u_i | u^{1:i-1}) \\ \prod_{i \in \mathcal{L}_X} q(u_i | u^{1:i-1}) P_{Y^{1:N}|U^{1:N}}(y^{1:N} | u^{1:N}).$$

2. The avg probability of error:

$$\mathbb{E}_{W^{1:N}}[P_e(W^{1:N})] \leq (N^3.k.O(2^{-N^\beta})) \text{ for } \beta < 0.5$$

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

**Proof:** Refer to appendix

**Remarks:**

1. We used linear MDS code to satisfy the distribution for each block. In symmetric case, linearity is not required.
2. Relation between rate  $R$  and  $k$  is same as in [5].
3. If the assumption  $|I_1 \cap I_2 \cap \dots \cap I_s| \geq |(\mathcal{H}_X \cup \mathcal{L}_X)^c|$  does not hold, we can use the universalizing technique of section IV with fewer recursions to produce partially universalized block to get the desired condition and use this scheme on the partially universalized block.

## VI. CONCLUSION

We proved that argmax rule can be used for encoding deterministic channels. We adapted two universal polar coding schemes to binary input asymmetric case using integrated scheme. We gave rigorous proof for asymptotic probability of error analysis for each scheme. We exploited stair case structure for scheme in section V and avoids side-channel for transmitting bits corresponding to synthetic channels which are not polarized to required level. Drawback of such schemes is large block-length. So open question is to design schemes which take short blocklength.

## VII. ACKNOWLEDGEMENT

This work is supported by NSF grant CCF-1116739 and CCF-1116820.

## REFERENCES

- [1] E. Arıkan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7):3051–3073, July 2009.
- [2] E. Arıkan. Source polarization. In *2010 IEEE International Symposium on Information Theory*, pages 899–903, June 2010.
- [3] R. A. Chou and M. R. Bloch. Using deterministic decisions for low-entropy bits in the encoding and decoding of polar codes. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1380–1385, Sept 2015.
- [4] 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 Transactions on Information Theory*, 62(7):4024–4038, July 2016.

- [5] S. H. Hassani and R. Urbanke. Universal polar codes. In *2014 IEEE International Symposium on Information Theory*, pages 1451–1455, June 2014.
- [6] J. Honda and H. Yamamoto. Polar coding without alphabet extension for asymmetric models. *IEEE Transactions on Information Theory*, 59(12):7829–7838, Dec 2013.
- [7] M. Mondelli, R. Urbanke, and S. H. Hassani. How to achieve the capacity of asymmetric channels. In *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 789–796, Sept 2014.
- [8] V. Taranalli, H. Uchikawa, and P. H. Siegel. On the capacity of the beta-binomial channel model for multi-level cell flash memories. *IEEE Journal on Selected Areas in Communications*, 34(9):2312–2324, Sept 2016.

## APPENDIX A PROOF OF THEOREM 1

For the decoder mentioned in section III, let  $\mathcal{E}_i$  is the error event for  $i$  th bit channel. The codeword  $u^{1:N}$  and received vector  $y^{1:N}$  belongs to  $\mathcal{E}_i$  then we have

$$\begin{aligned} P_{U_i|U^{1:i-1}Y^{1:N}}(u_i + 1|u^{1:i-1}y^{1:N}) \\ \geq P_{U_i|U^{1:i-1}Y^{1:N}}(u_i|u^{1:i-1}y^{1:N}) \end{aligned} \quad (4)$$

For every  $u^{1:i-1}$  define  $q(y|u^{1:i-1}) = 1$  and  $q(y+1|u^{1:i-1}) = 0$  where

$$y = \operatorname{argmax}_{x \in \{0,1\}} P_{U_i|U^{1:i-1}}(x|u^{1:i-1}) \quad (5)$$

Let  $\mathcal{E}$  be the error event. Hence  $\mathcal{E} = \cup_{i \in \mathcal{L}_X|Y} \mathcal{E}_i$ . The the probability of the error for given code that is for a particular random function set  $\lambda_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}$  and randomly generated frozen bit function  $f_r$ .

$$\begin{aligned} P_e(\lambda_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, f_r) \\ = \sum_{(u^{1:N}, y^{1:N})} 2^{-|I|} 1[\cap_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} (\lambda_i(u^{1:i-1}) = u_i)] \\ \prod_{i \in \mathcal{L}_X} q(u_i|u^{1:i-1}) 1[\cap_{i \in F} (u_i = f_r(i))] P_{Y^{1:N}|U^{1:N}}(y^{1:N}|u^{1:N}) 1[(u^{1:N}, y^{1:N}) \in \mathcal{E}] \end{aligned} \quad (6)$$

By independence of  $\lambda_i(u^{1:i-1})$  and frozen bit function  $f_r(i)$  for each  $i$ , we get

$$\begin{aligned} \mathbb{E}_{(\lambda_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, f_r)} [P_e(\lambda_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, f_r)] \\ = \sum_{(u^{1:N}, y^{1:N})} 2^{-(|I|+|F|)} \prod_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1}) \prod_{i \in \mathcal{L}_X} q(u_i|u^{1:i-1}) P_{Y^{1:N}|U^{1:N}}(y^{1:N}|u^{1:N}) 1[(u^{1:N}, y^{1:N}) \in \mathcal{E}] \end{aligned} \quad (7)$$

Let us define the probability distribution  $Q_{U^{1:N}, Y^{1:N}}$  as follows

$$\begin{aligned} Q_{U^{1:N}, Y^{1:N}}(u^{1:N}, y^{1:N}) \\ = 2^{-(|I|+|F|)} \prod_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1}) \prod_{i \in \mathcal{L}_X} q(u_i|u^{1:i-1}) P_{Y^{1:N}|U^{1:N}}(y^{1:N}|u^{1:N}) \end{aligned} \quad (8)$$

This implies that

$$\begin{aligned} \mathbb{E}_{(\lambda_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, f_r)} [P_e(\lambda_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, f_r)] \\ = Q_{U^{1:N}, Y^{1:N}}(\mathcal{E}) \\ \leq \|Q_{U^{1:N}, Y^{1:N}} - P_{U^{1:N}, Y^{1:N}}\| + P_{U^{1:N}, Y^{1:N}}(\mathcal{E}) \\ \leq \|Q_{U^{1:N}, Y^{1:N}} - P_{U^{1:N}, Y^{1:N}}\| + \sum_{i=1}^N P_{U^{1:N}, Y^{1:N}}(\mathcal{E}_i) \end{aligned} \quad (9)$$

From equation (60) of [6] we have  $P_{U^{1:N}, Y^{1:N}}(\mathcal{E}) \leq O(2^{-N^\beta})$ . Following lemma proves the total variation distance between the measure  $Q_{U^{1:N}, Y^{1:N}}$  and  $P_{U^{1:N}, Y^{1:N}}$  is  $O(2^{-N^{\beta'}})$  which completes the proof of theorem.

