Coding for Noisy Write-Efficient Memories

Qing Li
Computer Sci. & Eng. Dept.
Texas A & M University
College Station, TX 77843
qingli@cse.tamu.edu

Anxiao (Andrew) Jiang
CSE and ECE Departments
Texas A & M University
College Station, TX 77843
ajiang@cse.tamu.edu

***Abstract*—For nonvolatile memories such as flash memories and phase-change memories, endurance and reliability are both important challenges. Write-Efficient Memory (WEM) is an important rewriting model to solve the endurance problem. An optimal rewriting code has been proposed to approach the rewriting capacity of WEM.**

Aiming at jointly solving the endurance and the data reliability problem, this work focuses on a combined error correction and rewriting code for WEM. To that end, a new coding model, noisy WEM, is proposed here. Its noisy rewriting capacity is explored. An efficient coding scheme is constructed for a special case of noisy WEM. Its decoding and rewriting operations can be done in time $O(N \log N)$, with N as the length of the codeword, and it provides a lower bound to the noisy WEM's capacity.

I. Introduction

Nonvolatile memories (such as flash memories and phase-change memories (PCM)) are becoming ubiquitous nowadays. Besides the well-known endurance [10] problem, another serious challenge is the data reliability issue, e.g., retention error [16] in NAND flash memories and resistance drift [9] in PCM.

Write-efficient memory (WEM) [1] is a coding model that can be used to solve the endurance problem for nonvolatile memories. In WEM, codewords are partitioned into several disjointed sets, where codewords of the same set represent the same data. A cost constraint has to be satisfied during the rewriting (namely, updating the data stored in WEM).

WEM is a natural model for PCM [13], and can also be applied to flash memory when data representation scheme such as rank modulation [14] is used. In WEM, there is a cost associated with changing the level of a cell. For nonvolatile memories such as PCM, this cost is important because cells age with programming and have the endurance problem. An optimal code [15] has been proposed to achieve the rewriting capacity of WEM. However, rewriting codes combined with error correction are still limited [11], especially for WEM [7].

In this paper, we propose a joint error correction and rewriting scheme for WEM. While previous results are mainly for Write-Once Memories [3], our work focuses on WEM. We propose a new coding model, noisy WEM, and provide a characterization for its capacity. We present an efficient coding scheme based on polar codes [2] for a special case of noisy WEM. The scheme is related to the coding schemes in [12], [3] and [11]. We also provide a lower bound to noisy WEM's capacity, and experimentally verify the code construction's performance.

The rest of this paper is organized as follows. In Section II, we introduce noisy write-efficient memory model. In Section III, we present characterization of noisy rewriting capacity of noisy WEM. In Section IV, we present an efficient code construction for a special case of binary noisy WEM, and verify its performance experimentally. We conclude this paper in Section V.

II. Noisy Write-efficient memory model

In this section, we first introduce terms and notations used throughout the paper, and then formally present the definitions of noisy WEM and related parameters.

A. Terms and notations

Let $\mathcal{X} = \{0, 1, \cdots, q-1\}$ be the alphabet of a symbol stored in a cell. (For example, for PCM, it denotes the q levels of a cell.) $\forall x, y \in \mathcal{X}$, let the rewriting cost of changing a cell's level from x to y be $\varphi(x, y)$. Given N cells and $x_0^{N-1}, y_0^{N-1} \in \mathcal{X}^N$, let $\varphi(x_0^{N-1}, y_0^{N-1}) = \frac{1}{N}\sum_{i=0}^{N-1} \varphi(x_i, y_i)$ be the rewriting cost of changing the N cell levels from x_0^{N-1} to y_0^{N-1}.

Let $M \in \mathbb{N}$ and $\mathcal{D} = \{0, 1, \cdots, M-1\}$. We use $\mathcal{D}$ to denote the M possible values of the data stored in the N cells. Let the decoding function be $\mathbf{D} : \mathcal{X}^N \to \mathcal{D}$, which maps the N cells' levels to the data they represent. Let the rewriting function be $\mathbf{R} : \mathcal{X}^N \times \mathcal{D} \to \mathcal{X}^N$, which changes the N cells' levels to represent the new input data. Naturally, we require $\mathbf{D}(\mathbf{R}(x_0^{N-1}, i)) = i$ for any $x_0^{N-1} \in \mathcal{X}^N$ and $i \in \mathcal{D}$.

Assume the sequence of data written to the storage medium is $\{M_1, \cdots, M_t\}$, where we assume M_i for $1 \leq i \leq t$ is uniformly distributed over $\mathcal{D}$, and the average rewriting cost is $\bar{D} \stackrel{def}{=} \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} \varphi(x_0^{N-1}(i), \mathbf{D}(M_i, x_0^{N-1}(i)))$, where $x_0^{N-1}(i)$ is the current cell states before the i^{th} update. By assuming the stationary distribution of cell levels $x_0^{N-1}(i)$ is $\pi(x_0^{N-1})$, $\bar{D} = \sum_{x_0^{N-1}} \pi(x_0^{N-1}) \sum_{j\in\mathcal{D}} \bar{D}_j(x_0^{N-1})$, where $\bar{D}_j(x_0^{N-1})$ is the average rewriting cost of updating cell level states x_0^{N-1} to a codeword representing $j \in \mathcal{D}$.

