Fifty-first Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 2 - 3, 2013

Polar Codes are Optimal for Write-Efficient Memories

Qing Li
Department of Computer Science and Engineering
Texas A & M University
College Station, TX 77840
qingli@cse.tamu.edu

Anxiao (Andrew) Jiang
Department of Computer Science and Engineering
Texas A & M University
College Station, TX 77840
ajiang@cse.tamu.edu

***Abstract*—Write-efficient memories (WEM) [1] were introduced by Ahswede and Zhang as a model for storing and updating information on a rewritable medium with constraints. A coding scheme for WEM using recently proposed polar codes is presented. The coding scheme achieves rewriting capacity, and the rewriting and decoding operation can be done in time $O(N \log N)$, where N is the length of the codeword.**

I. Introduction

Write-efficient memories (WEM) are models for storing and updating information on a rewritable medium with constraints. WEM is widely used in data storage area: in flash memories, write-once memories (WOM) [11], and the recently proposed compressed rank modulation (CRM) [9] are examples of WEM; codes based on WEM were proposed for phase change memories [8]. The recently proposed scheme, that polar codes are constructed for WOM codes achieving capacity [3], motivates us to construct codes for WEM.

A. WEM with a maximal rewriting cost constraint

Let $\mathcal{X} = \{0, 1, ..., q-1\}$ be the storage alphabet. $\mathcal{R}_+ = [0, +\infty)$, and $\varphi : \mathcal{X} \times \mathcal{X} \to \mathcal{R}_+$ is the rewriting cost function. Suppose that a memory consists of N cells. Given one cell state, $x_0^{N-1} \stackrel{def}{=} (x_0, x_1, ..., x_{N-1}) \in \mathcal{X}^N$, and another cell state, $y_0^{N-1} \in \mathcal{X}^N$, the rewriting cost of changing x_0^{N-1} to y_0^{N-1} is $\varphi(x_0^{N-1}, y_0^{N-1}) = \sum_{i=0}^{N-1} \varphi(x_i, y_i)$.

Let $M \in \mathbb{N}$, and $\mathcal{D} = \{0, 1, ..., M-1\}$. Define the decoding function, $\mathbf{D} : \mathcal{X}^N \to \mathcal{D}$, i.e., $\mathbf{D}(x_0^{N-1}) = i$. Given the current cell state x_0^{N-1}, and data to rewrite, j, the rewriting function is defined as $\mathbf{R} : \mathcal{X}^N \times \mathcal{D} \to \mathcal{X}^N$ such that $\mathbf{D}(\mathbf{R}(x_0^{N-1}, j)) = j$.

Definition 1. [5] An (N, M, q, d) WEM code is a collection of subsets, $\mathcal{C} = \{\mathcal{C}_i : i \in \mathcal{D}\}$, where $\mathcal{C}_i \subseteq \mathcal{X}^N$, and $\forall x_0^{N-1} \in \mathcal{C}_i$ $\mathbf{D}(x_0^{N-1}) = i$, such that

- $\forall i \neq j$, $\mathcal{C}_i \bigcap \mathcal{C}_j = \emptyset$.
- $\forall j$, $\forall x_0^{N-1} \in \mathcal{C}$, $\exists y_0^{N-1} = \mathbf{R}(x_0^{N-1}, j)$ such that $\varphi(x_0^{N-1}, y_0^{N-1}) \leq Nd$.

The rewriting rate of an (N, M, q, d) WEM code is $\mathcal{R} = \frac{\log_2 M}{N}$, and the *rewriting capacity function*, $\mathcal{R}(q, d)$, is the largest d-admissible rate when $N \to \infty$.

Let $\mathcal{P}(\mathcal{X} \times \mathcal{X})$ be the set of joint probability distributions over $\mathcal{X} \times \mathcal{X}$. For a pair of random variables $(X, Y) \in (\mathcal{X}, \mathcal{X})$, let P_{XY} denote the joint probability distribution, P_Y denotes the marginal distribution, and $P_{Y|X}$ denotes the conditional probability distribution. $E(\cdot)$ denotes the expectation operator. If X is uniformly distributed over $\{0, 1, ..., q-1\}$, denote it as $X \sim U(q)$.

Let $\mathcal{P}(q, d) = \{P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{X}) : P_X = P_Y, E(\varphi(X, Y)) \leq d\}$. $\mathcal{R}(q, d)$ is determined as [5]: $\mathcal{R}(q, d) = \max_{P_{XY} \in \mathcal{P}(q,d)} H(Y|X)$.

For WOM codes, the cell state can only increase but not decrease. WOM codes are special cases of WEM codes: for a WOM cell if we update it from $x \in \mathcal{X}$ to $y \in \mathcal{X}$, the cost is measured by $\varphi(x, y) = 0$ if $y \geq x$, and ∞ otherwise. Therefore, WOM codes are such WEM codes with $\varphi(\cdot)$ defined previously, and d is equal to 0.

In this paper, we focus on *symmetric rewriting capacity function* $\mathcal{R}^s(q, d)$:

Definition 2. For $X, Y \in \mathcal{X}$ with P_X, P_Y and P_{XY}, and $\varphi : \mathcal{X} \times \mathcal{X} \to \mathcal{R}_+$, $\mathcal{R}^s(q, d) = \max_{P_{XY} \in \mathcal{P}^s(q,d)} H(Y|X)$, where $\mathcal{P}^s(q, d) \stackrel{def}{=} \{P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{X}) : P_X = P_Y, X \sim U(q), E(\varphi(X, Y)) \leq d\}$.

We present an example of WEM with $\mathcal{R}^s(q, d)$ below. $\mathcal{S}_{q,m}$ is the set of $\frac{(mq)!}{m!^q}$ permutations over $\{\overbrace{1, 1, ..., 1}^{m}, ..., \overbrace{q, q, ..., q}^{m}\}$. We abuse the notation of $u_0^{qm-1} \stackrel{def}{=} [u_0, u_1, ..., u_{qm-1}]$ to denote an element of $\mathcal{S}_{q,m}$, which denotes the mapping $i \to u_i$.

Example 3. A rewriting code for CRM with a maximal rewriting cost constraint, (q,m,M,d), is defined by replacing $\mathcal{X}^N$ by $\mathcal{S}_{q,m}$, $\varphi(\cdot)$ by Chebyshev distance between $u_0^{qm-1}, v_0^{qm-1} \in \mathcal{S}_{q,m}$, $d_\infty(u_0^{qm-1}, v_0^{qm-1}) \overset{def}{=} \max_{j\in\{0,1,...,qm-1\}} |u_j - v_j|$, and Nd by d in definition 1.

The above rewriting code is actually an instance of WEM: for $x, y \in \mathcal{X}$, let $\varphi(x,y) = 0$ if $|x-y| \le d$, and ∞ otherwise. Now the (q,m,M,d) CRM is an $(qm, M, q, 0)$ WEM with $\mathcal{X}^N$ replaced by $\mathcal{S}_{q,m}$, and $\varphi(\cdot)$ is defined previously.