**Lemma 3.** For any  $\beta' < \beta < 0.5$  for large enough polar block  $N$ ,

$$\|P_{U^{1:N}, Y^{1:N}} - Q_{U^{1:N}, Y^{1:N}}\| = O(2^{-N^{\beta'}}) \quad (10)$$

**Proof:**

$$B_1^N - A_1^N = \sum_{i=1}^N A_1^{i-1} B_i^N - \sum_{i=1}^N A_1^i B_{i+1}^N \quad (11)$$

$$= \sum_{i=1}^N (B_i - A_i) A_1^{i-1} B_{i+1}^N \quad (12)$$

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. For convenience, we omit the symbols of random variables in suffix for expressing probability of events in their respective measures. For example we express  $P(u_i|u^{1:i-1})$  and  $Q(u^{1:N}, y^{1:N})$  for  $P_{U_i|U^{1:i-1}}(u_i|u^{1:i-1})$  and  $Q_{U^{1:N}, Y^{1:N}}(u^{1:N}, y^{1:N})$  respectively.

$$\begin{aligned} 2\|P_{U^{1:N}, Y^{1:N}} - Q_{U^{1:N}, Y^{1:N}}\| \\ = \sum_{u^{1:N}, y^{1:N}} |Q(u^{1:N}, y^{1:N}) - P(u^{1:N}, y^{1:N})| \\ \stackrel{(a)}{=} \sum_{u^{1:N}, y^{1:N}} |Q(u^{1:N}) - P(u^{1:N})| |P(y^{1:N}|u^{1:N})| \\ \stackrel{(b)}{=} \sum_{u^{1:N}} |Q(u^{1:N}) - P(u^{1:N})| \\ \stackrel{(c)}{=} \sum_{u^{1:N}} \left| \prod_{i=1}^n Q(u_i|u^{1:i-1}) - \prod_{i=1}^N P(u_i|u^{1:i-1}) \right| \\ \stackrel{(d)}{=} \sum_{u^{1:N}} \left| \sum_{i=1}^N ((Q(u_i|u^{1:i-1}) - P(u_i|u^{1:i-1})) \right. \\ \left. \prod_{j=1}^{i-1} P(u_j|u_1^{j-1}) \prod_{j=i+1}^N Q(u_j|u_1^{j-1}) \right| \\ \stackrel{(e)}{\leq} \sum_{u^{1:N}} \sum_{i=1}^N (|Q(u_i|u^{1:i-1}) - P(u_i|u^{1:i-1})| P(u^{1:i-1}) \\ \prod_{j=i+1}^N Q(u_j|u_1^{j-1})) \end{aligned}$$

$$\begin{aligned}
& \stackrel{(f)}{=} \sum_{i=1}^N \sum_{u^{1:i-1}} (|Q(u_i|u^{1:i-1}) - P(u_i|u^{1:i-1})| P(u^{1:i-1})) \\
& \quad \prod_{j=i+1}^N Q(u_j|u^{1:j-1}) \\
& = \sum_{i=1}^N \sum_{u^{1:i-1}} \sum_{u_i} (|Q(u_i|u^{1:i-1}) - P(u_i|u^{1:i-1})| P(u^{1:i-1})) \\
& \quad \sum_{u_{i+1}^N} \prod_{j=i+1}^N Q(u_j|u_1^{j-1})
\end{aligned}$$

$$\begin{aligned}
& = \sum_{i=1}^N \sum_{u^{1:i-1}} \sum_{u_i} (|Q(u_i|u^{1:i-1}) - P(u_i|u^{1:i-1})| \\
& \quad P(u^{1:i-1}) \sum_{u_{i+1}^N} Q(u_{i+1}^N|u_1^i)) \\
& = \sum_{i=1}^N \sum_{u^{1:i-1}} 2P(u^{1:i-1}) \\
& \quad ||Q_{U_i|U^{1:i-1}=u^{1:i-1}} - P_{U_i|U^{1:i-1}=u^{1:i-1}}|| \\
& \stackrel{(g)}{=} \sum_{i \in F \cup I} \sum_{u^{1:i-1}} 2P(u^{1:i-1}) \\
& \quad ||Q_{U_i|U^{1:i-1}=u^{1:i-1}} - P_{U_i|U^{1:i-1}=u^{1:i-1}}|| \\
& \quad + \sum_{i \in D} \sum_{u^{1:i-1}} 2P(u^{1:i-1}) \\
& \quad ||Q_{U_i|U^{1:i-1}=u^{1:i-1}} - P_{U_i|U^{1:i-1}=u^{1:i-1}}||
\end{aligned}$$

(a) is true since  $P(y^{1:N}|u^{1:N}) = Q(y^{1:N}|u^{1:N})$ , (b) by summing over  $y^{1:N}$ , (c) by conditional probability law, (d) by equation (11), (e) by triangle inequality law, (f) exchanging order of summation, (g) is true because  $Q(u_i|u^{1:i-1}) = P(u_i|u^{1:i-1})$  for  $i \in (F \cup D \cup I)^c$ . Hence we get

$$\begin{aligned}
& \sum_{u^{1:i-1}} 2P(u^{1:i-1}) \\
& \quad ||Q_{U_i|U^{1:i-1}=u^{1:i-1}} - P_{U_i|U^{1:i-1}=u^{1:i-1}}|| \\
& \leq \sum_{i \in F \cup I \cup D} \sum_{u^{1:i-1}} 2P(u^{1:i-1}) \\
& \quad ||Q_{U_i|U^{1:i-1}=u^{1:i-1}} - P_{U_i|U^{1:i-1}=u^{1:i-1}}||
\end{aligned} \tag{13}$$

For  $i \in F \cup I$  evaluate

$$\begin{aligned}
& \sum_{u^{1:i-1}} 2P(u^{1:i-1}) \\
& \quad ||Q_{U_i|U^{1:i-1}=u^{1:i-1}} - P_{U_i|U^{1:i-1}=u^{1:i-1}}|| \\
& \stackrel{(a)}{\leq} \sum_{U^{1:i-1}} P(u^{1:i-1}) \\
& \quad \sqrt{(2 \ln 2) D(P_{U_i|U^{1:i-1}=u^{1:i-1}} || Q_{U_i|U^{1:i-1}=u^{1:i-1}})}
\end{aligned}$$

$$\begin{aligned}
& \stackrel{(b)}{\leq} [(2 \ln 2) \sum_{u^{1:i-1}} P(u^{1:i-1}) \\
& \quad D(P_{U_i|U^{1:i-1}=u^{1:i-1}} || Q_{U_i|U^{1:i-1}=u^{1:i-1}})]^{0.5} \\
& \stackrel{(c)}{\leq} \sqrt{(2 \ln 2)(1 - H(U_i|U^{1:i-1}))} \\
& \stackrel{(d)}{\leq} \sqrt{(2 \ln 2)(1 - Z(U_i|U^{1:i-1}))^2} \\
& \stackrel{(e)}{\leq} \sqrt{(2 \ln 2)(1 - (1 - 2^{-N^\beta})^2)} \\
& = O(2^{-N^{\beta'}})
\end{aligned}$$

(a) by pinsker inequality, (b) by jensen's inequality, (c)  $Q(u_i|u^{1:i-1}) = 0.5$  and by the formula of conditional entropy. (d) from equation (2). (e) for large enough  $N$  by polarization for  $i \in F \cup I$ ,  $Z(U_i|U^{1:i-1}) \geq 1 - 2^{-N^\beta}$ . Hence we have for  $i \in F \cup I$

$$\begin{aligned}
& \sum_{u^{1:i-1}} 2P(u^{1:i-1}) \\
& \quad ||Q_{U_i|U^{1:i-1}=u^{1:i-1}} - P_{U_i|U^{1:i-1}=u^{1:i-1}}|| \\
& \leq O(2^{-N^{\beta'}})
\end{aligned} \tag{14}$$

For  $i \in D$ ,

let  $p_{u^{1:i-1}} = \max\{P(0|u^{1:i-1}), P(1|u^{1:i-1})\}$

$$\begin{aligned}
& \sum_{u^{1:i-1}} 2P(u^{1:i-1}) \\
& \quad ||Q_{U_i|U^{1:i-1}=u^{1:i-1}} - P_{U_i|U^{1:i-1}=u^{1:i-1}}|| \\
& \stackrel{(a)}{\leq} \sum_{u^{1:i-1}} P(u^{1:i-1}) \\
& \quad \sqrt{(2 \ln 2) D(Q_{U_i|U^{1:i-1}=u^{1:i-1}} || P_{U_i|U^{1:i-1}=u^{1:i-1}})} \\
& \stackrel{(b)}{\leq} [(2 \ln 2) \sum_{u^{1:i-1}} P(u^{1:i-1}) \\
& \quad D(Q_{U_i|U^{1:i-1}=u^{1:i-1}} || P_{U_i|U^{1:i-1}=u^{1:i-1}})]^{0.5} \\
& \stackrel{(c)}{\leq} \sqrt{(2 \ln 2) \sum_{u^{1:i-1}} P(u^{1:i-1}) (-\log(p_{u^{1:i-1}}))} \\
& \stackrel{(d)}{\leq} \sqrt{(2 \ln 2) \sum_{u^{1:i-1}} P(u^{1:i-1}) (H(U_i|U^{1:i-1} = u^{1:i-1}))} \\
& = \sqrt{(2 \ln 2) (H(U_i|U^{1:i-1}))} \\
& \stackrel{(e)}{\leq} \sqrt{(2 \ln 2) (Z(U_i|U^{1:i-1}))} \\
& \stackrel{(f)}{\leq} \sqrt{(2 \ln 2) 2^{-N^\beta}} = O(2^{-N^{\beta'}})
\end{aligned}$$

