Forty-Ninth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 28 - 30, 2011

Multilevel Polarization of Polar Codes Over Arbitrary Discrete Memoryless Channels

Aria G. Sahebi and S. Sandeep Pradhan
Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, MI 48109, USA.
Email: ariaghs@umich.edu, pradhanv@umich.edu

***Abstract*—It is shown that the original construction of polar codes suffices to achieve the symmetric capacity of discrete memoryless channels with arbitrary input alphabet sizes. It is shown that in general, channel polarization happens in several, rather than only two, levels so that the synthesized channels are either useless, perfect or "partially perfect". Given a coset decomposition of the input alphabet, there exists a corresponding partially perfect channel whose outputs uniquely determine the coset where the channel input symbol belongs to. By a slight modification of the encoding and decoding rules, it is shown that perfect transmission of certain information letters over partially perfect channels is possible. It is also shown through an example that polar codes do not achieve the capacity of coset codes over arbitrary channels.**

I. Introduction

Polar codes were originally proposed by Arikan in [1] for discrete memoryless channels with binary input. Polar codes over binary input channels are shifted linear (coset) codes that achieve the symmetric capacity of these channels and are constructed based on the Kronecker power of the 2×2 matrix $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. It is known that non-binary codes outperform binary codes in certain communication settings. In order to construct capacity achieving codes over non binary channels, there have been attempts to extend polar coding techniques for channels of arbitrary input sizes. It is shown in [5] that polar codes achieve the symmetric capacity of the channel when the size of the input alphabet is a prime. For channels of arbitrary input alphabet sizes however, it is shown that the original construction of polar codes does not achieve the symmetric capacity of the channel and a randomized construction of polar codes based on permutations is proposed to achieve the symmetric capacity of the channel [5]. In another approach in [5], a multilevel code construction is proposed which is based on the decomposition of the composite input channel into channels of prime input alphabet size. In [2], It is shown that for channels for which the input alphabet size is a prime power, polar codes defined on the input alphabet can achieve the symmetric capacity without the need to use multilevel code construction methods. Independent from this work, Park and Barg [3] observed the multilevel polarization for channels of input alphabet a power of 2.

This work was supported by NSF grant CCF-0915619.

In this paper, we show that with a slight modification of the encoding and decoding rules, the original generator matrix used for binary input channels suffices to achieve the symmetric capacity of the channel. Coset codes, in general, do not achieve the symmetric capacity of channels and hence the constructed polar code cannot generally be a coset code. Although the code uses the same generator matrix as in the binary case, the constructed code is not a coset code due to the fact that the message tuples do not form a group.

The paper is organized as follows, In section II some definitions and basic facts are stated which are used in the paper. In section III, it is shown that polar codes achieve the symmetric capacity of channels with input alphabet size $q = p^r$ where p is a prime and r is an integer. This result is generalized to arbitrary channels in section IV. In section V, two examples of channels over $\mathbb{Z}_4$ are provided. The intent of the first example is to illustrate the multilevel polarization of the channel and in the second example we show that polar codes as coset codes (i.e. without the modification of the encoding rule), do not achieve the capacity of coset codes over the channel.

II. Preliminaries

1) Source and Channel Models: We consider discrete memoryless and stationary channels used without feedback. We associate two finite sets $\mathcal{X}$ and $\mathcal{Y}$ with the channel as the channel input and output alphabets. These channels can be characterized by a conditional probability law $W(y|x)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. The channel is specified by $(\mathcal{X}, \mathcal{Y}, W)$. Assuming a perfect source coding block applied prior to the channel coding, the source of information generates messages over the set $\{1, 2, \ldots, M\}$ uniformly.

2) Achievability and Capacity: A transmission system with parameters (n, M, τ) for reliable communication over a given channel $(\mathcal{X}, \mathcal{Y}, W)$ consists of an encoding mapping and a decoding mapping $e : \{1, 2, \ldots, M\} \to \mathcal{X}^n$, $f : \mathcal{Y}^n \to \{1, 2, \ldots, M\}$ such that for all $m = 1, 2, \ldots, M$,

$$\frac{1}{M} \sum_{m=1}^{M} W^n \left(f(Y^n) \neq m | X^n = e(m) \right) \leq \tau$$

Given a channel $(\mathcal{X}, \mathcal{Y}, W)$, the rate R is said to be achievable if for all $\epsilon > 0$ and for all sufficiently large n, there

exists a transmission system for reliable communication with parameters (n, M, τ) such that $\frac{1}{n}\log M \geq R - \epsilon,\ \tau \leq \epsilon$.

3) Symmetric Capacity and the Bhattacharyya Parameter: For a channel $(\mathcal{X}, \mathcal{Y}, W)$, the symmetric capacity is defined as $I^0 = I(X;Y)$ when the channel input X is uniformly distributed over $\mathcal{X}$ and Y is the output of the channel. i.e. for $q = |\mathcal{X}|$,

$$I^0(W) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \frac{1}{q} W(y|x) \log \frac{W(y|x)}{\sum_{\tilde{x}\in\mathcal{X}} \frac{1}{q} W(y|\tilde{x})}$$

The Bhattacharyya distance between two distinct input symbols x and $\tilde{x}$ is defined as

$$Z(W_{\{x,\tilde{x}\}}) = \sum_{y\in\mathcal{Y}} \sqrt{W(y|x)W(y|\tilde{x})}$$

and the average Bhattacharyya distance is defined as

$$Z(W) = \sum_{x,\tilde{x}\in\mathcal{X}, x\neq\tilde{x}} \frac{1}{q(q-1)} Z(W_{\{x,\tilde{x}\}})$$

4) Polar Codes: For any $N = 2^n$, the generator matrix for polar codes is defined as $G_N = B_N F^{\otimes n}$ where B_N is a permutation of rows, $F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $\otimes$ denotes the Kronecker product. The decoding algorithm for polar codes is a specific form of successive cancellation [1].

5) Polar Codes Over Abelian Groups: For any discrete memoryless channel, there always exists an *Abelian Group* structure defiend over the channel input alphabet. In general, for an Abelian group, there may not exist a multipliction operation and therefore, before using polar codes for channels of arbitrary input alphabet size, a generator matrix for codes over abelian groups needs to be properly defined. In section VII-A, a convention is introduced to generate codes over groups using $0-1$ valued generator matrices.

6) Group Codes: Let the channel input alphabet $\mathcal{X}$ be equipped with the structure of a finite Abelian group G of the same size. Then the channel is specified by $(G, \mathcal{Y}, W)$. A group code over G of length n for this channel is any *subgroup* of G^n. The group capacity of a channel $(G, \mathcal{Y}, W)$ is the maximum achievable rate using group codes over G for this channel. Group codes generalize the notion of linear codes over *fields* to channels with composite input alphabet sizes. A coset code is a shift of a group code by a constant vector.