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 $(S, X) \in (\mathcal{X}, \mathcal{X})$, let $P_{SX}, P_S, P_{X|S}$ denote the joint probability distribution, the marginal distribution, and the conditional probability distribution, respectively. $E(\cdot)$ denotes the expectation operator. If X is uniformly distributed over $\{0, 1, \cdots, q-1\}$, denote it by $X \sim U(q)$.

B. Noisy WEM with an average rewriting cost constraint

We formally present the definition of noisy WEM with an average rewriting cost constraint as follows:

Definition 1. An $(N, M, q, d)_{ave}$ noisy WEM code for the *storage channel* $\mathcal{P} = (\mathcal{X}, \mathcal{Y}, P_{Y|X}(y|x))$ consists of $\mathcal{D}$ and $\mathcal{C} = \bigcup_{i\in\mathcal{D}} \mathcal{C}_i$, where $\mathcal{C}_i \subseteq \mathcal{X}^N$ is the set of codewords representing data i. We require $\forall i \neq j$, $\mathcal{C}_i \bigcap \mathcal{C}_j = \emptyset$. (Here $P_{Y|X}(y|x)$ represents the transition probability of the noisy channel, which changes a cell's level from x to y.) It also consists of a rewriting function $\mathbf{R}(s_0^{N-1}, i)$ and a decoding function $\mathbf{D}(y_0^{N-1})$. Let $d \in \mathbb{R}^+$ be an upper bound to the average rewriting cost; namely, we require $\bar{D} \leq d$.

The noisy WEM model is illustrated in Fig. 1. Here the N-dimensional vector $s_0^{N-1} \in \mathcal{X}^N$ is the current cell states, and the message M is the new information to write, which is independent of s_0^{N-1}. The rewriter uses both s_0^{N-1} and M to choose a new codeword $x_0^{N-1} \in \mathcal{X}^N$, which will be programmed as the N cells' new states, such that the average rewriting cost satisfies the predefined cost constraint. The codeword x_0^{N-1} passes a noisy channel, and the noisy codeword $y_0^{N-1} \in \mathcal{X}^N$ is its output. The decoder can reliably decode y_0^{N-1} to recover the message M. The cell states s_0^{N-1} are drawn independently and identically from the probability distribution $P_S(s)$. The noisy channel is memoryless, and is characterized by the transition probabilities $P_{Y|X}(y|x)$.

Let $\lambda_i = Pr(\mathbf{D}(y_0^{N-1}) \neq i | x_0^{N-1} = \mathbf{R}(s_0^{N-1}, i))$ be the decoding error probability given data i. Let $\lambda^{(N)}$ be

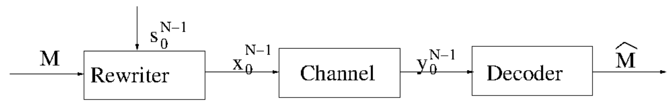

Fig. 1. The noisy WEM model. M, s_0^{N-1}, x_0^{N-1}, y_0^{N-1} and $\hat{M}$ are , respectively, the message, the current cell states, rewritten codeword, the noisy channel's output, and the estimated message.

$\max_{i\in\mathcal{D}} \lambda_i$. Let $\mathcal{R} = \frac{\log M}{N}$ be the code rate, and we say $\mathcal{R}$ is achievable if there exists a $(N, 2^{N\mathcal{R}}, q, d)_{ave}$ code such that $\lambda^{(N)} \to 0$ as $N \to \infty$. The *noisy rewriting capacity* $\mathcal{C}(q, d)_{ave}$ is the supremum of all achievable rates.

The noisy WEM problem is: given the average rewriting cost constraint d, find the maximal rate $\mathcal{R}$ of reliable rewriting supported by the rewriter and the decoder despite the noisy channel. Let $\mathcal{P}(q, d)$ be $\{P_{SX} \in \mathcal{P}(\mathcal{X} \times \mathcal{X}) : P_S = P_X, E(\varphi(S, X)) \leq d\}$. When there is no noise, $\mathcal{C}(q, d)_{ave}$ is $\mathcal{R}(q, d)_{ave} = \max_{P_{SX}\in\mathcal{P}(q,d)} H(X|S)$ [1].

C. Noisy WEM with a maximal rewriting cost constraint

The WEM code in definition 1 puts a constraint on the average rewriting cost. We now define a code with a maximal rewriting cost constraint.

Definition 2. An (N, M, q, d) noisy WEM code for the *storage channel* $\mathcal{P} = (\mathcal{X}, \mathcal{Y}, P_{Y|X}(y|x))$ consists of $\mathcal{D}$ and $\mathcal{C} = \bigcup_{i\in\mathcal{D}} \mathcal{C}_i$, where $\mathcal{C}_i \subseteq \mathcal{X}^N$ is the set of codewords representing data i. We require $\forall i \neq j$, $\mathcal{C}_i \bigcap \mathcal{C}_j = \emptyset$. It also consists of a rewriting function $\mathbf{R}(s_0^{N-1}, i)$ and a decoding function $\mathbf{D}(y_0^{N-1})$. Let $d \in \mathbb{R}^+$ be an upper bound to the maximal rewriting cost; namely, we require $\varphi(s_0^{N-1}, \mathbf{R}(s_0^{N-1}, i)) \leq d$ for any $s_0^{N-1} \in \mathcal{C}$ and $i \in \mathcal{D}$.

The rate $\mathcal{R}$ and noisy rewriting capacity $\mathcal{C}(q, d)$ can be defined similarly as before. When there is no noise, $\mathcal{C}(q, d)$ is $\mathcal{R}(q, d) = \mathcal{R}(q, d)_{ave}$ [1].