(a) by pinsker inequality, (b) jensen's inequality for concave functions. (c)  $Q(u_i|u^{1:i-1}) = 1$  when  $u_i = \operatorname{argmax}\{P(0|u^{1:i-1}), P(1|u^{1:i-1})\}$ . (d) is true since  $\log(\frac{p_{u^{1:i-1}}}{1-p_{u^{1:i-1}}}) > 0$ . (e) by equation (2), (f) is true since for

large enough  $N$  when  $i \in D$ ,  $Z(U_i|U^{1:i-1}) \leq 2^{-N^\beta}$ . Hence we have for  $i \in D$

$$\begin{aligned} & \sum_{u^{1:i-1}} 2P(u^{1:i-1}) \\ & \quad ||Q_{U_i|U^{1:i-1}=u^{1:i-1}} - P_{U_i|U^{1:i-1}=u^{1:i-1}}|| \\ & \leq O(2^{-N^{\beta'}}) \end{aligned} \quad (15)$$

For equations (13), (14) and (15) we complete the derivation of the lemma.

**Lemma 4.** For any random variables  $X, Y$  and  $Z$  defined over alphabets  $\mathcal{X} = \{0, 1\}$ ,  $\mathcal{Y}$  and  $\mathcal{Z}$  respectively

1.  $Z(X|Y) \leq Z(X)$
2. More generally  $Z(X|YZ) \leq Z(X|Y)$

**Proof:**

$$\begin{aligned} Z(X|Y) &= 2 \sum_{y \in \mathcal{Y}} P_Y(y) \sqrt{P_{X|Y}(0|y)P_{X|Y}(1|y)} \\ &\stackrel{(a)}{\leq} 2 \sqrt{\sum_{y \in \mathcal{Y}} P_Y(y) P_{X|Y}(0|y) P_{X|Y}(1|y)} \\ &= 2 \sqrt{\mathbb{E}_Y[P_{X|Y}(0|Y)P_{X|Y}(1|Y)]} \\ &\stackrel{(b)}{=} 2 \sqrt{\mathbb{E}_Y[P_{X|Y}(0|Y)] - \mathbb{E}_Y[P_{X|Y}^2(0|Y)]} \\ &\stackrel{(c)}{\leq} 2 \sqrt{\mathbb{E}_Y[P_{X|Y}(0|Y)] - (\mathbb{E}_Y[P_{X|Y}(0|Y)])^2} \\ &= 2 \sqrt{P_X(0) - P_X^2(0)} \\ &= 2 \sqrt{P_X(0)P_X(1)} = Z(X) \end{aligned} \quad (16)$$

(a) by concavity of  $\sqrt{x}$  and (b) by linearity of expectation. (c) by convexity of  $x^2$ . Therefore part 1 is proved.

$$\begin{aligned} & Z(X|Y, Z) \\ &= 2 \sum_{(y,z) \in \mathcal{Y} \times \mathcal{Z}} P_{Y,Z}(y, z) \\ & \quad \sqrt{P_{X|Y,Z}(0|y, z)P_{X|Y,Z}(1|y, z)} \\ &= 2 \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{z \in \mathcal{Z}} P_{Z|Y}(z|y) \\ & \quad \sqrt{P_{X|Y,Z}(0|y, z)P_{X|Y,Z}(1|y, z)} \\ &= \sum_{y \in \mathcal{Y}} P_Y(y) Z(X|Z, Y = y) \\ &\stackrel{(a)}{\leq} \sum_{y \in \mathcal{Y}} P_Y(y) Z(X|Y = y) \\ &= 2 \sum_{y \in \mathcal{Y}} P_Y(y) \sqrt{P_{X|Y}(0|y)P_{X|Y}(1|y)} \end{aligned} \quad (17)$$

(a) by part 1. Hence part 2 is proved.

**Lemma 5.** Let  $P_{X,Y}^j(x, y)$  be a joint distribution on  $(X, Y)$  supported on  $\mathcal{X} \times \mathcal{Y}$  for each  $j \in \mathcal{J}$ . Let  $Q(j)$  be the distribution on  $j$ . Define  $P_{X,Y}(x, y) = \sum_{j \in \mathcal{J}} Q(j)P_{X,Y}^j(x, y)$ . Then  $Z(X|Y) \geq \sum_{j \in \mathcal{J}} Q(j)Z^j(X|Y)$  where  $Z^j(X|Y) = 2 \sum_{y \in \mathcal{Y}} \sqrt{P_{X,Y}^j(0, y)P_{X,Y}^j(1, y)}$

**Proof:**

$$\begin{aligned} Z(X|Y) &= 2 \sum_{y \in \mathcal{Y}} \sqrt{P_{X,Y}(0, y)P_{X,Y}(1, y)} \\ &= -1 + \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in \mathcal{X}} \sqrt{P_{X,Y}(x, y)} \right]^2 \\ &\stackrel{(a)}{\geq} -1 + \sum_{y \in \mathcal{Y}} \sum_{j \in \mathcal{J}} Q(j) \left[ \sum_{x \in \mathcal{X}} \sqrt{P_{X,Y}^j(x, y)} \right]^2 \\ &= \sum_{j \in \mathcal{J}} Q(j) Z^j(X|Y) \end{aligned} \quad (18)$$

(a) by Minkowsky's inequality, when  $r < 1$  and  $a_{jk}$  is non-negative. Here  $r = 0.5$  and  $a_{jk} = \sqrt{P_{X,Y}^j(x, y)}$

$$\sum_{k \in \mathcal{K}} \left( \sum_{j \in \mathcal{J}} Q(j) a_{jk}^{\frac{1}{r}} \right)^r \geq \left[ \sum_{j \in \mathcal{J}} Q(j) \left( \sum_{k \in \mathcal{K}} a_{jk} \right)^{\frac{1}{r}} \right]^r$$

Hence proved.

**Lemma 6.** Let  $(X_1, Y_1)$  and  $(X_2, Y_2)$  independent random variable tuples may not be identically distributed.  $X_1$  and  $X_2$  are defined over  $\mathcal{X} = \{0, 1\}$ .  $Y_1$  and  $Y_2$  are distributed over alphabets  $\mathcal{Y}_1$  and  $\mathcal{Y}_2$ . Let  $U_1 = X_1 + X_2$  and  $U_2 = X_2$ . Then

1.  $Z(U_1|Y_1Y_2) \geq \max\{Z(X_1|Y_1), Z(X_2|Y_2)\}$
2.  $Z(U_2|U_1Y_1Y_2) = Z(X_1|Y_1)Z(X_2|Y_2)$