7) Notation: We denote by $O(\epsilon)$ any function of ϵ such that $O(\epsilon) \to 0$ as $\epsilon \to 0$. We denote by $a \approx b$ if $a = b + O(\epsilon)$. Given a partition $\{A_0, A_1, \cdots, A_r\}$ of the index set $[1, N]$ and sets T_t for $t = 0, \cdots, r$, the direct sum $\bigoplus_{i=1}^{I} T_t^{A_t}$ is defined as the set of all tuples $u_1^N = (u_1, \cdots, u_N)$ such that $u_i \in T_t$ whenever $i \in A_t$.

III. Polar Codes Over $\mathbb{Z}_{p^r}$ Rings

In this section, we consider channels of input alphabet size $q = p^r$ for some prime number p and a positive integer r. In this case, the input alphabet of the channel can be equipped with addition and multiplication modulo p^r to be considered as a *ring*. We prove the achievability of the symmetric capacity of these channels using polar codes and later in section IV we will generalize this result to channels of arbitrary input alphabet sizes.

A. $\mathbb{Z}_{p^r}$ Rings

Let $G = \mathbb{Z}_{p^r} = \{0, 1, 2, \cdots, p^r - 1\}$ be the input alphabet of the channel, where p is a prime and r is an integer. For $t = 0, 1, \cdots, r$, define the *subgroups* H_t of G as:

$$H_t = p^t G = \{0, p^t, 2p^t, \cdots, (p^{r-t} - 1)p^t\}$$

and for $t = 0, 1, \cdots, r$, define the subsets K_t of G as $K_t = H_t \backslash H_{t+1}$. Note that K_0 is the set of all invertible elements of G and $K_r = \{0\}$. One can sort the sets $K_0 < K_1 < \cdots < K_r$ in a decreasing order of "inversability" of its elements. Let T_t be a *transversal* of H_t in G i.e. T_t is a subset of G containing one and only one element from each coset (shift) of H_t in G. One valid choice for T_t is $\{0, 1, \cdots, p^t - 1\}$.

B. Recursive Channel Transformation

1) The Basic Channel Transforms: It has been shown in [1] that the error probability of polar codes over binary input channels is upper bounded by the sum of the Bhattacharyya parameters of certain channels defined by a recursive channel transformation. The same set of synthesized channels appear for polar codes over channels with arbitrary input alphabet sizes. The channel transformations are given by:

$$W^-(y_1, y_2|u_1) = \sum_{u_2\in G} \frac{1}{q} W(y_1|u_1 + u_2) W(y_2|u_2) \quad \text{for } y_1, y_2 \in \mathcal{Y} \text{ and } u_1 \in G \tag{1}$$

$$W^+(y_1, y_2, u_1|u_2) = \frac{1}{q} W(y_1|u_1 + u_2) W(y_2|u_2) \quad \text{for } y_1, y_2 \in \mathcal{Y} \text{ and } u_1, u_2 \in G \tag{2}$$

Repeating these operations n times recursively, we obtain $N = 2^n$ channels $W_N^{(1)}, \cdots, W_N^{(N)}$. For $i = 1, \cdots, N$, these channels are given by:

$$W_N^{(i)}(y_1^N, u_1^{i-1}|u_i) = \sum_{u_{i+1}^N \in G^{N-i}} \frac{1}{q^{N-1}} W^N(y_1^N | u_1^N G_N)$$

Where G_N is the generator matrix for polar codes.
For the case of binary input channels, it has been shown [1] that as $N \to \infty$, these channels polarize in the sense that their Bhattacharyya parameter is either close to zero (perfect channel) or close to one (useless channel). In the next part, we show that in general, when the input alphabet is a prime power, the polarization happens in multiple levels so that as $N \to \infty$ channels get useless, perfect or partially perfect.
For an integer n, let J be a uniform random variable over the set $\{1, 2, \cdots, N = 2^n\}$ and define the random variable $I^n(W)$ as

$$I^n(W) = I(W_N^{(J)}) \tag{3}$$

It has been shown in [5] that the process $I^0, I^1, I^2, \cdots$ is a martingale; hence $\mathbb{E}\{I^n\} = I^0$. Similarly, for an integer n, define the random variable $Z_d^n(W) = Z_d(W_N^{(J)})$ where

$$Z_d(W) = \frac{1}{q} \sum_{x \in G} \sum_{y \in \mathcal{Y}} \sqrt{W(y|x)W(y|x+d)} \tag{4}$$

This quantity is defined in [5].

2) Asymptotic Behavior of Synthesized Channels: It is shown in Lemma VII.1 that Z_d^n converges to a Bernoulli variable. The next lemma gives a sufficient condition for two processes Z_d^n and $Z_{d'}^n$ to converge to the same random variable.

Lemma III.1. *If $d, d' \in K_t$ for some $0 \leq t \leq r-1$, then Z_d^n and $Z_{d'}^n$ converge to the same bernoulli variable.*

Proof: It has been shown that Z_d^n and $Z_{d'}^n$ both converge to bernoulli random variables. It suffices to show that $Z_d^n \approx 1$ implies $Z_{d'}^n \approx 1$ and $Z_d^n \approx 0$ implies $Z_{d'}^n \approx 0$. First assume $Z_d^n \approx 1$. Lemma VII.4 implies that for all $y \in \mathcal{Y}$, if $x - x' \in \langle d \rangle = \langle d' \rangle$ then $W(y|x) \approx W(y|x')$. This and Lemma VII.8 in turn imply that $Z_{d'}^n \approx 1$.
Next, assume $Z_d^n \approx 0$ and assume for contradiction that $Z_{d'}^n \approx 1$. Same as above, the second assumption implies that $Z_d^n \approx 1$ which is a contradiction. ∎

For $t = 0 \cdots r-1$, pick an arbitrary element $k_t \in K_t$. The lemma above suggests that we only need to study Z_{k_t}'s rather than all Z_d's.

Lemma III.2. *if $Z_{k_t} \approx 1$ then $Z_{k_s} \approx 1$ for all $t \leq s \leq r-1$.*

Proof: Follows from lemma VII.4 and lemma VII.8 and the fact that $k_s \in \langle k_t \rangle$. ∎

This lemma implies that for the group $G = \mathbb{Z}_{p^r}$ all possible asymptotic cases are:

- **case 0:** $Z_{k_0} = 1, Z_{k_1} = 1, Z_{k_2} = 1, \cdots, Z_{k_{r-1}} = 1$.
- **case 1:** $Z_{k_0} = 0, Z_{k_1} = 1, Z_{k_1} = 1, \cdots, Z_{k_{r-1}} = 1$.
- **case 2:** $Z_{k_0} = 0, Z_{k_1} = 0, Z_{k_1} = 1, \cdots, Z_{k_{r-1}} = 1$.

  $\vdots$
- **case r:** $Z_{k_0} = 0, Z_{k_1} = 0, Z_{k_1} = 0, \cdots, Z_{k_{r-1}} = 0$.

Where for $t = 0, \cdots, r$, the case t happens with some probability p_t.
For $t = 0, \cdots, r$, define $Z^t(W_N^{(i)}) = \sum_{d \notin H_t} Z_d(W_N^{(i)})$. Note that $Z^t(W_N^{(i)}) = 0$ in case t. Next, We study the behavior of I^n in each of these asymptotic cases.