III. Characterizing Noisy rewriting capacity

In this section, we present the characterization of noisy rewriting capacity of $\mathcal{C}(q, d)$ and $\mathcal{C}(q, d)_{ave}$, respectively.

The characterization of $\mathcal{C}(q, d)$ is presented below. It is effectively the generalization of that of Gel'fand and Pinsker [8], which considers the problem without cost constraint. The direct part proof is based on random coding and typical sequences; the converse part is based on techniques of Fano's inequality [4] and Csiszár sum identity [5], and auxiliary random variables identification.

Lemma 3. For a given rewriting cost function $\varphi(\cdot,\cdot)$, $\mathcal{C}(q,d) = \max_{\substack{P_{U|S},P_{X|U,S}\\P_{SX}\in\mathcal{P}(q,d)}} \{I(Y;U) - I(U;S)\}$, where U is an auxiliary random variable, and $U \to (X,S) \to Y$ is a Markov chain.

Proof:

1) Background of strong typical-sequence: We first present some background about strong typical-sequences. For more details, interested readers are referred to [6].

Let x_0^{N-1} be a sequence with N elements drawn from $\mathcal{X}$. Define the type of x_0^{N-1} by $\pi(x|x_0^{N-1}) = \frac{|\{i:x_i=x\}|}{N}$. The set $\mathcal{T}_\epsilon^N(X)$ is defined as:

$$\mathcal{T}_\epsilon^N(X) = \{x_0^{N-1} : |\pi(x|x_0^{N-1}) - P_X(x)| \le \epsilon, \forall x\}.$$

That is, the set of sequences for which the empirical frequency is within ϵ of the probability $P_X(x)$ for every $x \in \mathcal{X}$.

Let (x_0^{N-1}, y_0^{N-1}) be a pair of sequences with elements drawn from $(\mathcal{X},\mathcal{Y})$. Define their joint type: $\pi(x,y|x_0^{N-1},y_0^{N-1}) = \frac{|\{i:(x_i,y_i)=(x,y)\}|}{N}$ for $(x,y) \in \mathcal{X}\times\mathcal{Y}$. We denote $\mathcal{T}_\epsilon^N(XY) = \{(x_0^{N-1},y_0^{N-1}) : |\pi(x,y|x_0^{N-1},y_0^{N-1}) - P_{XY}(x,y)| \le \epsilon, \forall (x,y)\}$.

For $x_0^{N-1} \in \mathcal{T}_\epsilon^N(X)$ and $P_{Y|X}$, we define the conditional typical sequence $\mathcal{T}_{Y|X}^N(x_0^{N-1}) = \{y_0^{N-1} : (x_0^{N-1},y_0^{N-1}) \in \mathcal{T}_\epsilon^N(XY)\}$.

The following results will be used:

i For a vector x_0^{N-1}, where x_i is chosen i.i.d. $\sim P_X$,
$$Pr(x_0^{N-1} \in \mathcal{T}_\epsilon^N(X)) \to 1 \text{ as } N \to \infty. \quad (1)$$

ii For vectors x_0^{N-1}, y_0^{N-1}, where (x_i,y_i) is chosen i.i.d. $\sim P_{XY}$,
$$Pr((x_0^{N-1},y_0^{N-1}) \in \mathcal{T}_\epsilon^N(XY)) \to 1 \text{ as } N \to \infty. \quad (2)$$

iii For $x_0^{N-1} \in \mathcal{T}_\epsilon^N(X)$, and y_0^{N-1} is independently chosen according to P_Y, then $Pr((x_0^{N-1},y_0^{N-1}) \in \mathcal{T}_{Y|X}^N(x_0^{N-1})) \in$
$$[2^{-N(I(X;Y)+\lambda)}, 2^{-N(I(X;Y)-\lambda)}], \quad (3)$$
for some $\lambda(\epsilon) > 0$ with $\lambda \to 0$ as $\epsilon \to 0$.

2) Proof of direct part:

a) The code construction: We present details of the direct part by a random code construction. For a $P_{U|S}(u|s)$, let $P_U(u)$ be the marginal probability distribution of U under the joint probability distribution of $P_S(s)P_{U|S}(u|s)$. Independently generate a set of 2^{NQ} vectors of u_0^{N-1} from the typical sequence $\mathcal{T}_\epsilon^N(U)$, where Q is to be determined. Randomly partition such 2^{NQ} vectors into 2^{NR} subsets each with size $2^{N(Q-R)}$:

$$\{\mathcal{A}_0, \mathcal{A}_1 ..., \mathcal{A}_{2^{NR}-1}\}.$$

b) Rewriting function and its analysis: For the rewriting part, given a current state vector $s_0^{N-1} \in \bigcup_{i=0}^{2^{NR}-1} \mathcal{A}_i$ and data to rewrite $j \in \mathcal{D}$, randomly choose a vector u_0^{N-1} from $\mathcal{A}_j$ that is jointly typical with s_0^{N-1} with $P_{U,S}$. Rewrite a vector x_0^{N-1} conditionally typical with s_0^{N-1} and u_0^{N-1}, $x_0^{N-1} \in \mathcal{T}_{X|U,S}^N(u_0^{N-1}, s_0^{N-1})$. Due to the definition of strong typical sequence, we know that $\varphi(s_0^{N-1}, x_0^{N-1} = \mathbf{R}(s_0^{N-1}, i)) \le ND + \epsilon$. Therefore, the successful rewriting depends on whether such u_0^{N-1} exist or not.