B. WEM with an average rewriting cost constraint

With deterministic $\mathbf{D}(x_0^{N-1})$ and $\mathbf{R}(x_0^{N-1}, i)$, suppose that message sequences $M_1, M_2, ..., M_t$ $(t \to \infty)$ are written into the memory medium, where $M_i \in \mathcal{D}$ is uniformly distributed, represented by $x_0^{N-1}(i)$, and it forms a *markov chain*. Its transition probability matrix $\mu(y_0^{N-1}|x_0^{N-1})_{y_0^{N-1}, x_0^{N-1} \in \mathcal{X}^N}$ is given by

$$\mu(y_0^{N-1}|x_0^{N-1}) = \begin{cases} \frac{1}{M} & \text{if } \exists i \ s.t. y_0^{N-1} = \mathbf{R}(x_0^{N-1}, i), \\ 0 & \text{otherwise.} \end{cases}$$

Denote the stationary distribution of x_0^{N-1} by $\pi(x_0^{N-1})$.

The average rewriting cost $\bar{D}$ is determined as follows:

$$\begin{aligned} \bar{D} &= \lim_{t\to\infty} \frac{1}{Nt} \sum_{i=1}^{t} E(\varphi(x_0^{N-1}(i), x_0^{N-1}(i+1))), \\ &= \frac{1}{N} \sum_{x_0^{N-1}} \pi(x_0^{N-1}) \sum_{y_0^{N-1}} \mu(y_0^{N-1}|x_0^{N-1}) \\ &\quad \varphi(x_0^{N-1}, y_0^{N-1}), \\ &= \sum_{x_0^{N-1}} \pi(x_0^{N-1}) \sum_j \bar{D}_j(x_0^{N-1}), \end{aligned} \tag{1}$$

where $\bar{D}_j(x_0^{N-1})$ is the average rewriting cost of updating from x_0^{N-1} to data j.

Definition 4. An $(N, M, q, d)_{ave}$ WEM code is a collection of subsets, $\mathcal{C} = \{\mathcal{C}_i : i \in \mathcal{D}\}$, where $\mathcal{C}_i \subseteq \mathcal{X}^N$, and $\forall x_0^{N-1} \in \mathcal{C}_i$ $\mathbf{D}(x_0^{N-1}) = i$, such that

- $\forall i \neq j, \mathcal{C}_i \bigcap \mathcal{C}_j = \emptyset$;
- The average rewriting cost $\bar{D} \le d$.

The rewriting rate of an $(N, M, q, d)_{ave}$ code is $\mathcal{R}_{ave} = \frac{\log_2 M}{N}$, and its rewriting capacity function, $\mathcal{R}_{ave}(q,d)$, is the largest d-admissible rate when $N \to \infty$. It is proved that $\mathcal{R}_{ave}(q,d) = \mathcal{R}(q,d)$ [1]. Similarly, we focus on the symmetric rewriting capacity function, $\mathcal{R}^s_{ave}(q,d)$, as defined in definition 2.

C. The outline

The connection between rate-distortion theorem and rewriting capacity theorem is presented in section II; The binary polar WEM codes with an average rewriting cost constraint and a maximal rewriting cost constraint are presented in subsection A and B of section III, respectively; The q-ary polar WEM codes, based on the recently proposed q-ary polar codes [10], are presented in subsection A and B of section IV for an average rewriting cost constraint and a maximal rewriting cost constraint, respectively; The conclusion is obtained in section V.

II. Lossy source coding and its duality with WEM

In this section, we present briefly background of lossy source coding and its duality with WEM, which inspires code constructions for WEM.

Let $\mathcal{X}$ also denote the variable space, and $\mathcal{Y}$ denotes the reconstruction space. Let $d : \mathcal{Y} \times \mathcal{X} \to \mathcal{R}_+$ denote the distortion function, and the distortion among a vector x_0^{N-1} and its reconstructed vector y_0^{N-1} is $d(x_0^{N-1}, y_0^{N-1}) = \frac{1}{N}\sum_{i=0}^{N-1} d(x_i, y_i)$.

A (q^{NR}, N) rate distortion code consists of a encoding function $f_N : \mathcal{X}^N \to \{0, 1, ..., q^{NR} - 1\}$ and a reproduction function $g_N : \{0, 1, ..., q^{NR} - 1\} \to \mathcal{Y}^N$. The associated distortion is defined as $E(d(X_0^{N-1}, g_N(f_N(X_0^{N-1}))))$, where the expectation is with respect to the probability distribution on $\mathcal{X}^N$. $R(q,D)$ is the infimum of rates R such that $E(d(X_0^{N-1}, g_N(f_N(X_0^{N-1}))))$ is at most D as $N \to \infty$.

Let $P(q,D) \overset{def}{=} \{P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : E(d(X,Y)) \le D\}$, and $R(q,D)$ is $\min_{P_{XY}\in P(q,D)} I(X;Y)$ [4].

We focus on the *double symmetric rate-distortion* for $(X,Y) \in (\mathcal{X} \times \mathcal{X})$, and $d(x,y)$, $R^S(q,D)$, which is $R^s(q,D) = \min_{P_{XY}\in P^s(q,D)} I(Y;X)$, where $P^s(q,D) \overset{def}{=} \{P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{X}) : P_X = P_Y, X \sim U(q), E(d(X,Y)) \le D\}$.

The duality between $\mathcal{R}^s(q,D)$ and $R^s(q,D)$ is captured by the following lemma, the proof of which is ommited due to being straightforward.

Lemma 5. With the same $d(\cdot)$ and $\varphi(\cdot)$,

$$\mathcal{R}^s(q,D) + R^s(q,D) = \log_2 q. \tag{2}$$

III. Polar codes are optimal for binary WEM

Inspired by lemma 5, we show that polar codes can be used to construct binary WEM codes with $\mathcal{R}^s(2,D)$ in a way related to the code construction for lossy source coding of [7] in this section.

A. A code construction for binary WEM with an average rewriting cost constraint

1) Background on polar codes [2]: Let $W : \{0,1\} \to \mathcal{Y}$ be a binary-input discrete memoryless channel for some output alphabet $\mathcal{Y}$. Let $I(W) \in [0,1]$ denote the mutual information between the input and output of W with a uniform distribution on the input. Let G_N denote n-th Kronecker product of $\left(\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right)$. Let the Bhattacharyya parameter $Z(W) = \sum_{y\in\mathcal{X}} \sqrt{W_{Y|X}(y|0)W_{Y|X}(y|1)}$.

The polar code, $C_N(F, u_F)$, $\forall F \subseteq \{0,1,...,N-1\}$, u_F denotes the subvector $u_i : i \in F$, and $u_F \in \{0,1\}^{|F|}$, is a linear code given by $C_N(F, u_F) = \{x_0^{N-1} = u_0^{N-1}G_N : u_{F^c} \in \{0,1\}^{|F^c|}\}$. The polar code ensemble, $C_N(F)$, $\forall F \subseteq \{0,1,...,N-1\}$, is $C_N(F) = \{C_N(F,u_F), \forall u_F \in \{0,1\}^{|F|}\}$.