**Proof:**

$$\begin{aligned} & P_{U_1|Y_1Y_2}(u_1|y_1y_2) \\ &= \sum_{u_2 \in \mathcal{X}} P_{U_1, U_2|Y_1Y_2}(u_1, u_2|y_1y_2) \\ &= \sum_{u_2 \in \mathcal{X}} P_{X_1X_2|Y_1Y_2}(u_1 + u_2, u_2|y_1y_2) \\ &= \sum_{u_2 \in \mathcal{X}} P_{X_1|Y_1}(u_1 + u_2|y_1) P_{X_2|Y_2}(u_2|y_2) \\ & P_{U_2|Y_1Y_2U_1}(u_2|y_1y_2u_1) \\ &= \frac{P_{U_1, U_2|Y_1Y_2}(u_1, u_2|y_1y_2)}{P_{U_1|Y_1Y_2}(u_1|y_1y_2)} \\ &= \frac{P_{X_1|Y_1}(u_1 + u_2|y_1) P_{X_2|Y_2}(u_2|y_2)}{\sum_{u_2 \in \mathcal{X}} P_{X_1|Y_1}(u_1 + u_2|y_1) P_{X_2|Y_2}(u_2|y_2)} \\ & P_{U_1Y_1Y_2}(u_1, y_1, y_2) \\ &= \sum_{u_2 \in \mathcal{X}} P_{U_1U_2Y_1Y_2}(u_1, u_2, y_1, y_2) \\ &= P_{U_1U_2Y_1Y_2}(u_1, 0, y_1, y_2) + P_{U_1U_2Y_1Y_2}(u_1, 1, y_1, y_2) \\ &= P_{U_2}(0) P_{U_1Y_1Y_2|U_2}(u_1, y_1, y_2|0) \\ & \quad + P_{U_2}(1) P_{U_1Y_1Y_2|U_2}(u_1, y_1, y_2|1) \end{aligned}$$

Let

$$\begin{aligned}
P_{U_1 Y_1 Y_2}^1(u_1, y_1, y_2) &= P_{U_1 Y_1 Y_2 | U_2}(u_1, y_1, y_2 | 0) \\
P_{U_1 Y_1 Y_2}^2(u_1, y_1, y_2) &= P_{U_1 Y_1 Y_2 | U_2}(u_1, y_1, y_2 | 1) \\
Z^1(U_1 | Y_1 Y_2) &= 2 \sum_{y_1 y_2} P_{Y_1 Y_2}^1(y_1 y_2) \\
&\quad \sqrt{P_{U_1 | Y_1 Y_2}^1(0 | y_1 y_2) P_{U_1 | Y_1 Y_2}^1(1 | y_1 y_2)} \\
&= 2 \sum_{y_1 y_2} P_{Y_1 Y_2 | U_2}(y_1 y_2 | 0) \\
&\quad \sqrt{P_{U_1 | Y_1 Y_2 U_2}(0 | y_1 y_2 0) P_{U_1 | Y_1 Y_2 U_2}(1 | y_1 y_2 0)} \\
&\stackrel{(a)}{=} 2 \sum_{y_1 y_2} P_{Y_1}(y_1) P_{Y_2 | U_2}(y_2 | 0) \\
&\quad \sqrt{P_{X_1 | Y_1 Y_2 U_2}(0 | y_1 y_2 0) P_{X_1 | Y_1 Y_2 U_2}(1 | y_1 y_2 0)} \\
&\stackrel{(b)}{=} 2 \sum_{y_1 y_2} P_{Y_1}(y_1) P_{Y_2 | U_2}(y_2 | 0) \\
&\quad \sqrt{P_{X_1 | Y_1}(0 | y_1) P_{X_1 | Y_1}(1 | y_1)} \\
&= 2 \sum_{y_1} P_{Y_1}(y_1) \sqrt{P_{X_1 | Y_1}(0 | y_1) P_{X_1 | Y_1}(1 | y_1)} \\
&= Z(X_1 | Y_1)
\end{aligned}$$

(a) is because  $Y_1$  is independent of  $U_2$ ,  $Y_2$  is independent of  $Y_1$  given  $U_2$ . (b) is true  $X_1$  is independent of  $Y_2 U_2$  given  $Y_1$ . Similarly we can easily prove that  $Z^2(U_1 | Y_1 Y_2) = Z(X_1 | Y_1)$ . Now Lemma 5 implies that  $Z(U_1 | Y_1 Y_2) \geq Z(X_1 | Y_1)$  exchanging the roles of  $(X_1, Y_1)$  and  $(X_2, Y_2)$ . We also get  $Z(U_1 | Y_1 Y_2) \geq Z(X_2 | Y_2)$ . Therefore  $Z(U_1 | Y_1 Y_2) \geq \max\{Z(X_2 | Y_2), Z(X_1 | Y_1)\}$ .  
 $Z(U_2 | Y_1 Y_2 U_1)$

$$\begin{aligned}
&= 2 \sum_{y_1 y_2 u_1 \in \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{X}} P_{U_1 Y_1 Y_2}(u_1 y_1 y_2) \\
&\quad \cdot \sqrt{P_{U_2 | U_1 Y_1 Y_2}(0 | u_1 y_1 y_2) P_{U_2 | U_1 Y_1 Y_2}(1 | u_1 y_1 y_2)} \\
&= 2 \sum_{y_1 y_2 u_1 \in \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{X}} \frac{P_{U_1 Y_1 Y_2}(u_1 y_1 y_2)}{P_{U_1 | Y_1 Y_2}(u_1 | y_1 y_2)} \\
&\quad [P_{X_1 | Y_1}(u_1 | y_1) P_{X_2 | Y_2}(0 | y_2) \\
&\quad P_{X_1 | Y_1}(u_1 + 1 | y_1) P_{X_2 | Y_2}(1 | y_2)]^{0.5} \\
&= 2 \sum_{y_1 y_2 u_1 \in \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{X}} P_{Y_1 Y_2}(y_1 y_2) \\
&\quad [P_{X_1 | Y_1}(u_1 | y_1) P_{X_1 | Y_1}(u_1 + 1 | y_1) \\
&\quad P_{X_2 | Y_2}(0 | y_2) P_{X_2 | Y_2}(1 | y_2)]^{0.5} \\
&= 2 \sum_{u_1 \in \mathcal{X}} \sum_{y_1 \in \mathcal{Y}_1} \sum_{y_2 \in \mathcal{Y}_2} P_{Y_1}(y_1) P_{Y_2}(y_2) \\
&\quad [P_{X_1 | Y_1}(u_1 | y_1) P_{X_2 | Y_2}(0 | y_2) \\
&\quad P_{X_1 | Y_1}(u_1 + 1 | y_1) P_{X_2 | Y_2}(1 | y_2)]^{0.5} \\
&= Z(X_1 | Y_1) Z(X_2 | Y_2)
\end{aligned}$$

**Proof of Lemma 1:**

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{1}{N} |\mathcal{H}_X \cap \mathcal{L}_{X|Y_l}| &= \lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{H}_X - \mathcal{L}_{X|Y_l}^c| \\
&\stackrel{(a)}{=} \lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{H}_X - \mathcal{H}_{X|Y_l}| \\
&\stackrel{(b)}{=} \lim_{N \rightarrow \infty} \frac{1}{N} (|\mathcal{H}_X| - |\mathcal{H}_{X|Y_l}|) \\
&= H(X) - H(X | Y_l) \\
&= I(X; Y_l)
\end{aligned}$$

(a) is true since we know that  $\lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{H}_{X|Y_l}| = \lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{L}_{X|Y_l}^c|$ . (b) is true since we know that  $\mathcal{H}_{X|Y_l} \subset \mathcal{H}_X$ , by Lemma 4 (conditioning reduces the bhattacharya parameter).