Next, we analyze the condition under which such u_0^{N-1} exists almost for sure. The probability that we do not have a vector from $\mathcal{A}_j$ that is jointly typical with s_0^{N-1} is $P(\mathcal{A}_j \bigcap \mathcal{T}_{U|S}^N(s_0^{N-1}) = \emptyset)$

$$\begin{aligned}
&\le P(\mathcal{A}_j \bigcap \mathcal{T}_{U|S}^N(s_0^{N-1}) = \emptyset | s_0^{N-1} \notin \mathcal{T}_\epsilon^N(S)) \\
&+ P(\mathcal{A}_j \bigcap \mathcal{T}_{U|S}^N(s_0^{N-1}) = \emptyset | s_0^{N-1} \in \mathcal{T}_\epsilon^N(S)), \quad (4)\\
&\le (1 - P(u_0^{N-1} \in \mathcal{A}_j \bigcap \mathcal{T}_{U|S}^N(s_0^{N-1}) \\
&| \quad s_0^{N-1} \in \mathcal{T}_\epsilon^N(S)))^{|\mathcal{A}_j|}, \quad (5)\\
&\le (1 - 2^{-N(I(S;U)+\lambda)})^{2^{N(Q-R)}}, \quad (6)\\
&\le \exp\{2^{-(Q-R-I(S;U)+\lambda)}\}, \quad (7)
\end{aligned}$$

where inequation (5) is based on the conclusions (1) and (3).

To ensure the success of rewriting, there have to be vectors in $\mathcal{A}_j$, which are jointly typical with s_0^{N-1}, and this implies that $Q - R > I(U;S)$.

c) Decoding function and its analysis: For the decoding part, given a received vector y_0^{N-1}, decode it is $j \in \mathcal{D}$ if there exists some $u_0^{N-1} \in \mathcal{A}_j$ jointly typical with y_0^{N-1} for a unique j.

Let us consider the probability of decoding error for rewriting data j, $P_e(j)$. There are two cases: one is there is no vector in $\mathcal{A}_j$ jointly typical with y_0^{N-1}, and the other is there are vectors in $\mathcal{A}_i$ for $i \ne j$ jointly typical with y_0^{N-1}. Based on union bound, $P_e(j)$

$$\begin{aligned}
&\le P((u_0^{N-1}, y_0^{N-1}) \in \mathcal{T}_\epsilon^N(UY) | u_0^{N-1} \in \mathcal{A}_i, i \ne j) \\
&+ P((u_0^{N-1}, y_0^{N-1}) \notin \mathcal{T}_\epsilon^N(UY) | u_0^{N-1} \in \mathcal{A}_j), \quad (8)\\
&\le \sum_{\substack{u_0^{N-1}\\ i\ne j}} P((u_0^{N-1}, y_0^{N-1}) \in \mathcal{T}_\epsilon^N(UY) | u_0^{N-1} \in \mathcal{A}_i), \quad (9)\\
&\le \sum_{\substack{u_0^{N-1}\in\mathcal{A}_i\\ i\ne j}} 2^{-N(I(U;Y)-\lambda)}, \quad (10)\\
&\le 2^{-N(I(U;Y)-Q-\lambda)}, \quad (11)
\end{aligned}$$

where inequation (9) is based on the conclusion (2), and inequation (10) is based on (3).

Therefore, the decoding error probability $P_e(j)$ and further λ^N approach 0 if $Q < I(U;Y)$.

3) Proof of the converse part: The converse part is as follows: $N\mathcal{R}$

$$
\begin{aligned}
&= H(M), && (12)\\
&= H(M|y_0^N) + I(M;y_0^N), && (13)\\
&\leq 1 + P_e NR + I(M;y_0^{N-1}), && (14)\\
&\leq 1 + P_e NR + \sum_{i=0}^{N-1} I(M;y_i|y_0^{i-1}), && (15)\\
&\leq \sum_{i=0}^{N-1} [I(y_i;s_{i+1}^{N-1}|M,y_0^{i-1}) - I(s_i;y_0^{i-1}|M,s_{i+1}^{N})]\\
&+ 1 + P_e NR + \sum_{i=0}^{N-1} I(M;y_i|y_0^{i-1}), && (16)\\
&= \sum_{i=0}^{N-1} [I(M,s_{i+1}^{N};y_i|y_0^{i-1}) - I(s_i;y_0^{i-1}|M,s_{i+1}^{N})]\\
&+ 1 + P_e NR, && (17)
\end{aligned}
$$

where

- (12) follows from the assumption that M is uniformly distributed among $\mathcal{D}$;
- (13) follows from the definition of mutual information;
- (14) follows from Fano's inequality [4], and $P_e = \frac{1}{2^{NR}} \sum_i \lambda_i$;
- (15) follows from chain rule for mutual information [4], and so does equation (17);
- (16) follows from Csiszár sum identity [5], which is

$$\sum_{i=0}^{N-1} I(A_i;B_{i+1}^{N-1}|A_0^{i-1}) = \sum_{i=0}^{N-1} I(B_i;A_0^{i-1}|B_{i+1}^{N-1}), \tag{18}$$

where in inequation (14) $A_i = y_i$, $B_i = s_i$, and by conditioning on M.