The secret of polar codes achieving $I(W)$ lies in how to select F: define $W_N^{(i)} : \{0,1\} \to \mathcal{Y}^N \times \{0,1\}^i$ as sub-channel i with input u_i, output (y_0^{N-1}, u_0^{i-1}) and transition probabilities $W_N^{(i)}(y_0^{N-1}, u_0^{i-1}|u_i) \overset{def}{=} \frac{1}{2^{N-1}} \sum_{u_{i+1}^{N-1}} W^N(y_0^{N-1}|u_0^{N-1})$, where $W^N(y_0^{N-1}|u_0^{N-1}) \overset{def}{=} \prod_{i=0}^{N-1} W(y_i|(u_0^{N-1}G_N)_i)$, and $(u_0^{N-1}G_N)_i$ denotes the i-th element of $u_0^{N-1}G_N$; The fraction of $\{W_N^{(i)}\}$ that are approaching noiseless, i.e., $Z(W_N^{(i)}) \leq 2^{-N^\beta}$ for $0 \leq \beta \leq \frac{1}{2}$, approaches $I(W)$; The F is chozen as indecies with large $Z(W_N^{(i)})$, that is $F \overset{def}{=} \{i \in \{0,1...,N-1\} : Z(W_N^{(i)}) \geq 2^{-N^\beta}\}$ for $\beta \leq 1/2$.

The encoding is done by a linear transformation, and the decoding is done by successive cancellation (SC).

2) Polar codes on lossy source coding: We sketch the binary polar code construction for lossy source coding [7] as follows.

Note that $R^s(2, D)$ can be obtained through the following optimization function:

$$\begin{aligned} \min \quad &: \quad I(Y;X), \\ \text{s.t.} \quad &: \quad \sum_x \frac{1}{2}P(y|x) = \sum_y \frac{1}{2}P(x|y) = \frac{1}{2}, \\ & \quad\quad \sum_x \sum_y \frac{1}{2}P(y|x)d(y,x) \leq D. \end{aligned} \tag{3}$$

Let $P^*(y|x)$ be the probability distribution minimizing the objective function of (3). $P^*(y|x)$ plays the role of a channel. By convention, we call $P^*(y|x)$ as *test channel*, and denote it as $W(y|x)$.

Now, we construct the source code with $R^s(2, D)$ using the polar code for $W(y|x)$, and denote the rate of the source code by R: set F as $N(1-R)$ sub-channels with the highest $Z(W_N^{(i)})$, set F^c as the remaining NR sub-channels, and set u_F to an arbitrary value.

A source codeword y_0^{N-1} is mapped to a codeword $x_0^{N-1} \in C_N(F, u_F)$, and x_0^{N-1} is described by the index $u_{F^c} = (x_0^{N-1}G_N^{-1})_{F^c}$.

The reproduction process is done as follows: we do SC encoding scheme, $\hat{u}_0^{N-1} = \hat{U}(y_0^{N-1}, u_F)$, that is for each k in the range 0 till $N-1$:

1) If $k \in F$, set $\hat{u}_k = u_k$;
2) Else, set $\hat{u}_k = m$ with the posterior $P(m|\hat{u}_0^{i-1}, y_0^{N-1})$.

The reproduction codeword is $\hat{u}_0^{N-1}G_N$.

Thus, the average distortion $D_N(F, u_F)$ (over the source codeword y_0^{N-1} and the encoder randomness for the code $C_N(F, u_F)$) is :

$$\sum_{y_0^{N-1}} P(y_0^{N-1}) \sum_{\hat{u}_{F^c}} \prod_{i\in F^c} P(\hat{u}_i|\hat{u}_0^{i-1}, y_0^{N-1})$$

$$d(y_0^{N-1}, \hat{u}_0^{N-1}G_N),$$

where $\hat{u}_F = u_F$.

The expectation of $D_N(F, u_F)$ over the uniform choice of u_F, $D_N(F)$, is $D_N(F) = \sum_{u_F} \frac{1}{2^{|F|}} D_N(F, u_F)$. Let $Q_{U_0^{N-1}, Y_0^{N-1}}$ denote the distribution defined by $Q_{Y_0^{N-1}}(y_0^{N-1}) = P(y_0^{N-1})$, and $Q_{U_0^{N-1}|Y_0^{N-1}}$ by

$$Q(u_i|u_0^{i-1}, y_0^{N-1}) = \begin{cases} \frac{1}{2}, & \text{if } i \in F, \\ P(u_i|u_0^{i-1}, y_0^{N-1}), & \text{otherwise.} \end{cases}$$

Thus, $D_N(F)$ is equivalent to $E_Q(d(y_0^{N-1}, u_0^{N-1}G_N))$, where $E_Q(\cdot)$ denotes the expectation with respect to the distribution $Q_{U_0^{N-1}, Y_0^{N-1}}$.

It is proved that $D_N(F) \leq D + O(2^{-N^\beta})$ for $0 \leq \beta \leq \frac{1}{2}$, the rate of the above scheme is $R = \frac{|F^c|}{N}$, and polar codes achieve the rate-distortion bound by Theorem 3 of [7].

On the other hand, Theorem 2 of [3] further states the strong converse result of the rate distortion theory. More precisely, if y_0^{N-1} is uniformly distributed over $\{0,1\}^N$, then $\forall \delta > 0$, $0 < \beta < \frac{1}{2}$, N sufficiently large, and with the above SC encoding process and the induced Q, $Q(d(u_0^{N-1}G_N, y_0^{N-1}) \geq D + \delta) < 2^{-N^\beta}$. That is, for $\forall y_0^{N-1}$, the above reproduction process obtains $x_0^{N-1} = u_0^{N-1}G_N$ such that the distortion $d(x_0^{N-1}, y_0^{N-1})$ is less than D almost by sure.

3) The code construction: We focus on the code construction with *symmetric rewriting cost function*, which satisfies $\forall x, y, z \in \{0,1\}, \varphi(x,y) = \varphi(x+z, y+z)$, where $+$ is over GF(2).

To construct codes for WEM with $\mathcal{R}^s(2,D)$, we utilize its related form $R^s(2,D)$ in (2) and the test channel $W(y|x)$ for $R^s(2,D)$. Note that $W(y|x)$ is a binary symmetric channel.

The code construction for $(N,M,2,D)_{ave}$ with rate $\mathcal{R}$ is presented in Algorithm III.1:

Algorithm III.1 A code construction for $(N,M,2,D)_{ave}$ WEM

1: Set F as $N\mathcal{R}$ sub-channels with the highest $Z(W_N^{(i)})$, and set F^c as the remaining $N(1-\mathcal{R})$ sub-channels.
2: The $(N,M,2,D)_{ave}$ code is $\mathcal{C} = \{\mathcal{C}_i : \mathcal{C}_i = C_N(F, u_F(i))\}$, where $u_F(i)$ is the binary representation form of i for $i \in \{0,1,...,M-1\}$.