Lemma III.3. *For a channel $(\mathbb{Z}_{p^r}, \mathcal{Y}, W)$ and for $t = 0, 1, \cdots, r$, if $Z_{k_0} \approx 0, Z_{k_1} \approx 0, \cdots, Z_{k_{t-1}} \approx 0, Z_{k_t} \approx 1, \cdots, Z_{k_{r-1}} \approx 1$, then $I^0(W) \approx t \log p$.*

Proof: We first prove the statement for the case where the approximate equalities are replaced with equalities. From Lemma VII.4, we have $W(y|x) = W(y|\tilde{x})$ if $x - \tilde{x} \in H_t$ and from Lemma VII.9 we have $W(y|x)W(y|\tilde{x}) = 0$ if $x - \tilde{x} \in K_0 \cup K_1 \cup K_{t-1} = G \backslash H_t$. Therefore, for $y \in \mathcal{Y}$ with positive probability $p_Y(y)$, $W(y|x)$ is uniform over a coset of H_t and zero over all other cosets. i.e. for all $y \in \mathcal{Y}$, there exists a coset C_t^y of H_t such that $\frac{1}{q}W(y|x) = \frac{p_Y(y)}{|C_t^y|} = \frac{p_Y(y)}{p^{r-t}}$ for $y \in C_t^y$ and $W(y|x) = 0$ otherwise. The mutual information is equal to

$$\begin{aligned}
I^0(W) &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \frac{1}{q} W(y|x) \log \frac{W(y|x)}{\sum_{\tilde{x} \in \mathcal{X}} \frac{1}{q} W(y|\tilde{x})} \\
&= \sum_{y \in \mathcal{Y}} \sum_{x \in C_t^y} \frac{1}{q} W(y|x) \log \frac{W(y|x)}{\sum_{\tilde{x} \in C_t^y} \frac{1}{q} W(y|\tilde{x})} \\
&= \sum_{y \in \mathcal{Y}} \sum_{x \in C_t^y} \frac{p_Y(y)}{p^{r-t}} \log \frac{\frac{p^r p_Y(y)}{p^{r-t}}}{\sum_{\tilde{x} \in C_t^y} \frac{p_Y(y)}{p^{r-t}}} \\
&= \sum_{y \in \mathcal{Y}} p^{r-t} \frac{p_Y(y)}{p^{r-t}} \log \frac{\frac{p^r p_Y(y)}{p^{r-t}}}{p^{r-t} \frac{p_Y(y)}{p^{r-t}}} \\
&= \sum_{y \in \mathcal{Y}} p_Y(y) \log p^t = t \log p
\end{aligned}$$

The lemma is proved considering the continuity of the mutual information. ∎

We have shown that the process I^n converges to an $r+1$ valued discrete random variable: $I = t \log p$ with probability p_t for $t = 0, \cdots, r$.

3) Summary of Channel Transformation: For the channel $(\mathbb{Z}_{p^r}, \mathcal{Y}, W)$, the convergence of the processes I^n and $(Z^t)^n$ for $t = 0, 1, \cdots, r$ implies that for all $\epsilon > 0$, there exists a number N and a partition $\{A_0^\epsilon, A_1^\epsilon, \cdots, A_r^\epsilon\}$ of $[1, N]$ such that for $t = 0, \cdots, r$ and $i \in A_t^\epsilon$, $I(W_N^{(i)}) = t \log(p) + O(\epsilon)$ and $Z^t(W_N^{(i)}) = O(\epsilon)$. Moreover, as $\epsilon \to 0$, $\frac{|A_t^\epsilon|}{N} \to p_t$ for some probabilities $p_0, \cdots, p_r$.

C. Encoding and Decoding

In the original construction of polar codes, we fix the input symbols corresponding to useless channels and send information symbols over perfect channels. Here, since the channels do not polarize into two levels, the encoding is slightly different and we send "some" information bits over "partially perfect" channels. At the encoder, if $i \in A_t^\epsilon$ for some $t = 0, \cdots, r$, the information symbol is chosen from the transversal T_t arbitrarily and not from the whole set G. As we will see later, the channel $W_N^{(i)}$ is perfect for symbols chosen from T_t and perfect decoding is possible at the decoder. Let $\mathcal{X}_N^\epsilon = \bigoplus_{t=0}^r T_t^{A_t^\epsilon}$ be the set of all valid input sequences. For the sake of analysis, as in the binary case, the message u_1^N is dithered with a uniformly distributed random vector $b_1^N \in \bigoplus_{t=0}^r H_t^{A_t^\epsilon}$ revealed to both the encoder and the decoder. A message $v_1^N \in \mathcal{X}_N^\epsilon$ is encoded to the vector $x_1^N = (v_1^N + b_1^N)G_N$. Note that $u_1^N = v_1^N + b_1^N$ is uniformly distributed over G^N.
At the decoder, after observing the output vector y_1^N, for $t = 0, \cdots, r$ and $i \in A_t$, use the following decoding rule:

$$\hat{u}_i = f_i(y_1^N, \hat{u}_1^{i-1}) = \arg\max_{g \in T_t} W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|g)$$

And finally, the message is recovered as $v_1^N = u_1^N - b_1^N$. The total number of valid input sequences is equal to

$$2^{NR} = \prod_{t=0}^{r} |T_t|^{|A_t|} = \prod_{t=0}^{r} p^{t|A_t|} = \prod_{t=0}^{r} p^{tp_t N}$$

Therefore, the rate is equal to $R = \sum_{t=0}^{r} p_t t \log p$. On the other hand, since I^n is a martingale [5], we have $\mathbb{E}\{I^\infty\} = I^0$. Since $\mathbb{E}\{I^\infty\} = \sum_{t=0}^{r} p_t t \log p$ we observe that the rate R is equal to the symmetric capacity I^0. We will see in the next section that this rate is achievable.

D. Error Analysis

Let B_i be the event that the first error occurs when the decoder decodes the ith symbol:

$$\begin{aligned} B_i = &\big\{(u_1^N, y_1^N) \in G^N \times \mathcal{Y}^N | u_1^{i-1} = f_1^{i-1}(u_1^N, y_1^N) \\ &, u_i \neq f_i(u_1^N, y_1^N)\big\} \\ &\subseteq \big\{(u_1^N, y_1^N) \in \mathcal{X}_N^\epsilon \times \mathcal{Y}^N | u_i \neq f_i(u_1^N, y_1^N)\big\} \end{aligned}$$

For $t = 0, \cdots, r$ and $i \in A_t$, define

$$\begin{aligned} E_i = \Big\{&(u_1^N, y_1^N) \in \mathcal{X}_N^\epsilon \times \mathcal{Y}^N | W_N^{(i)}(y_1^N, u_1^{i-1}|u_i) \\ &\leq W_N^{(i)}(y_1^N, u_1^{i-1}|\tilde{u}_i) \text{ for some } \tilde{u}_i \in T_t\Big\} \end{aligned}$$

Lemma III.4. *For $t = 0, \cdots, r$ and $i \in A_t$, $P(E_i) \leq qZ^t(W_N^{(i)}) = O(\epsilon)$.*