Define $u_i = (M, s_{i+1}^{N-1}, y_0^{i-1})$ and we know that $u_i \to (x_i, s_i) \to y_i$, we continue equation (15): $I(M,s_{i+1}^{N-1};y_i|y_0^{i-1}) - I(s_i;y_0^{i-1}|M,s_{i+1}^{N-1})$

$$
\begin{aligned}
&= H(y_i|y_0^{i-1}) - H(y_i|u_i)\\
&- H(s_i|M,s_{i+1}^{N-1}) + H(s_i|u_i),\\
&\leq H(y_i) - H(y_i|u_i) - H(s_i) + H(s_i|u_i), && (19)\\
&= I(y_i;u_i) - I(s_i;u_i),
\end{aligned}
$$

where inequation (19) is based on the fact that s_i is independent of M and s_{i+1}^{N-1}. ∎

The next lemma presents us the characterization of $\mathcal{C}_{ave}(q,d)$, which is the same as $\mathcal{C}(q,d)$. We omit its proof as any code for (N,M,q,d) is a code for $(N,M,q,d)_{ave}$, therefore $\mathcal{R}_{ave} \geq \mathcal{C}(q,d)$. The converse part is the same as previous one.

Lemma 4. For a given rewriting cost function $\varphi(\cdot,\cdot)$, $\mathcal{C}_{ave}(q,d) = \mathcal{C}(q,d) = \max_{\substack{P_{U|S},P_{X|U,S}\\ P_{SX}\in\mathcal{P}(q,d)}} \{I(Y;U) - I(U;S)\}$, where $U \to (X,S) \to Y$ is a Markov chain.

IV. A code construction for binary degraded and symmetric Noisy WEM

Let $\mathcal{P}^s(q,d) = \{P_{SX} \in \mathcal{P}(\mathcal{X}\times\mathcal{X}) : P_S = P_X, S \sim U(q), E(\varphi(S,X)) \leq d\}$ be the set of joint probabilities with uniform marginal distributions. Let *symmetric rewriting capacity* be defined as $\mathcal{R}^s(q,d) = \max_{P_{SX}\in\mathcal{P}^s(q,d)} H(X|S)$. Let W_{SX} be $\arg\max_{P_{XS}\in\mathcal{P}^s(q,d)} H(X|S)$. We call $\mathcal{W} = (\mathcal{X},\mathcal{X},W_{X|S})$ the *WEM channel*.

We say $\mathbb{Q} = (\mathcal{X},\mathcal{Z},Q_{Z|X})$ is *degraded* with respect to $\mathbb{W} = (\mathcal{X},\mathcal{Y},W_{Y|X})$ (which we denote by $\mathbb{Q} \preccurlyeq \mathbb{W}$) if there exists a channel $\mathbb{P} = (\mathcal{Y},\mathcal{Z},P_{Y|Z})$ such that for all $z \in \mathcal{Z}$ and $x \in \mathcal{X}$, we have $Q_{Z|X}(z|x) = \sum_{y\in\mathcal{Y}} W_{Y|X}(y|x)\cdot P_{Z|Y}(z|y)$.

In this section, we consider symmetric rewriting capacity, and present a code construction for noisy WEM when the WEM channel $\mathcal{W}$ is degraded with respect to the symmetric storage channel $\mathcal{P}$. We focus on binary cells, namely, $|\mathcal{X}| = 2$. We call such WEM a *binary degraded and symmetric noisy WEM*. (Note that when the flipping rates meet $W_{X|S} > P_{Y|X}$, the degradation condition is naturally satisfied.)

A. A nested polar code construction for binary degraded and symmetric noisy WEM with an average rewriting cost constraint

1) A brief introduction to binary polar codes [2]: Let $W : \mathcal{X} \to \mathcal{Y}$ be a binary-input discrete memoryless (DMC) channel. Let $G_2^{\otimes n}$ be n-th Kronecker product of $\left(\begin{smallmatrix}1 & 0\\ 1 & 1\end{smallmatrix}\right)$. Let $Z(W) = \sum_{y\in\mathcal{Y}} \sqrt{W_{Y|X}(y|0)W_{Y|X}(y|1)}$.

The polar code $C_N(F,u_F)$ is $\{x_0^{N-1} = u_0^{N-1}G_2^{\otimes n} : u_{F^c} \in \{0,1\}^{|F^c|}\}$, where $\forall F \subseteq \{0,1,\cdots,N-1\}$, u_F is the subvector $u_i : i \in F$, and $u_F \in \{0,1\}^{|F|}$.

Let $W_N^{(i)} : \{0,1\} \to \mathcal{Y}^N \times \{0,1\}^i$ be a sub-channel with $W_N^{(i)}(y_0^{N-1},u_0^{i-1}|u_i) \stackrel{def}{=} \frac{1}{2^{N-1}} \sum_{u_{i+1}^{N-1}} \prod_{i=0}^{N-1} W_{Y|X}(y_i|(u_0^{N-1}G_2^{\otimes n})_i)$, and $(u_0^{N-1}G_2^{\otimes n})_i$ denotes the i-th element of $u_0^{N-1}G_2^{\otimes n}$.

2) 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 the XOR operation over GF(2). (Many cost functions, such as the Hamming-distance based cost function satisfies this constraint.)