**Proof of Lemma 2:**

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

(a) by Lemma 6.

$$\begin{aligned}
Z(V'_{y_j} | U^{1:x_j-1} V^{1:y_j-1} U'_{x_j}) &\stackrel{(a)}{=} Z(U_{x_j} | U^{1:x_j-1}) Z(V_{y_j} | V^{1:y_j-1}) \\
&\geq (1 - 2^{-N^\beta})(1 - 2^{-N^\beta}) \\
&\stackrel{(b)}{\geq} 1 - 2^{-N^{\beta'}}
\end{aligned}$$

(a) by Lemma 6. (b) can be verified true for any  $\beta' < \beta$  for large enough  $N$ .

$$\begin{aligned}
Z(U'_{x_j} | U^{1:x_j-1} V^{1:y_j-1} Y_1^{1:2N}) &\stackrel{(a)}{\geq} \max\{Z(U_{x_j} | U^{1:x_j-1} Y_1^{1:N}), \\
&\quad Z(V_{y_j} | U^{1:y_j-1} Y_1^{N+1:2N})\} \\
&\geq 1 - 2^{-N^\beta}
\end{aligned}$$

(a) by Lemma 6..

$$\begin{aligned}
Z(V'_{y_j} | U^{1:x_j-1} V^{1:y_j-1} U'_{x_j} Y_1^{1:2N}) &\stackrel{(a)}{=} Z(U_{x_j} | U^{1:x_j-1} Y_1^{1:N}) \\
&\quad Z(V_{y_j} | V^{1:y_j-1} Y_1^{N+1:2N}) \\
&\leq Z(U_{x_j} | U^{1:x_j-1} Y_1^{1:N}) \\
&\leq 2^{-N^\beta}
\end{aligned}$$

(a) by Lemma 6.

$$\begin{aligned}
Z(U'_{x_j} | U^{1:x_j-1} V^{1:N+y_j-1} Y_2^{1:2N}) &\stackrel{(a)}{\geq} \max\{Z(U_{x_j} | U^{1:x_j-1} Y_2^{1:N}), \\
&\quad Z(U_{y_j} | V^{1:N+y_j-1} Y_2^{N+1:2N})\} \\
&\geq 1 - 2^{-N^\beta}
\end{aligned}$$

(a) by Lemma 6..

$$\begin{aligned}
Z(V'_{y_i} | U^{1:x_i-1} V^{1:y_i-1} U'_{x_i} Y_2^{1:2N}) \\
&\stackrel{(a)}{=} Z(U_{x_i} | U^{1:x_i-1} Y_2^{1:N}) \\
&\quad Z(V_{y_i} | V^{1:y_i-1} Y_2^{N+1:2N}) \\
&\leq Z(U_{y_i} | V^{1:y_i-1} Y_2^{N+1:2N}) \\
&\leq 2^{-N^\beta}
\end{aligned}$$

Lemma 6.

## APPENDIX B PROOF OF THEOREM 2

Let  $(U'^{1:N}, V'^{1:N})$  be a codeword corresponding to message  $M$ . In our code construction we have one to one linear relation from  $(U'^{\mathcal{H}_X}, V'^{\mathcal{H}_X})$  to  $(U^{\mathcal{H}_X}, V^{\mathcal{H}_X})$  let it be  $H_{2|\mathcal{H}_X}|$ . For  $i \notin \mathcal{H}_X$  we have  $U'_i = U_i$  and  $V'_i = V_i$ . Let the transform which maps  $((U'^{1:N}, V'^{1:N}))$  to  $((U^{1:N}, V^{1:N}))$  be  $H_{2N}$ . For word  $(u'^{1:N}, v'^{1:N})$  the corresponding word with  $H_{2N}$  transform be  $(u^{1:N}, v^{1:N})$ . The probability for this word to be selected as codeword and received vector becomes  $y^{1:2N}$  is

$$\begin{aligned}
&2^{-(2|I_1 \cup I_2| + |\mathcal{Y}|)} 1[\cap_{i \in (\mathcal{H}_X - (I_1 \cap I_2))} \{f_{r_1}(i) = u'_i\}] \\
&1[\cap_{i \in (\mathcal{H}_X - ((I_1 \cap I_2) \cup \mathcal{Y}))} \{f_{r_2}(i) = v'_i\}] 1[\cap_{i \in (\mathcal{H}_X \cap \mathcal{L}_X)^c} \{\lambda^1(u^{1:i-1}) = u_i\}] 1[\cap_{i \in (\mathcal{H}_X \cap \mathcal{L}_X)^c} \{\lambda^2(v^{1:i-1}) = v_i\}] \\
&] \Pi_{i \in \mathcal{L}_X} (q(u_i | u^{1:i-1}) q(v_i | v^{1:i-1})) P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) P_{Y^{N+1:2N} | v^{1:N}}(y^{N+1:2N} | v^{1:N})
\end{aligned} \quad (19)$$

This is true because we have  $P_{U^{1:N} Y^{1:N}} = P_{V^{1:N} Y^{N+1:2N}}$  from equation 3. Then the probability of error for the  $i$ th bit channel of block  $b$  be  $\mathcal{E}_i^b$  for  $b \in \{1, 2\}$ . If  $(u'^{1:N}, v'^{1:N}, y^{1:2N})$  belongs to  $\mathcal{E}_i^1$  and  $i \in I_1 \cap I_2$  then

$$\begin{aligned}
&P_{U_i | U^{1:i-1} Y^{1:N}}(u_i + 1 | u^{1:i-1} y^{1:N}) \\
&\geq P_{U_i | U^{1:i-1} Y^{1:N}}(u_i | u^{1:i-1} y^{1:N})
\end{aligned} \quad (20)$$

and similarly belongs to  $\mathcal{E}_i^2$  and  $i \in I_1 \cap I_2$  then

$$\begin{aligned}
&P_{U_i | U^{1:i-1} Y^{1:N}}(v_i + 1 | v^{1:i-1} y^{1:N}) \\
&\geq P_{U_i | U^{1:i-1} Y^{1:N}}(v_i | v^{1:i-1} y^{1:N})
\end{aligned} \quad (21)$$

and belongs to  $\mathcal{E}_i^2$  for  $i \in \mathcal{Y}$ , for such  $i$ ,  $i$  is equal to  $y_j$  for some  $j$  in  $[G]$  then

$$\begin{aligned}
&P_{V'_{y_j} | U^{1:x_j-1} U'_{x_j} V^{1:y_j-1} Y^{1:2N}}(u_i + 1 | u^{1:x_j-1} u'_{x_j} v^{1:y_j-1} y^{1:2N}) \\
&\geq P_{V'_{y_j} | U^{1:x_j-1} U'_{x_j} V^{1:y_j-1} Y^{1:2N}}(u_i | u^{1:x_j-1} u'_{x_j} v^{1:y_j-1} y^{1:2N})
\end{aligned} \quad (22)$$

Let  $\mathcal{E} = \{\cup_{i \in I_1 \cap I_2} \mathcal{E}_i^1\} \cup \{\cup_{i \in (I_1 \cap I_2) \cup \mathcal{Y}} \mathcal{E}_i^2\}$ .

Now the probability of error for a given code is that for a given  $f_{r_1}, f_{r_2}, \lambda^1_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, \lambda^2_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}$

$$\begin{aligned}
&P_e(\lambda^1_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, \lambda^2_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, f_{r_1}, f_{r_2}) \\
&= \sum_{(u'^{1:N}, v'^{1:N}, y^{1:2N})} 2^{-(2|I_1 \cup I_2| + |\mathcal{Y}|)} 1[\cap_{i \in (\mathcal{H}_X - (I_1 \cap I_2))} \{f_{r_1}(i) = u'_i\}] \\
&1[\cap_{i \in (\mathcal{H}_X - ((I_1 \cap I_2) \cup \mathcal{Y}))} \{f_{r_2}(i) = v'_i\}] 1[\cap_{i \in (\mathcal{H}_X \cap \mathcal{L}_X)^c} \{\lambda^1(u^{1:i-1}) = u_i\}] 1[\cap_{i \in (\mathcal{H}_X \cap \mathcal{L}_X)^c} \{\lambda^2(v^{1:i-1}) = v_i\}] \\
&\Pi_{i \in \mathcal{L}_X} (q(u_i | u^{1:i-1}) q(v_i | v^{1:i-1})) P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) P_{Y^{N+1:2N} | v^{1:N}}(y^{N+1:2N} | v^{1:N}) \\
&1[(u'^{1:N}, v'^{1:N}, y^{1:2N}) \in \mathcal{E}]
\end{aligned} \quad (23)$$