Proof: For $u_i, \tilde{u}_i \in T_t$, if $u_i \neq \tilde{u}_i$, then $u_i, \tilde{u}_i$ are not in the same coset of H_t and hence $u_i - \tilde{u}_i \notin H_t$. Similarly, $u_i - \tilde{u}_i \notin H_s$ for $s = t, t+1, \cdots, r$.

$$\begin{aligned} P(E_i) &= \sum_{u_1^N, y_1^N} \frac{1}{q^N} W_N(y_1^N, u_1^N) 1_{E_i}(u_1^N, y_1^N) \\ &\leq \sum_{u_1^N, y_1^N} \frac{1}{q^N} W_N(y_1^N, u_1^N) \sum_{\tilde{u}_i \neq u_i} \sqrt{\frac{W_N^{(i)}(y_1^N, u_1^{i-1}|\tilde{u}_i)}{W_N^{(i)}(y_1^N, u_1^{i-1}|u_i)}} \\ &= \sum_{u_i \in T_t} \sum_{\tilde{u}_i \in T_t, \tilde{u}_i \neq u_i} \frac{1}{q} Z_{\{u_i, \tilde{u}_i\}}(W_N^{(i)}) \end{aligned}$$

Note that for $d = u_i - \tilde{u}_i$, $Z_{\{u_i,\tilde{u}_i\}}(W_N^{(i)}) \leq qZ_d(W_N^{(i)})$. Since $d \in K_0 \cup K_1 \cup \cdots K_{t-1} = G \backslash H_t$,

$$Z_{\{u_i,\tilde{u}_i\}}(W_N^{(i)}) \leq qZ^t(W_N^{(i)}) = O(\epsilon)$$

■

The probability of block error is given by $P(err) = \sum_{t=0}^{r} \sum_{i \in A_t} P(B_i)$. Since $B_i \subseteq E_i$, we get

$$P(err) \leq \sum_{t=0}^{r} \sum_{i \in A_t} qZ^t(W_N^{(i)}) = O(\epsilon)$$

Therefore, the probability of block error goes to zero as $\epsilon \to 0$.

IV. Polar Codes Over Abelian Groups

For any channel input alphabet size there always exist an Abelian group of the same size. In this section, we generalize the result of the previous section to channels of arbitrary input alphabet size.

A. Abelian Groups

Let the Abelian group G be the input alphabet of the channel. It is a standard fact that any Abelian group can be decomposed as a direct sum of $\mathbb{Z}_{p^r}$ groups. Let $G = \bigoplus_{l=1}^{L} R_l$ with $R_l = \mathbb{Z}_{p_l^{r_l}}$ where p_l's are prime and r_l's are integers. For $t = (t_1, t_2, \cdots, t_L)$ with $t_l \in \{0, 1, \cdots, r_l\}$, there exists a corresponding subgroup H of G defined by $H = \bigoplus_{l=1}^{L} p_l^{t_l} R_l$. Let T_H be a transversal of H in G.

B. Recursive Channel Transformation

1) The Basic Channel Transforms: The synthesized channels $W_N^{(i)}$ and the process $I^n(W)$ are defined exactly the same way as the $\mathbb{Z}_{p^r}$ case through Equations 1 and 3.

2) Asymptotic Behavior of Synthesized Channels: Define $Z_d^n(W)$ same as 4, it has been shown that For all $d \neq 0$, $Z_d^\infty(W)$ is $0-1$ valued.

Lemma IV.1. *If $Z_{d_1}(W) \approx 1$ and $Z_{d_2}(W) \approx 1$, then $Z_{d'}(W) \approx 1$ for $d' \in \langle d_1, d_2 \rangle$.*

Proof: Immediate from Lemma VII.2 and Lemma VII.6. ■

It has been shown that all $Z_d(W)$'s are $0-1$ valued. Let $d_1, d_2, \cdots, d_m$ be the set od all d's with $Z_d = 1$. It has been shown in Lemma IV.1 that if $Z_{d_1}(W) \approx 1$ and $Z_{d_2}(W) \approx 1$, then $Z_{d'}(W) \approx 1$ for $d' \in \langle d_1, d_2 \rangle$ where $\langle d_1, d_2 \rangle$ is the subgroup generated by d_1 and d_2. This lemma is generalizable to the case where $Z_{d_1}(W) \approx 1, Z_{d_2}(W) \approx 1, \cdots, Z_{d_m}(W) \approx 1$. In this case we have $Z_{d'}(W) \approx 1$ for $d' \in \langle d_1, d_2, \cdots, d_m \rangle$. This means $Z_d = 1$ whenever $d \in H$ where H is the subgroup of G generated by $d_1, d_2, \cdots, d_m$ and $Z_d = 0$ otherwise. Hence all possible asymptotic cases can be indexed by subgroups of G. i.e. for any $H \leq G$, one possible asymptotic case is

- **Case H:** $Z_d(W) = \begin{cases} 1 & \text{if } d \in H; \\ 0 & \text{Otherwise.} \end{cases}$

For $H \leq G$, define $Z^H(W_N^{(i)}) = \sum_{d \notin H} Z_d(W_N^{(i)})$. Note that $Z^H(W_N^{(i)}) = O(\epsilon)$ in case H.
Next, We study the behavior of I^n in each of these cases.

Lemma IV.2. *For a channel $(G, \mathcal{Y}, W)$ and for a subgroup $H = \bigoplus_{l=1}^{L} p_l^{t_l} R_l$ of G, if $Z_d \approx 1$ for $d \in H$ and $Z_d \approx 0$ for $s \notin H$, then $I^0(W) \approx \log \frac{|G|}{|H|} = \sum_{l=1}^{L} t_l \log p_l$.*

Proof: Considering the continuity of the mutual information, it suffices to show the lemma when the approximate equalities are replaced with equalities. From Lemma VII.4, we have $W(y|x) = W(y|\tilde{x})$ if $x - \tilde{x} \in H$ and from Lemma VII.9 we have $W(y|x)W(y|\tilde{x}) = 0$ if $x - \tilde{x} \in G \backslash H_t$. Therefore for $y \in \mathcal{Y}$ with positive probability $p_Y(y)$, $W(y|x)$ is uniform over a coset of H and zero over all other cosets. i.e. for all $y \in \mathcal{Y}$, there exists a coset C_H^y of H such that $\frac{1}{q} W(y|x) = \frac{p_Y(y)}{|C_H^y|} = \frac{p_Y(y)}{\prod_{l=1}^{L} p^{r-t_l}}$ for $y \in C_H^y$ and $W(y|x) = 0$

otherwise. The mutual information is equal to

$$\begin{aligned}
I^0(W) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\frac{1}{q}W(y|x)\log\frac{W(y|x)}{\sum_{\tilde{x}\in\mathcal{X}}\frac{1}{q}W(y|\tilde{x})} \\
&= \sum_{y\in\mathcal{Y}}\sum_{x\in C_H^y}\frac{1}{q}W(y|x)\log\frac{W(y|x)}{\sum_{\tilde{x}\in C_H^y}\frac{1}{q}W(y|\tilde{x})} \\
&= \sum_{y\in\mathcal{Y}}\sum_{x\in C_H^y}\frac{p_Y(y)}{\prod_{l=1}^L p^{r-t_l}}\log\frac{\frac{\prod_{l=1}^L p^{r_l}p_Y(y)}{\prod_{l=1}^L p^{r-t_l}}}{\sum_{\tilde{x}\in C_H^y}\frac{p_Y(y)}{p^{r-t}}} \\
&= \sum_{y\in\mathcal{Y}}p_Y(y)\log\prod_{l=1}^L p^{t_l} = \sum_{l=1}^L t_l\log p_l = \log\frac{|G|}{|H|}
\end{aligned}$$

■

We have shown that the process I^n converges to a discrete random variable: $I = \log\frac{|G|}{|H|}$ with probability p_H for $H \le G$.