Clearly, since G_N is of full rank [2], $\forall u_F(i) \neq u_F(j)$, $C_N(F,u_F(i)) \bigcap C_N(F,u_F(j)) = \emptyset$.

The rewriting operation and the decoding operation are defined in Algorithm III.2 and Algorithm III.3, respectively, where the dither g_0^{N-1} is inspired by [3].

Algorithm III.2 The rewriting operation $y_0^{N-1} = \mathbf{R}(x_0^{N-1}, i)$.

1: Let $v_0^{N-1} = x_0^{N-1} + g_0^{N-1}$, where g_0^{N-1} is known both to the decoding function and to the rewriting function, it is chosen such that v_0^{N-1} is uniformly distributed over $\{0,1\}^N$, and $+$ is over GF(2).
2: SC encoding v_0^{N-1}, and this results $u_0^{N-1} = \hat{U}(v_0^{N-1}, u_F(i))$ and $\hat{y}_0^{N-1} = u_0^{N-1}G_N$.
3: $y_0^{N-1} = \hat{y}_0^{N-1} + g_0^{N-1}$.

Algorithm III.3 The decoding operation $u_F(i) = \mathbf{D}(x_0^{N-1})$.

1: $y_0^{N-1} = x_0^{N-1} + g_0^{N-1}$.
2: $u_F(i) = (y_0^{N-1}G_N^{-1})_F$.

Lemma 6. $\mathbf{D}(\mathbf{R}(y_0^{N-1}, i)) = i$ holds for each rewriting.

Proof: From the rewriting operation, $y_0^{N-1} = \hat{y}_0^{N-1} + g_0^{N-1} = u_0^{N-1}G_N + g_0^{N-1} = \hat{U}(v_0^{N-1}, u_F(i))G_N + g_0^{N-1}$, from the decoding function $\hat{U}(v_0^{N-1}, u_F(i))G_N + g_0^{N-1} + g_0^{N-1}$, which is $\hat{U}(v_0^{N-1}, u_F(i))G_N$, thus the decoding result is i. ■

4) The average rewriting cost analysis: From $y_0^{N-1} = \mathbf{R}(x_0^{N-1}, i)$, we know that $y_0^{N-1} = \hat{U}(v_0^{N-1}, u_F(i))G_N + g_0^{N-1} = \hat{U}(x_0^{N-1} + g_0^{N-1}, u_F(i))G_N + g_0^{N-1}$, thus $\varphi(x_0^{N-1}, y_0^{N-1})$ is $\varphi(x_0^{N-1}, \hat{U}(x_0^{N-1} + g_0^{N-1}, u_F(i))G_N + g_0^{N-1})$, which is $\varphi(x_0^{N-1} + g_0^{N-1}, \hat{U}(x_0^{N-1} + g_0^{N-1}, u_F(i))G_N)$ due to $\varphi(\cdot)$ being symmetric. Denote $w_0^{N-1} = x_0^{N-1} + g_0^{N-1}$, thus $\varphi(x_0^{N-1}, y_0^{N-1}) = \varphi(w_0^{N-1}, \hat{U}(w_0^{N-1}, u_F(i))G_N)$.

The average rewriting cost $\bar{D}$

$$\begin{aligned} &= \lim_{t\to\infty} \frac{1}{Nt} \sum_{i=1}^{t} E(\varphi(x_0^{N-1}(i), x_0^{N-1}(i+1))), \\ &= \lim_{t\to\infty} \frac{1}{Nt} \sum_{i=1}^{t} E(\varphi(w_0^{N-1}, \\ &\quad \hat{U}(w_0^{N-1}, u_F(M_{i+1}))G_N)), \\ &= \sum_{w_0^{N-1}} \pi(w_0^{N-1}) \sum_j \bar{D}_j(w_0^{N-1}). \end{aligned} \tag{4}$$

Let us focus on $\bar{D}_j(w_0^{N-1})$, which is the average (in this case, over the probability of rewriting to data j and the randomness of the encoder) rewriting cost of updating w_0^{N-1} to a codeword representing j. Thus $\bar{D}_j(w_0^{N-1}) =$

$$\frac{1}{2^{|F|}} \sum_{u_{F^c}} \prod_{i\in F^c} P(u_i|u_0^{i-1}, w_0^{N-1})\varphi(w_0^{N-1}, u_0^{N-1}G_N).$$

Therefore, interpreting $\varphi(\cdot)$ as $d(\cdot)$, $\bar{D}$ is actually the average distortion over the ensemble $C_N(F)$, $D_N(F)$.

The following lemma from [7] is to bound $\bar{D}$:

Lemma 7. [7] Let $\beta < \frac{1}{2}$ be a constant and let $\sigma_N = \frac{1}{2N}2^{-N^\beta}$. When the polar code for the source code with $R^s(2,D)$ is constructed with F:

$$F = \{i \in \{0,1,...,N-1\} : Z(W_N^{(i)}) \geq 1 - 2\sigma_N^2\},$$

then $D_N(F) \leq D + O(2^{-N^\beta})$, where D is the average rewriting cost constraint.

Therefore, with the same β, σ_N, F, and polar code ensemble $C_N(F)$, $\bar{D} \leq D + O(2^{-N^\beta})$.

According to [2], $\lim_{N=2^n, n\to\infty} \frac{|F^c|}{N} =$

$$\begin{aligned} &\lim_{N\to\infty} \frac{|\{i \in \{0,1,...,N\} : Z(W_N^{(i)}) \leq 2^{-2^{n\beta}}\}|}{N} \\ &= I(W) = R^s(2,D), \end{aligned}$$

thus this implies that for N sufficiently large $\exists$ a set F such that $\frac{|F^c|}{N} \geq R^s(2,D) - \epsilon$, $\forall \epsilon > 0$. In other words, the rate of the constructed WEM code, $\mathcal{R} = \frac{|F|}{N} = 1 - \frac{|F^c|}{N} \leq \mathcal{R}^s(2,D) + \epsilon$.

The complexity of the decoding and the rewriting operation is of the order $O(N \log N)$ according to [2].

We conclude the theoretical performance of the above polar WEM code as follows:

Theorem 8. For a binary symmetric rewriting cost function $\varphi : \mathcal{X} \times \mathcal{X} \to \mathcal{R}_{+}$. Fix a rewriting cost D and $0 < \beta < \frac{1}{2}$. For any rate $\mathcal{R} < \mathcal{R}^s(2, D)$, there exists a sequence of polar WEM codes of length N and rate $R \leq \mathcal{R}$, so that under the above rewriting operation, $\bar{D}$ satisfies $\bar{D} \leq D + O(2^{-N^{\beta}})$. The decoding and rewriting operation complexity of the codes is $O(N \log N)$.