Algorithm IV.1 A code construction for binary degraded and symmetric noisy WEM and storage channel $\mathcal{P}$

1. Let $C_N(F_{\mathcal{P}}, u_{F_{\mathcal{P}}})$ be a polar code [2] designed for the storage channel $\mathcal{P}$, where $F_{\mathcal{P}} = \{i \in \{0, 1, \cdots, N-1\} : Z(\mathcal{P}_N^{(i)}) \geq 2^{-N^\beta}\}$ and $u_{F_{\mathcal{P}}}$ is set to 0.
2. Let $C_N(F_{\mathcal{W}}, u_{F_{\mathcal{W}}})$ be a polar code designed for the WEM channel $\mathcal{W}$, where $F_{\mathcal{W}} = \{i \in \{0, 1, \cdots, N-1\} : Z(\mathcal{W}_N^{(i)}) \geq 2^{-N^\beta}\}$ and $F_{\mathcal{P}} \subseteq F_{\mathcal{W}}$.
3. The $(N, M, 2, d)_{ave}$ code is $\mathcal{C} = C_N(F_{\mathcal{P}}, u_{F_{\mathcal{P}}}) = \{C_N(F_{\mathcal{W}}/F_{\mathcal{P}}, u_{F_{\mathcal{W}/F_{\mathcal{P}}}}(i))\}$, where $u_{F_{\mathcal{W}/F_{\mathcal{P}}}}(i)$ is the binary representation form of $i \in \{0, \cdots, M-1\}$.

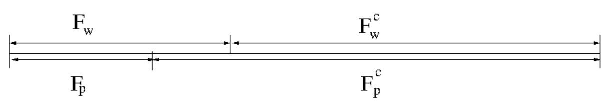

Fig. 2. Illustration of relationship between $F_{\mathcal{W}}$ and $F_{\mathcal{P}}$. The line represents the indexes, and $F_{\mathcal{W}}$ and $F_{\mathcal{P}}$ are the frozen set for the WEM channel $\mathcal{W}$ and the storage channel $\mathcal{P}$, respectively.

The code construction is presented in Algorithm IV.1, where we use nested polar codes (i.e., the polar code for channel coding [2] and the polar code for WEM [15]) to design noisy WEM codes. The fact that $F_{\mathcal{P}} \subseteq F_{\mathcal{W}}$ follows from [12, lemma 4.7]. Fig. 2 presents us a pictorial presentation of $F_{\mathcal{W}}$ and $F_{\mathcal{P}}$.

The rewriting function is presented in Algorithm IV.2. It is very similar to that of [15] except for how to set u_F.

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 a common and uniformly distributed message, and $+$ is over GF(2).
2. Apply SC (Successive Cancellation) encoding [12] to v_0^{N-1}, and this results in a vector $u_0^{N-1} = \hat{U}(v_0^{N-1}, u_{F_{\mathcal{W}/F_{\mathcal{P}}}}(i))$, that is, $u_j =$
$$\begin{cases} (u_{F_{\mathcal{W}/F_{\mathcal{P}}}}(i))_j & \text{if } j \in F_{\mathcal{W}/F_{\mathcal{P}}} \\ 0 & \text{if } j \in F_{\mathcal{P}} \\ m & \text{with probability } \frac{\mathcal{W}(u_0^{j-1}, v_0^{j-1}|m)}{\sum_{m'} \mathcal{W}(u_0^{j-1}, v_0^{N-1}|m')} \end{cases}$$
and $\hat{y}_0^{N-1} = u_0^{N-1} G_2^{\otimes n}$.
3. $y_0^{N-1} = \hat{y}_0^{N-1} + g_0^{N-1}$.

The decoding function is presented in Algorithm IV.3, where we use the SC (Successive Cancellation) decoding [2] to assure $\lambda^{(N)} \to 0$ as $N \to 0$.

Algorithm IV.3 The decoding operation $u_{F_{\mathcal{W}/F_{\mathcal{P}}}}(i) = \mathbf{D}(x_0^{N-1})$.

1. $\hat{y}_0^{N-1} = x_0^{N-1} + g_0^{N-1}$.
2. Apply SC decoding to $\hat{y}_0^{N-1}$, and this results in y_0^{N-1}, i.e., $y_j =$
$$\begin{cases} 0 & \text{if } j \in F_{\mathcal{P}} \\ \arg_m \max \mathcal{P}_N^{(j)}(y_0^{j-1}, \hat{y}_0^{N-1}|m) & \text{else} \end{cases}$$
3. $u_{F_{\mathcal{W}/F_{\mathcal{P}}}}(i) = (y_0^{N-1}(G_2^{\otimes n})^{-1})_{u_{F_{\mathcal{W}/F_{\mathcal{P}}}}}$.

3) Theoretical performance analysis:

a) Code analysis:

Theorem 5. For a binary degraded symmetric noisy WEM, fix d, $\forall \mathcal{R} \leq \mathcal{R}^s(2, d)_{ave} - H(P_{Y|X})$ and any $0 < \beta < \frac{1}{2}$, there exists a sequence of nested polar codes of length N with rates $R \leq \mathcal{R}$ so that under the above rewriting and decoding operations, $\bar{D} \leq d + O(2^{-N^\beta})$, $\lambda^{(N)} \leq O(2^{-N^\beta})$, and the rewriting as well as the decoding operation complexity is $O(N \log N)$.

Proof: Let ϵ and $0 < \beta < \frac{1}{2}$ be some constants. $F_{\mathcal{P}}$ and $F_{\mathcal{W}}$ are $F_{\mathcal{W}} = \{i : Z(\mathcal{W}_N^{(i)}) \geq 2^{-N^\beta}\}$, $F_{\mathcal{P}} = \{i : Z(\mathcal{P}_N^{(i)}) \geq 2^{-N^\beta}\}$. Based on [12, lemma 2.6] $\lim_{n \to \infty} Pr(Z(\mathcal{W}_N^{(i)}) \geq 2^{-N^\beta}) = 1 - I(\mathcal{W}) = \mathcal{R}^s(2, d)$, thus $\frac{|F_{\mathcal{W}}|}{N} \leq \mathcal{R}^s(2, d) + \epsilon$ for sufficiently large N. Similarly, $\frac{|F_{\mathcal{P}}|}{N} \geq H(P_{Y|X}) - \epsilon$ for sufficiently large N.