3) Summary of Channel Transformation: For the channel $(G, \mathcal{Y}, W)$, the convergence of the processes I^n and $(Z^H)^n$ for $t = 0, 1, \cdots, r$ implies that for all $\epsilon > 0$, there exists a number N and a partition $\{A_H^\epsilon | H \le G\}$ of $[1, N]$ such that for $H \le G$ and $i \in A_H^\epsilon$, $I(W_N^{(i)}) = \log\frac{|G|}{|H|} + O(\epsilon)$ and $Z^H(W_N^{(i)}) = O(\epsilon)$. Moreover, as $\epsilon \to 0$, $\frac{|A_H^\epsilon|}{N} \to p_H$ for some probabilities $p_H, H \le G$.

C. Encoding and Decoding

At the encoder, if $i \in A_H^\epsilon$ for some $H \le G$, the information symbol is chosen from the transversal T_H arbitrarily. Let $\mathcal{X}_N^\epsilon = \bigoplus_{H\le G} T_H^{A_H^\epsilon}$ be the set of all valid input sequences. As in the $\mathbb{Z}_{p^r}$ case, the message u_1^N is dithered with a uniformly distributed random vector $b_1^N \in \bigoplus_{H\le G} H^{A_H^\epsilon}$ revealed to both the encoder and the decoder. A message $v_1^N \in \mathcal{X}_N^\epsilon$ is encoded to the vector $x_1^N = (v_1^N + b_1^N)G_N$. Note that $u_1^N = v_1^N + b_1^N$ is uniformly distributed over G^N.

At the decoder, after observing the output vector y_1^N, for $H \le G$ and $i \in A_H^\epsilon$, use the following decoding rule:

$$\hat{u}_i = f_i(y_1^N, \hat{u}_1^{i-1}) = \arg\max_{g\in T_H} W_N^{(i)}(y_1^N, \hat{u}_1^{i-1}|g)$$

And finally, the message is recovered as $v_1^N = u_1^N - b_1^N$. The total number of valid input sequences is equal to

$$2^{NR} = \prod_{H\le G}|T_H|^{|A_H|} = \prod_{H\le G}\left(\frac{|G|}{|H|}\right)^{|A_H|}$$

Therefore the rate is equal to $R = \sum_{H\le G}\frac{|A_H|}{N}\log\frac{|G|}{|H|}$. On the other hand, since I^n is a martingale [5], we have $\mathbb{E}\{I^\infty\} = I^0$. Since $\mathbb{E}\{I^\infty\} = \sum_{H\le G} p_H \log\frac{|G|}{|H|}$ we observe that the rate R converges to the symmetric capacity I^0 as $\epsilon \to 0$. We will see in the next section that this rate is achievable.

D. Error Analysis

Let B_i be the event that the first error occurs when the decoder decodes the ith symbol:

$$\begin{aligned}
B_i = &\big\{(u_1^N, y_1^N) \in G^N \times \mathcal{Y}^N | u_1^{i-1} = f_1^{i-1}(u_1^N, y_1^N) \\
&, u_i \ne f_i(u_1^N, y_1^N)\big\} \\
&\subseteq \big\{(u_1^N, y_1^N) \in G^N \times \mathcal{Y}^N | u_i \ne f_i(u_1^N, y_1^N)\big\}
\end{aligned}$$

For $H \le G$ and $i \in A_H^\epsilon$, define

$$\begin{aligned}
E_i = &\Big\{(u_1^N, y_1^N) \in G^N \times \mathcal{Y}^N | W_N^{(i)}(y_1^N, u_1^{i-1}|u_i) \\
&\le W_N^{(i)}(y_1^N, u_1^{i-1}|\tilde{u}_i) \text{ for some } \tilde{u}_i \in T_H\Big\}
\end{aligned}$$

Lemma IV.3. *For $H \le G$ and $i \in A_H^\epsilon$, $P(E_i) \le qZ^H(W_N^{(i)}) = O(\epsilon)$.*

Proof: For $u_i, \tilde{u}_i \in T_H$, if $u_i \ne \tilde{u}_i$, then $u_i, \tilde{u}_i$ are not in the same coset of H and hence $u_i - \tilde{u}_i \notin H$.

$$\begin{aligned}
P(E_i) &= \sum_{u_1^N, y_1^N}\frac{1}{q^N}W_N(y_1^N, u_1^N)1_{E_i}(u_1^N, y_1^N) \\
&\le \sum_{u_1^N, y_1^N}\frac{1}{q^N}W_N(y_1^N, u_1^N)\sum_{\tilde{u}_i\ne u_i}\sqrt{\frac{W_N^{(i)}(y_1^N, u_1^{i-1}|\tilde{u}_i)}{W_N^{(i)}(y_1^N, u_1^{i-1}|u_i)}} \\
&= \sum_{u_i\in T_H}\sum_{\tilde{u}_i\in T_H, \tilde{u}_i\ne u_i}\frac{1}{q}Z_{\{u_i,\tilde{u}_i\}}(W_N^{(i)})
\end{aligned}$$

Note that for $d = u_i - \tilde{u}_i$, $Z_{\{u_i,\tilde{u}_i\}}(W_N^{(i)}) \le qZ_d(W_N^{(i)})$. Since $d \notin H$,

$$Z_{\{u_i,\tilde{u}_i\}}(W_N^{(i)}) \le q\left(Z^H(W_N^{(i)})\right) + O(\epsilon)$$

■

The probability of block error is given by $P(err) = \sum_{H\le G}\sum_{i\in A_H^\epsilon} P(E_i)$. Since $B_i \subseteq E_i$, we get

$$P(err) \le \sum_{H\le G}\sum_{i\in A_H^\epsilon} qZ^H(W_N^{(i)}) = O(\epsilon)$$

Therefore, the probability of block error goes to zero as $\epsilon \to 0$.

V. Examples

In this part, we consider two examples of channels over $\mathbb{Z}_4$. In the first example, the recursive equations for $I^n(W)$ is found and the multi-level polarization of the channel is observed. It is also shown that polar codes achieve the group capacity of this specific channel. The intent of the second example is to show that in general polar codes do not achieve the group capacity of channels. Note that since transversals are not subgroups, the polar code used above is not a group code. In order to obtain a group code based on polar codes we can use perfect channels and fix useless and partially perfect channels.