B. A code construction for binary WEM with a maximal rewriting cost constraint

The code construction, the rewriting operation and the decoding operation are exactly the same as Algorithm III.1, Algorithm III.2, and Algorithm III.3, respectively.

The rewriting capacity is guaranteed by Lemma 5, the decoding and rewriting operation complexity is the same as polar codes, and the rewriting cost is obtained due to the strong converse result of rate distortion theory, i.e., Theorem 2 of [3]. Thus, we have:

Theorem 9. For a binary symmetric rewriting cost function $\varphi : \mathcal{X} \times \mathcal{X} \to \mathcal{R}_{+}$. Fix a rewriting cost D, δ, and $0 < \beta < \frac{1}{2}$. For any rate $\mathcal{R} < \mathcal{R}^s(2, D)$, there exists a sequence of polar WEM codes of length N and rate $R \leq \mathcal{R}$, so that under the above rewriting operation and the induced probability distribution Q, the rewriting cost between a current codeword $\forall y_0^{N-1}$ and its updated codeword x_0^{N-1} satisfies $Q(\varphi(y_0^{N-1}, x_0^{N-1}) \geq D + \delta) < 2^{-N^{\beta}}$. The decoding and rewriting operation complexity of the codes is $O(N \log N)$.

IV. Polar codes are optimal for q-ary WEM, $q = 2^r$

In this section, we extend the previous binary scheme to q-ary WEM ($q = 2^r$), considering the length of polar codes, which is $N = 2^n$.

A. A code construction for q-ary WEM with an average rewriting cost constraint, $q = 2^r$

1) Background of q-ary polar codes, $q = 2^r$ [10]: The storage alphabet is $\mathcal{X}$, $|\mathcal{X}| = q$, and for $x \in \mathcal{X}$, $(x_0, x_1, ..., x_{r-1})$ is its binary representation. Let $W : \mathcal{X} \to \mathcal{Y}$ be a discrete memoryless channel. $I(W)$ is $\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \frac{1}{q} W(y|x) \log_2 \frac{W(y|x)}{\sum_{x' \in \mathcal{X}} \frac{1}{q} W(y|x')}$.

The sub-channel i, $W_N^{(i)}$, is defined by $W_N^{(i)}(y_0^{N-1}, u_0^{i-1}|u_i) \overset{def}{=} \frac{1}{q^{N-1}} \sum_{u_{i+1}^{N-1}} W^N(y_0^{N-1}|u_0^{N-1})$, where $W^N(y_0^{N-1}|u_0^{N-1}) \overset{def}{=} \prod_{i=0}^{N-1} W(y_i|(u_0^{N-1} G_N)_i)$.

Consider the following *bit channel* and its Bhattacharyya: Fix $k \in \{0, 1, ..., r\}$, and $W^{[r-k]}(y|u) = \frac{1}{2^k} \sum_{x : x_k^{r-1} = u} W(y|x)$, where $y \in \mathcal{Y}, u \in \{0,1\}^{r-k}$, and $x = (x_0, x_1, ..., x_{r-1}) \in \mathcal{X}$; Define $Z(W_{\{x,x'\}}) = \sum_{y \in \mathcal{Y}} \sqrt{W(y|x) W(y|x')}$, let $Z_v(W) = \frac{1}{2^r} \sum_{x \in \mathcal{X}} Z(W_{\{x, x+v\}})$ for $v \in \mathcal{X} \backslash \{0\}$, and $Z_i(W) = \frac{1}{2^i} \sum_{v \in \mathcal{X}_i} Z_v(W)$, where $i = 0, 1, ..., r-1$, and $\mathcal{X}_i = \{v \in \mathcal{X} : i = \arg \max_{0 \leq j \leq r-1} v_j \neq 0\}$ with the binary representation of v, $(v_0, v_1, ..., v_{r-1})$.

$Z_i(W_N^{(j)})$ $\forall i \in \{0, 1, ..., r-1\}$ and $j \in \{0, 1, ..., N-1\}$ converges to $Z_{i,\infty} \in \{0, 1\}$, and with probability one $(Z_{0,\infty}, Z_{1,\infty}, ..., Z_{r-1,\infty})$ takes one of the values $(Z_{0,\infty} = 1, ..., Z_{k-1,\infty} = 1, Z_{k,\infty} = 0, ..., Z_{r-1,\infty} = 0)$ $\forall k = 0, 1, ..., r-1$, i.e., Theorem 1.b of [10]. $j \in \mathcal{A}_{k,n}$ iff $(Z_0(W_N^{(j)}), Z_1(W_N^{(j)}), ..., Z_{r-1}(W_N^{(j)})) \in \mathcal{R}_k(\epsilon)$, where $\mathcal{R}_k(\epsilon) \overset{def}{=} (\prod_{i=0}^{k-1} D_1) \times (\prod_{i=k}^{r-1} D_0)$, $D_0 = [0, \epsilon)$, $D_1 = (1 - \epsilon, 1]$. The channel polarizes in the sense that $\frac{\sum_{i \in \{0,1,...,N-1\}} |\mathcal{A}_{k_i,n}|}{rN} \to I(W)$ as $N \to \infty$.

For $u_0^{N-1} \in \mathcal{X}^N$, we also represent it by its binary form, that is $u_0^{N-1} = (u_{I(0,0)}, ..., u_{I(0,r-1)}, ..., u_{I(N-1,0)}, ..., u_{I(N-1,r-1)}) \in \{0,1\}^{rN}$, where $u_{I(i,0)}, ..., u_{I(i,r-1)}$ is the binary representation of $u_i \in \mathcal{X}$, and $I(i,j) = i \times r + j$. $\forall i \in \mathcal{A}_{k_i,n}$, let the frozen bit set be determined by $F = \{I(i,j) : i \in \{0, 1, ..., N-1\}, j \in \{0, 1, ..., k_i - 1\}\} \subseteq \{0, 1, ..., rN - 1\}$. Frozen bits for u_0^{N-1} are defined as $u_F \in \{0,1\}^{|F|}$ with the subvector $u_i : i \in F$.

Finally, the polar code, $C_N(F, u_F)$, $\forall F \subseteq \{0, 1, ..., rN - 1\}$ and $u_F \in \{0,1\}^{|F|}$, is a linear code given by $C_N(F, u_F) = \{x_0^{N-1} = u_0^{N-1} G_N : \forall u_{F^c} \in \{0,1\}^{|F^c|}\}$. The polar code ensemble, $C_N(F)$, $\forall F \subseteq \{0, 1, ..., rN - 1\}$, is $C_N(F) = \{C_N(F, u_F) : \forall u_F \in \{0,1\}^{|F|}\}$.

2) The code construction: We focus on the q-ary WEM code construction with a *symmetric rewriting cost function*, which satisfies $\forall x, y, z \in \{0, 1, ..., q-1\}, \varphi(x, y) = \varphi(x + z, y + z)$, where $+$ is over GF(q).