By taking expectation with respect to  $f_{r_1}, f_{r_2}, \lambda^1_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, \lambda^2_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}$

$$\begin{aligned}
&\mathbb{E}[P_e(\lambda^1_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, \lambda^2_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, f_{r_1}, f_{r_2})] = \\
&\sum_{(u'^{1:N}, v'^{1:N}, y^{1:2N})} 2^{-2|\mathcal{H}_X|} \Pi_{i \in \mathcal{L}_X} (q(u_i | u^{1:i-1}) q(v_i | v^{1:i-1})) \Pi_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1}) \\
&P_{U_i | U^{1:i-1}}(v_i | v^{1:i-1}) P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) P_{Y^{N+1:2N} | v^{1:N}}(y^{N+1:2N} | v^{1:N})
\end{aligned} \quad (24)$$

Let us introduce measure of  $Q_{U^{1:N} V^{1:N} Y^{1:2N}}$  equivalently  $Q_{U^{1:N} V^{1:N} Y^{1:2N}}$ .

$$\begin{aligned}
&Q_{U^{1:N} V^{1:N} Y^{1:2N}}(u'^{1:N}, v'^{1:N}, y^{1:2N}) \\
&= Q_{U^{1:N} V^{1:N} Y^{1:2N}}(u^{1:N}, v^{1:N}, y^{1:2N}) \\
&= 2^{-2|\mathcal{H}_X|} \Pi_{i \in \mathcal{L}_X} (q(u_i | u^{1:i-1}) q(v_i | v^{1:i-1})) \Pi_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1}) \\
&P_{U_i | U^{1:i-1}}(v_i | v^{1:i-1}) P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) P_{Y^{N+1:2N} | v^{1:N}}(y^{N+1:2N} | v^{1:N})
\end{aligned} \quad (25)$$

This implies that

$$\begin{aligned}
&Q_{U^{1:N} V^{1:N} Y^{1:2N}}(\mathcal{E}) \\
&= \mathbb{E}[P_e(\lambda^1_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, \lambda^2_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}, f_{r_1}, f_{r_2})]
\end{aligned} \quad (26)$$

By marginalization, we can get

$$\begin{aligned}
&Q_{U^{1:N} Y^{1:N}}(u^{1:N}, y^{1:N}) \\
&= 2^{-|\mathcal{H}_X|} \Pi_{i \in \mathcal{L}_X} q(u_i | u^{1:i-1}) \Pi_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1}) P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N})
\end{aligned} \quad (27)$$

$$\begin{aligned}
&Q_{V^{1:N} Y^{N+1:2N}}(v^{1:N}, y^{N+1:2N}) \\
&= 2^{-|\mathcal{H}_X|} \Pi_{i \in \mathcal{L}_X} q(v_i | v^{1:i-1}) \Pi_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} P_{U_i | U^{1:i-1}}(v_i | v^{1:i-1}) P_{Y^{1:N} | U^{1:N}}(y^{N+1:2N} | v^{1:N})
\end{aligned} \quad (28)$$

This implies

$$\begin{aligned}
&Q_{U^{1:N} V^{1:N} Y^{1:2N}}(u^{1:N}, v^{1:N}, y^{1:2N}) \\
&= Q_{U^{1:N} Y^{1:N}}(u^{1:N}, y^{1:N}) Q_{V^{1:N} Y^{N+1:2N}}(v^{1:N}, y^{N+1:2N})
\end{aligned} \quad (29)$$

This implies that,  $U^{1:N}, Y^{1:N}$  and  $V^{1:N}Y^{N+1:2N}$  are independent and identically distributed with respect to measure  $Q$ . So we have from equation  $Q_{U^{1:N}Y^{1:N}} = Q_{V^{1:N}Y^{N+1:2N}}$  and also from equation 3 we get

$$\begin{aligned} & \|Q_{U^{1:N}Y^{1:N}} - P_{U^{1:N}Y^{1:N}}\| \\ &= \|Q_{V^{1:N}Y^{N+1:2N}} - P_{V^{1:N}Y^{N+1:2N}}\| \end{aligned} \quad (30)$$

$$\begin{aligned} & Q_{U^{1:N}V^{1:N}Y^{1:2N}}(\mathcal{E}) \\ & \leq \|Q_{U^{1:N}V^{1:N}Y^{1:2N}} - P_{U^{1:N}V^{1:N}Y^{1:2N}}\| \\ & \quad + P_{U^{1:N}V^{1:N}Y^{1:2N}}(\mathcal{E}) \\ & \leq \|Q_{U^{1:N}V^{1:N}Y^{1:2N}} - P_{U^{1:N}V^{1:N}Y^{1:2N}}\| \\ & \quad + P_{U^{1:N}V^{1:N}Y^{1:2N}}(\mathcal{E}) \\ & \leq \|Q_{U^{1:N}V^{1:N}Y^{1:2N}} - P_{U^{1:N}V^{1:N}Y^{1:2N}}\| \\ & \quad + \sum_{i \in I_1 \cap I_2} P_{U^{1:N}V^{1:N}Y^{1:2N}}(\mathcal{E}_i^1) \\ & \quad + \sum_{i \in (I_1 \cap I_2) \cup \mathcal{Y}} P_{U^{1:N}V^{1:N}Y^{1:2N}}(\mathcal{E}_i^2) \end{aligned} \quad (31)$$

Now we bound each term in the last inequality of the summation. Consider

$$\begin{aligned} & \|Q_{U^{1:N}V^{1:N}Y^{1:2N}} - P_{U^{1:N}V^{1:N}Y^{1:2N}}\| \\ &= \frac{1}{2} \sum_{(u^{1:N}, v^{1:N}, y^{1:2N})} |Q_{U^{1:N}V^{1:N}Y^{1:2N}}(u^{1:N}, v^{1:N}, y^{1:2N}) - P_{U^{1:N}V^{1:N}Y^{1:2N}}(u^{1:N}, v^{1:N}, y^{1:2N})| \\ &= \frac{1}{2} \sum_{(u^{1:N}, v^{1:N}, y^{1:2N})} |Q_{U^{1:N}Y^{1:N}}(u^{1:N}, y^{1:N}) \\ & \quad Q_{V^{1:N}Y^{N+1:2N}}(v^{1:N}, y^{N+1:2N}) \\ & \quad - P_{U^{1:N}Y^{1:N}}(u^{1:N}, y^{1:N}) \\ & \quad P_{V^{1:N}Y^{N+1:2N}}(v^{1:N}, y^{N+1:2N})| \\ &= \frac{1}{2} \sum_{(u^{1:N}, v^{1:N}, y^{1:2N})} |Q_{U^{1:N}Y^{1:N}}(u^{1:N}, y^{1:N}) \\ & \quad Q_{V^{1:N}Y^{N+1:2N}}(v^{1:N}, y^{N+1:2N}) \\ & \quad - Q_{U^{1:N}Y^{1:N}}(u^{1:N}, y^{1:N}) \\ & \quad P_{V^{1:N}Y^{N+1:2N}}(v^{1:N}, y^{N+1:2N}) \\ & \quad + Q_{U^{1:N}Y^{1:N}}(u^{1:N}, y^{1:N}) \\ & \quad P_{V^{1:N}Y^{N+1:2N}}(v^{1:N}, y^{N+1:2N}) \\ & \quad - P_{U^{1:N}Y^{1:N}}(u^{1:N}, y^{1:N}) \\ & \quad P_{V^{1:N}Y^{N+1:2N}}(v^{1:N}, y^{N+1:2N})| \end{aligned}$$

$$\begin{aligned} & \stackrel{(a)}{\leq} \frac{1}{2} \sum_{(u^{1:N}, v^{1:N}, y^{1:2N})} |Q_{U^{1:N}Y^{1:N}}(u^{1:N}, y^{1:N})| \\ & \quad (Q_{V^{1:N}Y^{N+1:2N}}(v^{1:N}, y^{N+1:2N}) - \\ & \quad P_{V^{1:N}Y^{N+1:2N}}(v^{1:N}, y^{N+1:2N}))| \end{aligned}$$