A. Example 1

We consider a channel with input $\mathbb{Z}_4$ where in the output there exists an erasure for each coset of each nontrivial subgroup. The channel is depicted in Figure 1.

Define $H_0 = \{0, 1, 2, 3\}$, $H_1 = \{0, 2\}$ and $H_2 = \{0\}$ and define $K_0 = \{1, 3\}$, $K_1 = \{2\}$ and $K_2 = \{0\}$.

For this channel we have:

$$\begin{aligned}
I^0 &= I(X;Y) = 2 - \epsilon - 2\lambda \\
I_2^0 &= I(X;Y|X \in H_1) = 1 - (\epsilon + \lambda) \\
(I_2')^0 &= I(X;Y|X \in 1 + H_1) = 1 - (\epsilon + \lambda) = I_2^0
\end{aligned}$$

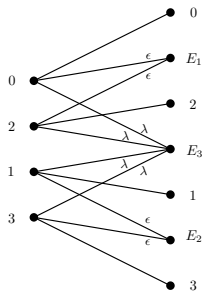

Fig. 1. Channel 1

The capacity of group codes over this symmetric channel is equal to [4]:

$$C = \min(I_4^0, (I_2 + I_2')^0) = \min(2 - \epsilon - 2\lambda, 2 - 2\epsilon - 2\lambda) \\ = 2 - 2\epsilon - 2\lambda$$

Given a sequence of bits $b_1 b_2 \cdots b_n$, let $N = 2^n$ and $i = (b_1 b_2 \cdots b_n)_{10}$ and define $I(W^{b_1 b_2 \cdots b_n}) = I(W_N^{(i)})$. we can find $I(W^{b_1 b_2 \cdots b_n})$ using the following recursion: Define $\epsilon_0 = \epsilon$ and $\lambda_0 = \lambda$. for $i = 1, \cdots, n$, if $b_i = 1$, let

$$\epsilon_i = \epsilon_{i-1}^2 + 2\epsilon_{i-1}\lambda_{i-1}$$
$$\lambda_i = \lambda_{i-1}^2$$

and if $b_i = 0$, let

$$\epsilon_i = 2\epsilon_{i-1} - \left(\epsilon_{i-1}^2 + 2\epsilon_{i-1}\lambda_{i-1}\right)$$
$$\lambda_i = 2\lambda_{i-1} - \lambda_{i-1}^2$$

Then we have $I(W^{b_1 b_2 \cdots b_n}) = 2 - \epsilon_n - 2\lambda_n$.
All possible cases for this channel are

- **case 0** $Z_1^\infty = Z_3^\infty = 1, Z_2^\infty = 1 \Rightarrow I^\infty = 0$
- **case 1** $Z_1^\infty = Z_3^\infty = 0, Z_2^\infty = 1 \Rightarrow I^\infty = 1$
- **case 2** $Z_1^\infty = Z_3^\infty = 0, Z_2^\infty = 0 \Rightarrow I^\infty = 2$

This result agrees with the asymptotic behavior of I^n predicted by the recursion. Figures 2 and 3 show the three level polarization of the mutual information as n grows.

Define $I_2(W^{b_1 b_2 \cdots b_n}) = I(W_N^{(i)} | X \in H_1)$ and $I_2'(W^{b_1 b_2 \cdots b_n}) = I(W_N^{(i)} | X \in 1 + H_1)$. For this channel, we can show that $I_2(W^{b_1 b_2 \cdots b_n}) = I_2'(W^{b_1 b_2 \cdots b_n}) = 1 - (\epsilon_n + \lambda_n)$ and conclude that $(I_2 + I_2')^n$ is a martingale. This observation provides us with an Ad-hoc way to find the probabilities p_t, $t = 0, 1, 2$ of the limit random variable I^∞ for this simple channel. We can show the following for the final states:

- **case 0** $\Rightarrow (I_2 + I_2')^\infty = 0$
- **case 1** $\Rightarrow (I_2 + I_2')^\infty = 0$
- **case 2** $\Rightarrow (I_2 + I_2')^\infty = 2$

Therefore we obtain the following three equations:

$$\mathbb{E}\{I_4^\infty\} = p_0 \cdot 0 + p_1 \cdot 1 + p_2 \cdot 2 = I_4^0 = 2 - \epsilon - 2\lambda$$
$$\mathbb{E}\{I_4^\infty\} = p_0 \cdot 0 + p_1 \cdot 0 + p_2 \cdot 2 = (I_2 + I_2')^0 = 2 - 2\epsilon - 2\lambda$$
$$p_0 + p_1 + p_2 = 1$$

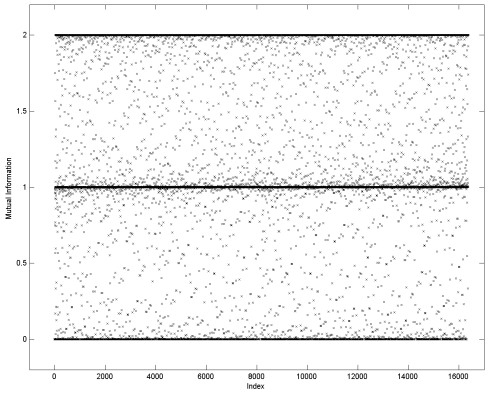

Fig. 2. The behavior of I^n for $N = 2^{12}$. The three solid lines are the three discrete values of I^∞ with positive probability.

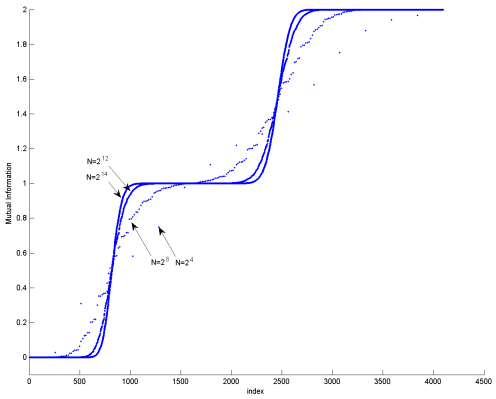

Fig. 3. The asymptotic behavior of I^n for $N = 2^4, 2^8, 2^{12}, 2^{14}$ when the data is sorted.

Solving this system of equations, we obtain:

$$p_2 = 1 - \epsilon - \lambda = C/2$$
$$p_1 = I_4^0 - (I_2 + I_2')^0$$
$$p_0 = 1 - \left(I_4^0 - (I_2 + I_2')^0/2\right)$$

We see that the fraction of perfect channels is equal to the group capacity of the channel and therefore, polar codes achieve the capacity of group codes for this channel.