Similar to the binary case, to construct codes for WEM with $\mathcal{R}^s(q, D)$, we utilize its related form $R^s(q, D)$ in Lemma 5 and its test channel $W(y|x)$.

The code construction is presented in Algorithm IV.1.

Algorithm IV.1 A code construction for $(N, M, q, D)_{ave}$ WEM

1: $\forall i \in \mathcal{A}_{k_i,n}$, $F = \{I(i,j) : i \in \{0, 1, ..., N-1\}, j \in \{0, 1, ..., k_i - 1\}\}$.
2: The $(N, M, q, D)_{ave}$ code is $\mathcal{C} = \{\mathcal{C}_i : \mathcal{C}_i = C_N(F, u_F(i))\}$, where $u_F(i)$ is the binary representation form of i for $i \in \{0, 1, ..., M-1\}$.

The WEM code rate is $\mathcal{R} = \frac{|F|}{rN}$, and the polar code rate is $R = \frac{|F^c|}{rN}$. Similarly, when R approaches $R^s(q,D)$, $\mathcal{R}$ approaches $\mathcal{R}^s(q,D)$ based on Lemma 5.

The rewriting operation and the decoding operation are defined in Algorithm IV.2 and Algorithm IV.3, respectively, where the SC encoding is a generalization of q-ary lossy source coding, and the dither g_0^{N-1} is inspired by [3] to make sure the uniform distribution of v_0^{N-1}.

Algorithm IV.2 The rewriting operation $y_0^{N-1} = \mathbf{R}(x_0^{N-1}, i)$.

1: Let $v_0^{N-1} = x_0^{N-1} + g_0^{N-1}$, where g_0^{N-1} is known both to the decoding function and to the rewriting function, it is chosen such that v_0^{N-1} is uniformly distributed, and $+$ is over GF(q).

2: SC encoding v_0^{N-1}, $\hat{u}_0^{N-1} = \hat{U}(v_0^{N-1}, u_F(i))$, that is for each k in the range 0 till $N-1$:

$$\hat{u}_j = \begin{cases} u_j & \text{if } j \in \mathcal{A}_{r,n}, \\ m & \text{with the posterior } P(m|\hat{u}_0^{j-1}, v_0^{N-1}), \end{cases}$$

where in the above $m = 0, 1, ..., q-1$ and if $j \in \mathcal{A}_{k,n}$ for $0 \le k \le r-1$, the first k bits of $\hat{u}_j$ are fixed, and let $\hat{y}_0^{N-1} = \hat{u}_0^{N-1} G_N$.

3: $y_0^{N-1} = \hat{y}_0^{N-1} - g_0^{N-1}$, and $-$ is over GF(q).

Algorithm IV.3 The decoding operation $u_F(i) = \mathbf{D}(x_0^{N-1})$.

1: $y_0^{N-1} = x_0^{N-1} + g_0^{N-1}$.

2: $u_F(i) = (y_0^{N-1} G_N^{-1})_F$.

The correctness of the above rewriting function can be verified similarly to Lemma 6.

3) The average rewriting cost analysis: Similar to the analysis of equation (4), we obtain that $\bar{D} = \sum_{w_0^{N-1}} \pi(w_0^{N-1}) \sum_j \bar{D}_j(w_0^{N-1})$.

In the following, we first focus on $\bar{D}_j(w_0^{N-1})$. Note that $\hat{u}_0^{N-1} = \hat{U}(v_0^{N-1}, u_F(j))$ is random, i.e., the SC encoding function may result in different outputs for the same input. More precisely, in step i of the SC encoding process, $i \in \bigcup_{k=0}^{r-1} \mathcal{A}_{k,n}$, $\hat{u}_i = m$ with the posterior $P(m|\hat{u}_0^{i-1}, v_0^{N-1})$, where if $i \in \mathcal{A}_{k,n}$, the first k bits of $\hat{u}_i$ are fixed and known. This implies that the probability of picking a vector u_0^{N-1} with $\hat{\mathcal{A}}_{r,n}$ given v_0^{N-1} with $\mathcal{A}_{r,n}$ is equal to

$$\begin{cases} 0 & \text{if } \hat{\mathcal{A}}_{r,n} \neq \mathcal{A}_{r,n}, \\ \prod_{i \in \bigcup_{k=0}^{r-1} \mathcal{A}_{k,n}} P(u_i|u_0^{i-1}, v_0^{N-1}) & \text{otherwise,} \end{cases}$$

where in the second case $\forall i \in \mathcal{A}_{k,n}$, the first k bits of u_i are fixed.

Therefore, the average (in this case, over the probability of rewriting to data j and the randomness of the encoder) rewriting cost of updating w_0^{N-1} to a codeword representing j, $\bar{D}_j(w_0^{N-1})$, is

$$\begin{aligned} = \; & \frac{1}{2^{|F|}} \sum_{u_{F^c}} \prod_{i \in \bigcup_{k=0}^{r-1} \mathcal{A}_{k,n}} P(u_i|u_0^{i-1}, w_0^{N-1}) \\ & \varphi(w_0^{N-1}, u_0^{N-1} G_N), \end{aligned}$$

where $u_F = u_F(j)$, $u_{F^c} \in \{0,1\}^{|F^c|}$, and the summation over u_{F^c} takes care of $\mathcal{A}_{r,n}$ and the fact that $i \in \mathcal{A}_{k,n}$, the first k bits of u_i are fixed.

Thus, we obtain that $\bar{D}$

$$\begin{aligned} = \; & \sum_{w_0^{N-1}} \pi(w_0^{N-1}) \sum_j \bar{D}_j(w_0^{N-1}), \\ = \; & \sum_{w_0^{N-1}} \pi(w_0^{N-1}) \sum_{u_F(j)} \frac{1}{2^{|F|}} \sum_{u_F^c} \prod_{i \in \bigcup_{k=0}^{r-1} \mathcal{A}_{k,n}} \\ & P(u_i|u_0^{i-1}, w_0^{N-1}) \varphi(w_0^{N-1}, u_0^{N-1} G_N), \\ < \; & \sum_{w_0^{N-1}} \pi(w_0^{N-1}) \frac{1}{q^{|\mathcal{A}_{r,n}|}} \sum_{u_0^{N-1}} \prod_{i \in \bigcup_{k=0}^{r-1} \mathcal{A}_{k,n}} \\ & P(u_i|u_0^{i-1}, w_0^{N-1}) \varphi(w_0^{N-1}, u_0^{N-1} G_N). \end{aligned} \quad (5)$$

Let $Q_{U_0^{N-1}, W_0^{N-1}}$ denote the distribution defined by $Q_{W_0^{N-1}}(w_0^{N-1}) = \pi(w_0^{N-1})$, and $Q_{U_0^{N-1}|W_0^{N-1}}$ defined by

$$Q(u_i|u_0^{i-1}, w_0^{N-1}) = \begin{cases} \frac{1}{q} & \text{if } i \in \mathcal{A}_{r,n}, \\ P(u_i|u_0^{i-1}, w_0^{N-1}) & \text{otherwise.} \end{cases}$$

Then, inequation (5) is equivalent to $\bar{D} < E_Q(\varphi(w_0^{N-1}, u_0^{N-1} G_N))$, where $E_Q(\cdot)$ denotes the expectation with respect to the distribution $Q_{U_0^{N-1}, W_0^{N-1}}$. Similarly, let $E_P(\cdot)$ denote the expectation with respect to the distribution $P_{U_0^{N-1}, W_0^{N-1}}$.