As mentioned, $F_{\mathcal{P}} \subseteq F_{\mathcal{W}}$, thus $R = \frac{|F_{\mathcal{W}}| - |F_{\mathcal{P}}|}{N} \leq \mathcal{R}^s(2, d) - H(P_{Y|X})$.

Since the rewriting function is similar to that of [15], the rewriting cost is guaranteed by [15, theorem 8], i.e., $\bar{D} \leq d + O(2^{-N^\beta})$.

The error probability $\lambda^{(N)}$ is guaranteed by [2, theorem 4], that is $\lambda^{(N)} \leq \sum_{i \in F_{\mathcal{P}}^c} Z(\mathcal{P}_N^{(i)}) \leq O(2^{-N^\beta})$. ■

b) Covering radius of polar codes: In the following, we present *covering radius* to theoretically upper bound the maximal rewriting cost when the cost metric is the Hamming distance between old and new cell levels.

Let the polar code ensemble be $C_N(F) = \{C_N(F, u_F) : u_F \in \{0, 1\}^{|F|}\}$. Let the covering radius of $C_N(F)$ (which we denote $c_H(C_N(F))$) be $\max_{i,j} \min_{\substack{x_0^{N-1} \in C_N(F, u_F(i)) \\ y_0^{N-1} \in C_N(F, u_F(j))}} d_H(x_0^{N-1}, y_0^{N-1})$, where $d_H(x_0^{N-1}, y_0^{N-1})$ is the Hamming distance between x_0^{N-1} and y_0^{N-1}.

Lemma 6. $c_H(C_N(F)) = \min_{l \in F^c} 2^{wt(l)}$, where $wt(l)$ is the number of ones in (i.e., Hamming weight of) the binary representation of l.

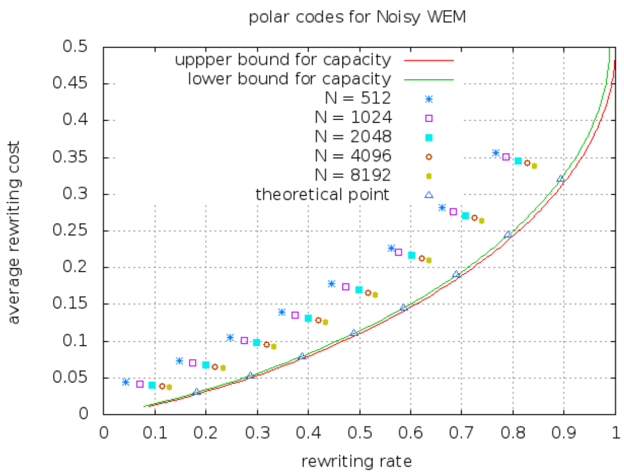

Fig. 3. Experimental performance for noisy WEM with an average cost constraint for polar code with various lengths, where the x-axis is the rewriting rate, the y-axis the average rewriting cost, and the theoretical points are those points (R, d) ($R \in \{0.2, 0.3, \cdots, 0.9\}$) satisfying $R = H(d) - H(0.001)$.

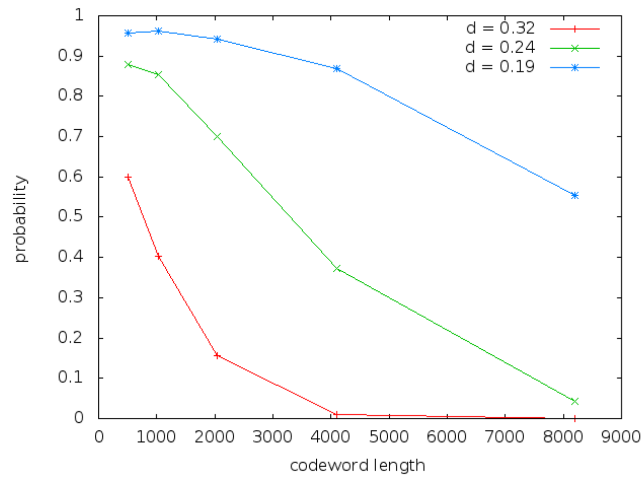

Fig. 4. Experimental performance for noisy WEM with a maximal cost constraint with $d = 0.32, 0.24$ and 0.19, respectively, where the x-axis is the codeword length, and y-axis is the empirical probability $Q(\varphi(y_0^{N-1}, x_0^{N-1}) \geq 1.1d)$.

Proof: $c_H(C_N(F))$

$$\begin{aligned}
&= \max_{i,j} \min_{\substack{x_0^{N-1} \in C_N(F, u_F(i)) \\ y_0^{N-1} \in C_N(F, u_F(j))}} d_H(x_0^{N-1}, y_0^{N-1}), \\
&= \max_{i,j} \min_{\substack{x_0^{N-1} \in C_N(F, u_F(i)) \\ y_0^{N-1} \in C_N(F, u_F(j))}} wt(x_0^{N-1} - y_0^{N-1}), \\
&= \max_{i,j} \min_{z_0^{N-1} \in C_N(F, u_F(i) + u_F(j))} wt(z_0^{N-1}), \\
&= \max_{k} \min_{l \in F^c} 2^{wt(l)}, \qquad (20) \\
&= \min_{l \in F^c} 2^{wt(l)},
\end{aligned}$$

where eq. (20) is based on [12, lemma 6.2], i.e., the minimal distance of polar code $C_N(F, u_F)$ is $\min_{l \in F^c} 2^{wt(l)}$. ∎

The above results can be generalized to the following polar codes $C_{N,M}(F) = \bigcup_{i=0}^{M-1} C_N(F, u_F(i))$, where $\{u_F(i)\}$ forms a group under binary operations in GF(2).