B. Example 2

The channel is depicted in Figure 4. For this channel, when $\lambda = 0.2$ we have:

$$I^0 = I(X;Y) = 0.6390$$
$$(I_2^0 + I_2')^0 = 0.2161$$

The rate $R = \min(I_4^0, (I_2 + I_2')^0) = 0.2161$ is achievable using group codes over this channel [4].
For this channel we have three possible asymptotic case:

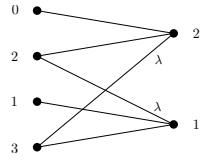

Fig. 4. Channel 2

- **case 0** $Z_1^\infty = 1, Z_2^\infty = 1 \Rightarrow I^\infty = 0, (I_2 + I_2')^\infty = 0$
- **case 1** $Z_1^\infty = 0, Z_2^\infty = 1 \Rightarrow I^\infty = 0, (I_2 + I_2')^\infty = 1$
- **case 2** $Z_1^\infty = 0, Z_2^\infty = 0 \Rightarrow I^\infty = 0, (I_2 + I_2')^\infty = 2$

Therefore we obtain the following three equations:

$$\begin{aligned}
\mathbb{E}\{I_4^\infty\} &= p_0 \cdot 0 + p_1 \cdot 1 + p_2 \cdot 2 \\
\mathbb{E}\{I_4^\infty\} &= p_0 \cdot 0 + p_1 \cdot 0 + p_2 \cdot 2 \\
p_0 + p_1 + p_2 &= 1
\end{aligned}$$

Therefore, the group capacity of polar codes over this channel is equal to $2p_2 = \mathbb{E}\{I_4^\infty\}$. In order to show that polar codes do not achieve the rate R, we show that $\mathbb{E}\{(I_2 + I_2')^\infty\} < R = (I_2 + I_2')^0$. For this channel we can find $\mathbb{E}\{(I_2 + I_2')^1\} = 0.2063$ which is strictly less than $(I_2 + I_2')^0$. The following lemma implies $\mathbb{E}\{I_2^\infty\} \leq \mathbb{E}\{I_2^1\} < R = (I_2 + I_2')^0$ and completes the proof.

Lemma V.1. *For a channel $(\mathbb{Z}_4, \mathcal{Y}, W)$, the process $(I_2 + I_2')^n, n = 0, 1, 2, \cdots$ is a supermartingale.*

Proof: Let X, U_1 and U_2 be uniform random variables over $\mathbb{Z}_4$ and let $X_1 = U_1 + U_2$ and $X_2 = U_2$. Let Y, Y_1 and Y_2 be the output of the channel when the input is X, X_1 and X_2 respectively. Note that the variable X can be represented by a pair $(\hat{X}, \tilde{X})$ where $\hat{X}$ takes value in $\{0, 1\}$ and $\tilde{X}$ takes value in $\{0, 2\}$. Define the variables $\hat{X}_1$, $\tilde{X}_1$, $\hat{X}_2$ and $\tilde{X}_2$ similarly. Note that

$$\begin{aligned}
\frac{1}{2}(I_2 + I_2') &= \frac{1}{2}\Big(I(\tilde{X}; Y|\hat{X} = 0) + I(\tilde{X}; Y|\hat{X} = 1)\Big) \\
&= I(\tilde{X}; Y|\hat{X}) = I(\hat{X}\tilde{X}; Y) - I(\hat{X}; Y) \\
&= I(X; Y) - I(\hat{X}; Y)
\end{aligned}$$

Since $I(X; Y)$ is a martingale it suffices to show that $I(\hat{X}; Y)$ is a submartingale. We have

$$\begin{aligned}
I(\hat{U}_2; W^+) &= I(\hat{U}_2; Y_1 Y_2 U_1) = I(\hat{U}_2; Y_1 Y_2 \hat{U}_1 \tilde{U}_1) \\
&= I(\hat{U}_2; Y_1 Y_2 \hat{U}_1) + I(\hat{U}_2; \tilde{U}_1 | Y_1 Y_2 \hat{U}_1) \\
&\geq I(\hat{U}_2; Y_1 Y_2 \hat{U}_1) \\
I(\hat{U}_1; W^-) &= I(\hat{U}_1; Y_1 Y_2)
\end{aligned}$$

Therefore,

$$\begin{aligned}
I(\hat{U}_2; W^+) + I(\hat{U}_1; W^-) &\geq I(\hat{U}_2; Y_1 Y_2 \hat{U}_1) + I(\hat{U}_1; Y_1 Y_2) \\
&= I(\hat{U}_1 \hat{U}_2; Y_1 Y_2) \\
&= I(\hat{X}_1 \hat{X}_2; Y_1 Y_2) = 2I(\hat{X}; W)
\end{aligned}$$

■

VI. Conclusion

It has been shown that the original construction of polar codes suffices to achieve the symmetric capacity of discrete memoryless channels with arbitrary input alphabet sizes. It is shown that in general, channel polarization happens in several levels so that some synthesized channels are partially perfect and there needs to be a modification of the coding scheme to exploit these channels. It is also shown that polar codes do not achieve the group capacity of arbitrary channels.

References

[1] E. Arikan. "Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels". *IEEE Transactions on Information Theory*, 55(7):3051–3073, 2009.

[2] R. Mori and T. Tanaka. "Channel Polarization on q-ary Discrete Memoryless Channels by Arbitrary Kernels". *Proceedings of IEEE International Symposium on Information Theory*, July 2010. Austin, TX.

[3] W. Park and A. Barg. Multilevel polarization for nonbinary codes and parallel channels. 2011. http://arxiv.org/abs/1107.4965.

[4] A. G. Sahebi and S. S. Pradhan. "On the Capacity of Abelian Group Codes Over Discrete Memoryless Channels". Feb. 2011. Online: http://arxiv.org/abs/1102.3243.

[5] E. Sasoglu, E. Telatar, and E. Arikan. "Polarization for arbitrary discrete memoryless channels". *IEEE Information Theory Workshop*, Dec. 2009. Lausanne, Switzerland.

VII. Appendices

A. Polar Codes Over Abelian Groups

Given a $k \times n$ matrix G_n of 0's and 1's, one can construct a group code as follows: Given any message tuple $u^k \in G^k$, encode it to $u^k \cdot G_n$. Where the elements of G_n determine whether an element of u^k appears as a summand in the encoded word or not. For example consider the generator matrix

$$G_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Then $u^4 \cdot G_4$ is defined as

$$[u_1 u_2 u_3 u_4] \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} u_1 + u_2 + u_3 + u_4 \\ u_3 + u_4 \\ u_2 + u_4 \\ u_4 \end{pmatrix}$$

Using this convention, we can define a group code based on a given binary matrix without actually defining a multiplication operation for the group.