The following three lemmas are already proved in [7] for $q = 2$ and in [6] for primary q, they extend trivially to $q = 2^r$, and we omit their proofs.

Lemma 10. $\sum_{w_0^{N-1}, u_0^{N-1}} |Q(w_0^{N-1}, u_0^{N-1}) - P(w_0^{N-1}, u_0^{N-1})| \le \sum_{i \in \mathcal{A}_{r,n}} \sum_{u_i=0}^{q-1} E_P(|\frac{1}{q} - P(u_i|u_0^{i-1}, w_0^{N-1})|)$.

Lemma 11. Let F be chosen such that for $i \in \mathcal{A}_{r,n}$,

$$\sum_{u_i=0}^{q-1} E_P(|\frac{1}{q} - P(u_i|u_0^{i-1}, w_0^{N-1})|) \leq \sigma_N.$$

Then, the average rewriting cost is bounded by

$$\frac{1}{N} E_Q(\varphi(w_0^{N-1}, u_0^{N-1} G_N)) \leq$$

$$\frac{1}{N} E_P(\varphi(w_0^{N-1}, u_0^{N-1} G_N)) + |\mathcal{A}_{r,n}| d_{max} \sigma_N,$$

where $d_{max} \overset{def}{=} \max_{x,y} \varphi(x, y)$.

Lemma 12. $E_P(\varphi(w_0^{N-1}, u_0^{N-1} G_N)) = ND$.

The following lemma, which is a modification of Lemma 5 [6], presents that it is sufficient to choose the set F as those bits with indexes $I(i, j)$ for which $i \in \{0, 1, ..., N-1\}, j \in \{0, 1, ..., k_i - 1\}$ with $i \in \mathcal{A}_{k_i,n}$.

Lemma 13. If $i \in \mathcal{A}_{r,n}$, that is $(Z_0(W_N^{(i)}), ..., Z_{r-1}(W_N^{(i)})) \in \mathcal{R}_r(\epsilon)$, and let $\epsilon_N \geq r\sqrt{\epsilon}$, then

$$\sum_{u_i=0}^{q-1} E_P(|\frac{1}{q} - P(u_i|u_0^{i-1}, w_0^{N-1})|) \leq \sqrt{2\epsilon_N}.$$

Proof: By Pinsker's inequality, for two distribution functions P and Q defined on $\mathcal{X}$, $||P - Q||_1 \leq \sqrt{2D(P||Q)}$, where $D(P||Q)$ is the Kullback-Leibler divergence between two distributions, that is $D(P||Q) = \sum_i \log_2(\frac{P(i)}{Q(i)}) P(i)$, we obtain that:

$$\begin{aligned}
&\sum_{u_i=0}^{q-1} E_P|\tfrac{1}{q} - P(u_i|u_0^{i-1}, w_0^{N-1})| \\
&\overset{P(u_i)=\frac{1}{q}}{=} \sum_{w_0^{N-1}, u_0^i} |P(u_i)P(u_0^{i-1}, w_0^{N-1}) \\
&\qquad -P(u_i|u_0^{i-1}, w_0^{N-1})P(u_0^{i-1}, w_0^{N-1})|, \\
&= \sum_{w_0^{N-1}, u_0^i} |P(u_i)P(u_0^{i-1}, w_0^{N-1}) \\
&\qquad -P(u_0^i, w_0^{N-1})|, \\
&\leq \sqrt{2 \sum_{w_0^{N-1}, u_0^i} \log_2 \frac{P(u_0^i, w_0^{N-1})}{P(u_i)P(u_0^{i-1}, w_0^{N-1})}} \\
&\qquad \sqrt{P(u_0^i, w_0^{N-1})}, \\
&= \sqrt{2I(W_N^{(i)})}, \\
&\leq \sqrt{2\epsilon_N},
\end{aligned}$$

where the last inequality is based on the conclusion of Lemma 1 [10], that is if $(Z_0(W_N^{(i)}), ..., Z_{r-1}(W_N^{(i)})) \in \mathcal{R}_k(\epsilon)$ $\forall k = 0, 1, ..., r$, then $|I(W_N^{(i)}) - (r - k)| \leq \gamma$ with $\gamma \geq \max(k\sqrt{\epsilon}, (2^{r-k} - 1)\epsilon \log_2 e)$. ∎

Therefore, the performance of the above polar WEM code can be concluded as follows:

Theorem 14. For a q-ary ($q = 2^r$) symmetric rewriting cost function $\varphi : \mathcal{X} \times \mathcal{X} \to \mathcal{R}_+$. Fix a rewriting cost D and $0 < \beta < \frac{1}{2}$. For any rate $\mathcal{R} < \mathcal{R}^s(q, D)$, there exists a sequence of polar WEM codes of length N and rate $R \leq \mathcal{R}$, so that under the above rewriting operation, $\bar{D}$ satisfies $\bar{D} \leq D + O(2^{-N^\beta})$. The decoding and rewriting operation complexity of the codes is $O(N \log N)$.

B. A code construction for q-ary WEM with a maximal rewriting cost constraint, $q = 2^r$

Similarly, the code construction, the rewriting operation and the decoding operation are exactly the same as Algorithm IV.1, Algorithm IV.2, and Algorithm IV.3, respectively. Next, we mainly focus on its performance.

Theorem 15. For a q-ary ($q = 2^r$) symmetric rewriting cost function $\varphi : \mathcal{X} \times \mathcal{X} \to \mathcal{R}_+$. Fix a rewriting cost D, δ, and $0 < \beta < \frac{1}{2}$. For any rate $\mathcal{R} < \mathcal{R}^s(q, D)$, there exists a sequence of polar WEM codes of length N and rate $R \leq \mathcal{R}$, so that under the above rewriting operation and the induced probability distribution Q, the rewriting cost between a current codeword $\forall y_0^{N-1}$ and its updated codeword x_0^{N-1} satisfies $Q(\varphi(y_0^{N-1}, x_0^{N-1}) \geq D + \delta) < 2^{-N^\beta}$. The decoding and rewriting operation complexity of the codes is $O(N \log N)$.

Proof: We mainly focus on the rewriting cost analysis. The proof of this part is based on the *ϵ-strong typical sequence* [4] and Theorem 4 and 5 of [3]. We give the sketch as follows.