$$\begin{aligned} & + |(Q_{U^{1:N}Y^{1:N}}(u^{1:N}, y^{1:N}) - \\ & P_{U^{1:N}Y^{1:N}}(u^{1:N}, y^{1:N}))| \\ & P_{V^{1:N}Y^{N+1:2N}}(v^{1:N}, y^{N+1:2N}) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{2} \sum_{(v^{1:N}, y^{N+1:2N})} |Q_{V^{1:N}Y^{N+1:2N}}(v^{1:N}, y^{N+1:2N}) - \\ & \quad P_{V^{1:N}Y^{N+1:2N}}(v^{1:N}, y^{N+1:2N})| \\ & \quad + \frac{1}{2} \sum_{(u^{1:N}, y^{1:N})} |Q_{U^{1:N}Y^{1:N}}(u^{1:N}, y^{1:N}) - \\ & \quad P_{U^{1:N}Y^{1:N}}(u^{1:N}, y^{1:N})| \\ &= \|Q_{V^{1:N}Y^{N+1:2N}} - P_{V^{1:N}Y^{N+1:2N}}\| + \\ & \quad \|Q_{U^{1:N}Y^{1:N}} - P_{U^{1:N}Y^{1:N}}\| \\ & \stackrel{(b)}{=} 2\|Q_{U^{1:N}Y^{1:N}} - P_{U^{1:N}Y^{1:N}}\| = O(2^{-N^{\beta'}}) \end{aligned} \quad (32)$$

(a) is true by triangular inequality. (b) from equation 30. For  $i \in I_1 \cap I_2$ ,  $P(\mathcal{E}_i^b)$  can be bound as equation (60) in [6]. Hence

$$P(\mathcal{E}_i^b) \leq 2^{-N^{\beta}}. \quad (33)$$

For block 2 and  $i \in \mathcal{Y}$  there exists a  $j \in [G]$  such that  $i = y_j$ .

$$\begin{aligned} & P_{U^{1:N}, V^{1:N}Y^{1:2N}}(\mathcal{E}_i^1) \\ & \leq \sum_{u^{1:N}x_j v^{1:N}y_j-1 y^{1:2N}} P(u^{1:N}x_j v^{1:N}y_j-1, y^{1:2N}) \\ & \quad P(v'_{y_j} | u^{1:N}x_j v^{1:N}y_j-1, y^{1:2N}) [P(v'_{y_j} | u^{1:N}x_j v^{1:N}y_j-1, \\ & \quad y^{1:2N}) \leq P(v'_{y_j} + 1 | u^{1:N}x_j v^{1:N}y_j-1, y^{1:2N})] \\ & \leq \sum_{u^{1:N}x_j v^{1:N}y_j-1 y^{1:2N}} P(u^{1:N}x_j v^{1:N}y_j-1, y^{1:2N}) \\ & \quad P(v'_{y_j} | u^{1:N}x_j v^{1:N}y_j-1, y^{1:2N}) \\ & \quad \sqrt{\frac{P(v'_{y_j} + 1 | u^{1:N}x_j v^{1:N}y_j-1, y^{1:2N})}{P(v'_{y_j} | u^{1:N}x_j v^{1:N}y_j-1, y^{1:2N})}} \\ & \leq Z(V'_{y_j} | U^{1:N}x_j-1 V^{1:N}y_j-1 Y^{1:2N}) \\ & \stackrel{(a)}{\leq} 2^{-N^{\beta'}} \end{aligned} \quad (34)$$

(a) by Lemma 2.

Equations 31, 32, 33 and 34 conclude that

$$\mathbb{E}[P_e(\lambda_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}^1, \lambda_{(\mathcal{H}_X \cup \mathcal{L}_X)^c}^2, f_{r_1}, f_{r_2})] = O(2^{-N^{\beta'}})$$

## APPENDIX C PROOF OF THEOREM 3

### Step1:

Consider any polar block in the stair case structure which is completely in full column regime. This polar block corresponds to  $N$  different full columns. To get distribution on encoded codeword  $U^{1:N}$  for this polar block, first we compute the conditional distribution of  $\mathbb{P}(U_i = u_i | U^{1:i-1} = u^{1:i-1}, W^{1:N})$  for all  $i$  in the block.

If  $i \in \mathcal{L}_X$  we have by encoding rule

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

If  $i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c$  by 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  $i \in \mathcal{H}_X - I'$

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

This is true because in these indices we place bits of codeword  $m$  of linear MDS code corresponding to that column. We replaced the zero bits in those positions which have zeros in all codewords of  $\mathcal{M}$  with parity ( $X$ ) of information bits of the column to get modified codeword  $m$ . Hence for such code, at a given position, there are equal no. of zeros and ones in all codewords. Hence  $U_i$  is Ber(0.5) random variable and is only dependent on the information bits corresponding to that column of linear MDS codeword and independent to  $U^{1:i-1}$  of that block (previous columns).

If  $i \in I'$

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

Here we assign  $U_i$  as binary sum of parity of information bits corresponding to that column, let it be  $X$  and already encoded bit in the block corresponding to the  $g(i)$  in that column, let it be  $Y$ .  $U_i = X \oplus Y$ .  $X$  is Ber(0.5) and  $Y$  is Ber( $p$ ) for some  $p$  and both  $X$  and  $Y$  are independent clearly  $X$  is parity sum which is independently generated with respect to  $Y$ . Observe that  $U_i$  is independent of  $Y$ , because for a given  $Y$ , the distribution of  $X$  is again Ber(0.5), when constant is added to Ber(0.5) random variable we get Ber(0.5) random variable. Hence distribution of  $U_i$  is Ber(0.5) given random variable  $Y$ . It is noteworthy that  $U_i$  is independent of all the previously encoded bits of previous columns which implies that it is independent of  $U^{1:i-1}$  of that block. To see this let us denote  $\bar{P}_i$  be random vector denoting the encoded bits of previous columns.

$$\begin{aligned} & \mathbb{P}(U_i = u_i | \bar{P}_i, W^{1:N}) \\ &= \sum_{y \in \{0,1\}} \mathbb{P}(U_i = u_i, Y = y | \bar{P}_i, W^{1:N}) \\ &= \sum_{y \in \{0,1\}} \mathbb{P}(Y = y | \bar{P}_i, W^{1:N}) \\ & \quad \cdot \mathbb{P}(U_i = u_i | Y = y, \bar{P}_i, W^{1:N}) \\ &\stackrel{(a)}{=} \sum_{y \in \{0,1\}} \mathbb{P}(Y = y | \bar{P}_i, W^{1:N}) \mathbb{P}(U_i = u_i | Y = y) \\ &\stackrel{(b)}{=} \sum_{y \in \{0,1\}} \mathbb{P}(Y = y | \bar{P}_i, W^{1:N}) \mathbb{P}(U_i = u_i) \\ &= \mathbb{P}(U_i = u_i) \end{aligned} \quad (35)$$

(a) is true by Markovity, given  $Y$ ,  $U_i$  is independent of encoded bits of previous columns. (b) by independence of  $U_i$  with respect to  $Y$ . By combining everything for a block encoding in full column regime  $U^{1:N}$  is distributed as

$$\begin{aligned} & \mathbb{P}(U^{1:N} = u^{1:N}, Y^{1:N} = y^{1:N} | W^{1:N}) \\ &= \prod_{i \in [N]} \mathbb{P}(U_i = u_i | U^{1:i-1} = u^{1:i-1}, W^{1:N}) \\ & \quad P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) \\ &= 2^{-|\mathcal{H}_X|} \prod_{i \in \mathcal{L}_X} q(u_i | u^{1:i-1}) \prod_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} \\ & \quad P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1}) P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) \end{aligned} \quad (36)$$