B. Lemmas For $Z_d(W)$

Lemma VII.1. *For all $d \neq 0$, $Z_d^\infty(W)$ is $0 - 1$ valued.*

Proof: It has been proved in [5] for $d = \arg\max_{a \neq 0} Z_a(W)$. The proof for an arbitrary d is similar. ■

Lemma VII.2. *If $Z_d(W) \approx 1$, then for all $x \in G$ and $y \in \mathcal{Y}$*

$$W(y|x) \approx W(y|x + d)$$

in the sense that, if $Z_d(W) > 1-\epsilon$ *then*

$$|W(y|x) - W(y|x+d)| < O(\epsilon),\ \forall x \in G, \forall y \in \mathcal{Y}$$

Lemma VII.3. *For constants* $b \geq a \geq 0$*, with* $|b-a| > \delta > 0$*,*

$$\sqrt{ab} \leq \frac{a+b}{2} - f(\delta)$$

for some increasing function $f : [0,1] \to \mathbb{R}$ *which is* $O(\delta)$*.*

Proof: We need to find the solution to the maximization problem

$$\min_{|x-a|>\delta} \frac{a+x}{2} - \sqrt{ax}$$

where a is assumed to be a constant. We have

$$\frac{\partial}{\partial x}\left[\frac{a+x}{2} - \sqrt{ax} = \frac{1}{2} - \frac{a}{2\sqrt{ax}}\right] > 0$$

for all $x > a+\delta$. Therefore the minimum is attained by $x = a+\delta$. Therefore,

$$\frac{a+b}{2} - \sqrt{ab} \geq \frac{a+(a+\delta)}{2} - \sqrt{a(a+\delta)}$$

Next, we will minimize over $0 \leq a \leq 1$.

$$\frac{\partial}{\partial a}\left[a + \frac{\delta}{2} - \sqrt{a(a+\delta)}\right] < 0$$

Therefore, the minimum is attained for $a = 1$. We have shown that whenever $|b-a| > \delta$, $\sqrt{ab} \leq \frac{a+b}{2} - f(\delta)$ where $f(\delta) = 1 + \frac{\delta}{2} - \sqrt{1+\delta}$ is an increasing $O(\delta)$ function. ∎

Proof: (of Lemma VII.2) Let $f^{-1}(\epsilon) : [0, \frac{3}{2} - \sqrt{2}] \to \mathbb{R}$ be the local inverse of $f(\cdot)$. Note that $f^{-1}(\cdot)$ is also $O(\epsilon)$. We will prove that $O(\epsilon) = f^{-1}(q\epsilon)$ will satisfy the condition of the lemma. Assume for contradiction that for some $x \in G$ and some $y \in \mathcal{Y}$, $|W(y|x) - W(y|x+d)| \geq O(\epsilon)$. Then,

$$\begin{aligned} Z_d(W) &= \frac{1}{q}\sum_{x\in G}\sum_{y\in\mathcal{Y}} \sqrt{W(y|x)W(y|x+d)} \\ &\leq \frac{1}{q}\sum_{x\in G}\sum_{y\in\mathcal{Y}} \frac{W(y|x)+W(y|x+d)}{2} - \frac{f(f^{-1}(q\epsilon))}{q} \\ &= 1-\epsilon \end{aligned}$$

Which is a contradiction. We conclude the lemma. ∎

Lemma VII.4. *If* $Z_d(W) \approx 1$*, then for all* $y \in \mathcal{Y}$*, if* $x - x' \in \langle d\rangle = \{0, d, 2d, \cdots\}$ *then* $W(y|x) \approx W(y|x')$ *in the sense that, if* $Z_d(W) > 1-\epsilon$ *then for all* $x, x' \in G$ *such that* $x - x' \in \langle d\rangle$ *and for all* $y \in \mathcal{Y}$*,* $|W(y|x) - W(y|x')| < O(\epsilon)$*.*

Proof: Immediate from the above lemma. ∎

Lemma VII.5. *If* $Z_d(W) \approx 0$*, then for all* $x \in G$ *and* $y \in \mathcal{Y}$*, either* $W(y|x) \approx 0$ *or* $W(y|x+d) \approx 0$*. In the sense that, if* $Z_d(W) < \epsilon$ *then either* $W(y|x) \leq O(\epsilon)$ *or* $W(y|x+d) \leq O(\epsilon)$*.*

Proof: We will show that $O(\epsilon) = \sqrt{q}\epsilon$ will satisfy this condition. For all $x \in G$ and $y \in \mathcal{Y}$ we have,

$$\begin{aligned} W(y|x)W(y|x+d) &\leq \left(\sum_{x\in G}\sum_{y\in\mathcal{Y}} \sqrt{W(y|x)W(y|x+d)}\right)^2 \\ &= \left(qZ_d(W)\right)^2 \leq q^2\epsilon^2 \end{aligned}$$

If both $W(y|x)$ and $W(y|x+d)$ are greater than $O(\epsilon)$, we will get a contradiction. ∎

Lemma VII.6. *If for all* $x \in G$ *and* $y \in \mathcal{Y}$*,* $W(y|x) \approx W(y|x+d)$*, then* $Z_d(W) \approx 1$*. In the sense that, if*

$$|W(y|x) - W(y|x+d)| < \epsilon,\ \forall x \in G, \forall y \in \mathcal{Y}$$

then $Z_d(W) > 1 - O(\epsilon)$*.*

Lemma VII.7. *For constants* $b \geq a \geq 0$*, with* $|b-a| < \delta$*,*

$$\sqrt{ab} \geq \frac{a+b}{2} - \frac{\delta}{2}$$

Proof: Similar to above. ∎

Proof: (of Lemma VII.6) We will show that $O(\epsilon) = \frac{\epsilon|\mathcal{Y}|}{2}$ satisfies this condition.

$$\begin{aligned} 1 - Z_d(W) &= 1 - \frac{1}{q}\sum_{x\in G}\sum_{y\in\mathcal{Y}} \sqrt{W(y|x)W(y|x+d)} \\ &= \frac{1}{q}\sum_{x\in G}\sum_{y\in\mathcal{Y}} \Big(\frac{W(y|x)+W(y|x+d)}{2} \\ &\quad - \sqrt{W(y|x)W(y|x+d)}\Big) \\ &\leq \frac{1}{q}\sum_{x\in G}\sum_{y\in\mathcal{Y}} \frac{\epsilon}{2} = \frac{\epsilon|\mathcal{Y}|}{2} \end{aligned}$$

∎

Lemma VII.8. *If for all* $y \in \mathcal{Y}$ *and* $x - x' \in \langle d\rangle$*,* $W(y|x) \approx W(y|x+d)$*, then* $Z_d(W) \approx 1$*.*

Proof: Immediate from Lemma VII.6. ∎

Lemma VII.9. *If for all* $x \in G$ *and* $y \in \mathcal{Y}$*, either* $W(y|x) \approx 0$ *or* $W(y|x+d) \approx 0$*, then* $Z_d(W) \approx 0$*. In the sense that, if for all* $x \in G$ *and* $y \in \mathcal{Y}$*, either* $W(y|x) \leq \epsilon$ *or* $W(y|x+d) \leq \epsilon$*, then* $Z_d(W) < O(\epsilon)$*.*

Proof: We will show that $O(\epsilon) = \epsilon|\mathcal{Y}|$ satisfies this condition.

$$\begin{aligned} Z_d(W) &= \frac{1}{q}\sum_{x\in G}\sum_{y\in\mathcal{Y}} \sqrt{W(y|x)W(y|x+d)} \\ &\leq \frac{1}{q}\sum_{x\in G}\sum_{y\in\mathcal{Y}} \epsilon \\ &= \epsilon|\mathcal{Y}| \end{aligned}$$

∎