We recall ϵ-strong typical sequences $x_0^{N-1} \times y_0^{N-1} \in \mathcal{X}^N \times \mathcal{Y}^N$ with respect to $p(x, y)$ over $\mathcal{X} \times \mathcal{Y}$, and denote it by $A_\epsilon^{*(N)}(X, Y)$. We denote by $C(a, b|x_0^{N-1}, y_0^{N-1})$ the number of occurrences of a, b in x_0^{N-1}, y_0^{N-1} with the same indexes, and require the following. First, $\forall a, b \in \mathcal{X} \times \mathcal{Y}$ with $p(a, b) > 0$, $|C(a, b|x_0^{N-1}, y_0^{N-1})/N - p(a, b)| < \epsilon$. Second, $\forall a, b \in \mathcal{X} \times \mathcal{Y}$ with $p(a, b) = 0$, $C(a, b|x_0^{N-1}, y_0^{N-1}) = 0$.

In our case $y_0^{N-1} = u_0^{N-1} G_N$. Due to the full rank of G_N, there is a one-to-one correspondence between u_0^{N-1} and y_0^{N-1}. We say that $u_0^{N-1}, x_0^{N-1} \in A_\epsilon^{*(N)}(U, X)$ if $x_0^{N-1}, u_0^{N-1} G_N \in A_\epsilon^{*(N)}(X, Y)$ with respect to $\frac{1}{q} W(y|x)$, where $W(y|x)$ is the test channel.

The first conclusion is that for N sufficiently large, $Q(A_\epsilon^{*(N)}(U, X)) > 1 - 2^{-N^\beta}$ for $\forall 0 < \beta < \frac{1}{2}$, $\epsilon > 0$, which is a generalization of Theorem 4 of [3], and where $Q(A_\epsilon^{*(N)}(U, X)) = Q(\forall a, b :$

$|\frac{1}{N}C(a,b|u_0^{N-1}G_N, x_0^{N-1}) - \frac{1}{q}W(a|b)| \leq \epsilon)$. The outline is that based on Lemma 10 and Lemma 11 we obtain that

$$\sum_{u_0^{N-1}, x_0^{N-1} \in A_\epsilon^{*(N)}(U,X)} |Q(u_0^{N-1}, x_0^{N-1}) - P(u_0^{N-1}, x_0^{N-1})| \leq |\mathcal{A}_{r,n}|\sigma_N d_{max}.$$

Thus, we obtain

$$\begin{aligned} & |\sum_{u_0^{N-1}, x_0^{N-1}} Q(u_0^{N-1}, x_0^{N-1}) - P(u_0^{N-1}, x_0^{N-1})| \\ \leq & \sum_{u_0^{N-1}, x_0^{N-1}} |Q(u_0^{N-1}, x_0^{N-1}) - P(u_0^{N-1}, x_0^{N-1})|, \\ \leq & |\mathcal{A}_{r,n}|\sigma_N d_{max}, \end{aligned}$$

where in the above $u_0^{N-1}, x_0^{N-1} \in A_\epsilon^{*(N)}(U,X)$.

By lower bounding $P(A_\epsilon^{*(N)}(U,X)) = 1 - P(\exists a, b : |\frac{1}{N}C(a,b|u_0^{N-1}G_N, x_0^{N-1}) - \frac{1}{q}W(a|b)| \geq \epsilon) \geq 1 - 2q^2 e^{-2N\epsilon^2}$ based on Hoeffding's inequality, and $Q(A_\epsilon^{*(N)}(U,X)) \geq P(A_\epsilon^{*(N)}(U,X)) - |\mathcal{A}_{r,n}|\sigma_N d_{max}$, we obtain the desired result by setting $\sigma_N = \frac{2^{-N^\beta}}{2N d_{max}}$.

The second conclusion is that let $y_0^{N-1} = \mathbf{R}(x_0^{N-1}, i)$, then $\forall \delta > 0$, $0 < \beta < \frac{1}{2}$ and N sufficiently large, $Q(\varphi(y_0^{N-1}, x_0^{N-1})/N \geq D + \delta) < 2^{-N\beta}$, which is a generalization of Theorem 5 of [3]. The outline is that

$$\begin{aligned} & Q(\varphi(x_0^{N-1}, y_0^{N-1})/N \geq D + \delta) \\ \leq & Q(\varphi(x_0^{N-1}, y_0^{N-1})/N \geq D + \delta \\ & \bigcap x_0^{N-1}, y_0^{N-1} \in A_\epsilon^{*(N)}(X,Y)) \\ + & Q(x_0^{N-1}, y_0^{N-1} \notin A_\epsilon^{*(N)}(X,Y)), \\ \leq & 2^{-N\beta}, \end{aligned}$$

where the last inequality is based on the conclusion just obtained $Q(x_0^{N-1}, y_0^{N-1} \notin A_\epsilon^{*(N)}(X,Y)) < 2^{-N\beta}$, and that when $x_0^{N-1}, y_0^{N-1} \in A_\epsilon^{*(N)}(X,Y)$, for ϵ sufficiently small and N sufficiently large, $\varphi(y_0^{N-1}, x_0^{N-1})/N \leq D + \delta$. ■

V. Conclusion

Code constructions for WEM using recently proposed polar codes have been presented. Future work focuses on exploring error-correcting codes for WEM.

VI. Acknowledgement

This work was supported in part by the National Science Foundation (NSF) CAREER Award CCF-0747415 and the NSF Grant CCF-1217944.

References

[1] R. Ahlswede and Z. Zhang, "Coding for write-efficient memory," *Inform. Comput*, vol. 83, no. 1, pp. 80–97, Oct 1989.
[2] E. Arikan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans on Inf Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.
[3] D. Burshtein and A. Strugatski, "Polar write once memory codes," in *Proceedings of the 2012 IEEE international conference on Symposium on Information Theory (ISIT'12)*, Boston, MA, July 2012, pp. 1972–1975.
[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[5] F. Fu and R. W. Yeung, "On the capacity and error-correcting codes of write-efficient memories," *IEEE Trans on Inf Theory*, vol. 46, no. 7, pp. 2299–2314, Nov 2000.
[6] M. Karzand and E. Teletar, "Polar codes for q-ary source coding," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Austin, Texas, June 2010, pp. 909–912.
[7] S. B. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1751–1768, April 2010.
[8] L. A. Lastras-Montano, M. Franceschini, T. Mittelholzer, J. Karidis, and M. Wegman, "On the lifetime of multilevel memories," in *Proceedings of the 2009 IEEE international conference on Symposium on Information Theory (ISIT'09)*, Coex, Seoul, Korea, 2009, pp. 1224–1228.
[9] Q. Li, "Compressed rank modulation," in *Proc. 50th Annual Allerton Conference on Communication, Control and Computing (Allerton)*, Monticello, IL, October 2012.
[10] W. Park and A. Barg, "Polar codes for q-ary channels, $q = 2^r$," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Boston, MA, June 2012, pp. 2142–2146.
[11] R. L. Rivest and A. Shamir, "How to reuse a write-once memory," *Inf. Contr*, vol. 55, pp. 1–19, 1982.