This implies that,

$$\begin{aligned} & \mathbb{E}_{W^{1:N}} [\mathbb{P}(U^{1:N} = u^{1:N}, Y^{1:N} = y^{1:N} | W^{1:N})] \\ &= 2^{-|\mathcal{H}_X|} \prod_{i \in \mathcal{L}_X} q(u_i | u^{1:i-1}) \prod_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} \\ & \quad P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1}) P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) \end{aligned} \quad (37)$$

**Step2:**

Consider a block which is partly in right side of non full column side. For  $U_i$  in full column regime, the same conditional probability rule like above follows. For  $U_i$  in non full column regime, by encoding rule.

if  $i \in \mathcal{H}_X$

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

if  $i \in \mathcal{L}_X$

$$\mathbb{P}(U_i = u_i | U^{1:i-1} = u^{1:i-1}) = q(u_i | u^{1:i-1})$$

if  $i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c$

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

This implies again that,

$$\begin{aligned} & \mathbb{P}(U^{1:N} = u^{1:N}, Y^{1:N} = y^{1:N} | W^{1:N}) \\ &= \prod_{i \in [N]} \mathbb{P}(U_i = u_i | U^{1:i-1} = u^{1:i-1}, W^{1:N}) \\ & \quad P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) \\ &= 2^{-|\mathcal{H}_X|} \prod_{i \in \mathcal{L}_X} q(u_i | u^{1:i-1}) \prod_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} \\ & \quad P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1}) P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) \end{aligned} \quad (38)$$

This implies that,

$$\begin{aligned} & \mathbb{E}_{W^{1:N}} [\mathbb{P}(U^{1:N} = u^{1:N}, Y^{1:N} = y^{1:N} | W^{1:N})] \\ &= 2^{-|\mathcal{H}_X|} \prod_{i \in \mathcal{L}_X} q(u_i | u^{1:i-1}) \prod_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} \\ & \quad P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1}) P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) \end{aligned} \quad (39)$$

**Step3:**

Now if we consider a block which is partly in the left of stair case structure which is in non-full column regime and some part lies in the full-column regime. Assume that  $h$  encoded bits of the block are in left side of non full column regime, let those indices be the set  $H$  and encoded as  $W_i$  corresponding to index  $i$ .

$$\begin{aligned} & \mathbb{E}_{W^{1:N}} [\prod_{i \in H} 1(u_i = W_i)] \\ &= \mathbb{E}_{W^{1:N}} [1(\cap_{i \in H} (u_i = W_i))] \\ &\stackrel{(a)}{=} 2^{-|H \cap \mathcal{H}_X|} \prod_{i \in \mathcal{L}_X \cap H} q(u_i | u^{1:i-1}) \prod_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c \cap H} \\ & \quad P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1}) \end{aligned} \quad (40)$$

We get step (a) by marginalizing

$$\mathbb{E}_{W^{1:N}} [1(\cap_{i \in [N]} (u_i = W_i))] = 2^{-|\mathcal{H}_X|} \prod_{i \in \mathcal{L}_X} q(u_i | u^{1:i-1}) \prod_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1})$$

Now consider,

$$\begin{aligned} & \mathbb{P}(U^{1:N} = u^{1:N}, Y^{1:N} = y^{1:N} | W^{1:N}) \\ &= \prod_{i \in [N]} \mathbb{P}(U_i = u_i | U^{1:i-1} = u^{1:i-1}, W^{1:N}) \\ & \quad P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) \\ &= \prod_{i \in H} 1(u_i = W_i) 2^{-|\mathcal{H}_X - H|} \prod_{i \in \mathcal{L}_X - H} q(u_i | u^{1:i-1}) \\ & \quad \prod_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c - H} P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1}) \\ & \quad P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) \end{aligned} \quad (41)$$

This implies

$$\begin{aligned}
& \mathbb{E}_{W^{1:N}} [\mathbb{P}(U^{1:N} = u^{1:N}, Y^{1:N} = y^{1:N} | W^{1:N})] \\
&= \mathbb{E}_{W^{1:N}} [\prod_{i \in H} 1(u_i = W_i) 2^{-(|\mathcal{H}_X - H|)} \prod_{i \in \mathcal{L}_X - H} \\
&\quad q(u_i | u^{1:i-1}) \prod_{(\mathcal{H}_X \cup \mathcal{L}_X)^c - H} P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1})] \\
&\quad P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) \\
&= \mathbb{E}_{W^{1:N}} [\prod_{i \in H} 1(u_i = W_i) 2^{-(|\mathcal{H}_X - H|)} \prod_{i \in \mathcal{L}_X - H} \quad (42) \\
&\quad q(u_i | u^{1:i-1}) \prod_{(\mathcal{H}_X \cup \mathcal{L}_X)^c - H} P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1})] \\
&\quad P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) \\
&\stackrel{(a)}{=} 2^{-|\mathcal{H}_X|} \prod_{i \in \mathcal{L}_X} q(u_i | u^{1:i-1}) \prod_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} P_{U_i | U^{1:i-1}} \\
&\quad (u_i | u^{1:i-1}) P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N})
\end{aligned}$$

(a) by equation 40. Hence proved the part 1. For any block, we use the same decoding rule in section III to decode information bits. Let  $\mathcal{E}_b$  be the error event for the block in a stair case. Let  $P_b$  be a block error probability for the block.

$$\begin{aligned}
& P_b(W^{1:N}) \\
&= \sum_{(u^{1:N}, y^{1:N})} \mathbb{P}(U^{1:N} = u^{1:N}, Y^{1:N} = y^{1:N} | W^{1:N}) \quad (43) \\
&\quad 1((u^{1:N}, y^{1:N}) \in \mathcal{E}_b)
\end{aligned}$$

Taking expectation on both sides and by linearity of expectation we get,

$$\begin{aligned}
& \mathbb{E}_{W^{1:N}} [P_b(W^{1:N})] \\
&= \sum_{(u^{1:N}, y^{1:N})} \mathbb{E}_{W^{1:N}} [\mathbb{P}(U^{1:N} = u^{1:N}, Y^{1:N} = y^{1:N} | \\
&\quad W^{1:N}) 1((u^{1:N}, y^{1:N}) \in \mathcal{E}_b)] \\
&= \sum_{(u^{1:N}, y^{1:N})} 2^{-|\mathcal{H}_X|} \prod_{i \in \mathcal{L}_X} q(u_i | u^{1:i-1}) \prod_{i \in (\mathcal{H}_X \cup \mathcal{L}_X)^c} \quad (44) \\
&\quad P_{U_i | U^{1:i-1}}(u_i | u^{1:i-1}) P_{Y^{1:N} | U^{1:N}}(y^{1:N} | u^{1:N}) \\
&\quad 1((u^{1:N}, y^{1:N}) \in \mathcal{E}_b) \\
&= Q(\mathcal{E}_b)
\end{aligned}$$

$Q(\mathcal{E}_b)$  is bounded exactly similarly to the proof of Theorem 1. Hence we get part 2 by using union bound.

#### Encoding Complexity:

Encoding complexity consists of two factors, encoding polar block which takes  $O(N \log_2 N)$  real operations and encoding RS codeword. Encoding RS codeword can be done in  $O(N \log_2(N))$  binary operations over  $\text{GF}(2^q)$ . Addition and multiplication over this field takes  $q$  and  $q^{\log_2(3)}$  binary operations respectively. One RS codeword contains  $Nq$  bits. Hence the claim follows.

#### Decoding Complexity:

Decoding complexity consists of two factors, decoding polar block which takes  $O(N \log_2 N)$  real operations and decoding RS codeword. Decoding RS codeword can be done in  $O(N(\log_2(N))^2)$  operations over  $\text{GF}(2^q)$ . Addition and multiplication over this field takes  $q$  and  $q^{\log_2(3)}$  binary operations respectively. One codeword contains  $Nq$  bits. Hence the claim follows. This concludes part 3.