4) Experimental performance: The experimental performance is presented in Fig. 3, where the rewriting cost function is the Hamming distance between old and new cell states, the upper bound of $\mathcal{C}(2, d)$ is $H(d)$ [1], the storage channel $\mathcal{P}$ is the binary symmetric channel with flipping rate $p = 0.001$, and the lower bound is $H(d) - H(p)$. $\lambda^{(N)}$ is set to be around 10^{-5}.

We can see that the rates and the average rewriting costs approach those of points of $H(d) - H(0.001)$ as the length of codeword increases. Longer codewords are needed for further approaching the lower bound.

B. A nested polar code construction for binary degraded and symmetric noisy WEM with a maximal rewriting cost constraint

The code construction in Algorithm IV.1, the rewriting function in Algorithm IV.2 and the decoding function in Algorithm IV.3 can be applied to noisy WEM codes with a maximal rewriting cost constraint as well.

Similar to the analysis of Theorem 5, we obtain the following result for the theoretical performance of the proposed code construction.

Theorem 7. For a binary degraded symmetric noisy WEM, fix d, δ, $\forall \mathcal{R} \leq \mathcal{R}^s(2, d) - H(P_{Y|X})$ and any $0 < \beta < \frac{1}{2}$, there exists a sequence of nested polar codes of length N with rates $R \leq \mathcal{R}$, so that under the above rewriting operation and decoding operation, the probability that the rewriting cost between a current codeword $\forall y_0^{N-1}$ and its updated codeword x_0^{N-1} larger than $d + \delta$ is bounded by $Q(\varphi(y_0^{N-1}, x_0^{N-1}) \geq d + \delta) < 2^{-N^\beta}$, $\lambda^{(N)} \leq O(2^{-N^\beta})$, and the decoding and rewriting operations' complexity of the code is $O(N \log N)$.

We present our experimental results in Fig. 4. The rewriting cost function, storage channel $\mathcal{P}$, and $\lambda^{(N)}$ are the same as those of the previous subsection. We let $\delta = 0.1d$, and $d = 0.32$, 0.24, and 0.19, respectively. The empirical probability $Q(\varphi(y_0^{N-1}, x_0^{N-1}) \geq 1.1d)$ is presented in fig. 4. As predicted by Theorem 7 it decreases (nearly exponentially) as the length of codeword increases. However, even longer codewords are needed to make the probability to be truly negligible.

V. Concluding remarks

In this paper, we analyse the capacity for noisy WEM, and present a code construction for binary degraded and symmatric noisy WEM. The code construction is both theoretically analyzed and experimentally verified. We are interested in extending the code construction to q-ary cells, and to more general settings regarding channel degradation. Those remain as our future research directions.

VI. Acknowledgement

This work was supported in part by the NSF Grant CCF-1217944.

References

[1] R. Ahlswede and Z. Zhang, "Coding for write-efficient memory," *Information and Computation*, vol. 83, no. 1, pp. 80–97, October 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," *IEEE Transacation on Information Theory*, vol. 59, no. 8, pp. 5088–5101, August 2013.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[5] I. Criszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Transaction on Information Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[6] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Orlando, FL, USA: Academic Press, Inc., 1982.

[7] 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.

[8] S. I. Gel'fand and M. S. Pinsker, "Coding for Channel with Random Parameters," *Problems of Control Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[9] D. Ielmini, S. Lavizzari, D. Sharma, and A. Lacaita, "Physical interpretation, modeling and impact on phase change memory (PCM) reliability of resistance drift due to chalcogenide structural relaxation," *IEEE International Electron Devices Meeting, 2007. IEDM 2007*, pp. 939–942, 2007.

[10] A. Jiang, V. Bohossian, and J. Bruck, "Rewriting codes for joint information storage in flash memories," *IEEE Trans on Inf Theory*, vol. 56, no. 10, pp. 5300–5313, October 2010.

[11] A. Jiang, Y. Li, E. E. Gad, M. Lanberg, and J. Bruck, "Joint rewriting and error correction in write-once memories," in *Proc. IEEE International Symposium on Information Theory(ISIT)*, Istanbul, Turkey, June 2013, pp. 1067–1071.

[12] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, EPFL, 2010.

[13] 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.

[14] Q. Li, "Compressed Rank Modulation," in *Proc. 50th Annual Allerton Conference on Communication, Control and Computing (Allerton)*, Monticello, IL, October 2012.

[15] Q. Li and A. Jiang, "Polar codes are optimal for Write-efficient Memories," in *proc 51th Annual Allerton Conference on Communication, Control and Computing (Allerton)*, Monticello, IL, October 2013.

[16] Q. Li, A. Jiang, and E. F. Haratsch, "Noise Modeling and Capacity Analysis for NAND Flash Memories," in *Proc. IEEE International Symposium on Information Theory(ISIT)*, Honolulu, HI, June 2014.