Polar Coding for Noisy Write-Once Memories

Eyal En Gad[1], **Yue Li**[1,4], **Joerg Kliewer**[2], **Michael Langberg**[3],
Anxiao (Andrew) Jiang[4] and **Jehoshua Bruck**[1]
[1]*California Institute of Technology, Pasadena, CA 91125*
[2]*New Jersey Institute of Technology, Newark, NJ 07102*
[3]*SUNY at Buffalo, Buffalo, NY 14260*
[4]*Texas A&M University, College Station, TX 77843*
{eengad, yli, bruck}@caltech.edu jkliewer@njit.edu mikel@buffalo.edu ajiang@cse.tamu.edu

***Abstract*—We consider the noisy write-once memory (WOM) model to capture the behavior of data-storage devices such as flash memories. The noisy WOM is an asymmetric channel model with non-causal state information at the encoder. We show that a nesting of non-linear polar codes achieves the corresponding Gelfand-Pinsker bound with polynomial complexity.**

I. Introduction

Write-once memory is a model for data-storage devices, where a set of binary cells are used to store data, and the cell levels can only increase when the data is rewritten. This model was first proposed in [13], and it is related to earlier models for channel coding with state information [4], [11]. The WOM model was recently identified as a good model for flash memories, where rewriting delays expensive block erasures and by that leads to better preservation of cell quality and higher performance [8], [15].

In this paper we propose a coding scheme for WOM that incorporates error-correction capabilities to protect from noise in the process of writing data to the memory. We consider a stochastic model of writing noise due to Heegard [6], and propose the first coding scheme that achieves the capacity of the model with polynomial complexity. The coding scheme is based on a nesting of non-linear polar codes.

The majority of previous work on WOM codes do not consider error correction, e.g., in [14] where an algebraic zero-error capacity-achieving coding scheme was presented. A different capacity-achieving scheme, based on polar coding, was proposed in [1]. In contrast, there is little work available when errors are considered. For example, in [3] a capacity-achieving coding scheme was presented which corrects a negligible fraction of errors. Further, previous work by the authors [9] proposes a non-capacity-achieving solution based on high-rate polar codes. Other non-capacity-achieving error-correcting WOM schemes were proposed in [15], [16].

The main contribution of this work is the derivation of a nested polar coding scheme for the noisy WOM channel, with an asymptotically-optimal rate and polynomial complexity. As typically with polar coding schemes, the algorithmic complexity of the proposed scheme is given as $O(n \log n)$ under a decoding error probability of $2^{-\Omega(n^{1/2-\delta})}$ for a block length n and any $\delta > 0$.

Comparing with the capacity-achieving polar WOM coding scheme of [1], we make two important contributions. First, the scheme of this paper poses a capability to correct errors, while the scheme in [1] did not have this capability. Second, the scheme in [1] requires a large amount of randomness to be shared between the encoder and the decoder. The scheme that we propose in this paper requires a significantly smaller amount of randomness to be shared. In fact, under the reasonable approximation of the cell states to be i.i.d., the requirement for shared randomness is completely removed. The reduction of shared randomness is done by the use of non-linear polar codes as proposed in [7].

We note that the coding scheme presented in this paper is very similar to the scheme presented in [5], for communication in broadcast channels. While we were unaware of the result of [5] and developed the scheme independently, this paper also has two contributions that were not shown in [5]. First, we connect the scheme to the application of data storage and flash memories, that was not considered in the previous work. Second, the analysis in [5] holds for channels that exhibit a certain technical condition of degradation. In this paper we show that the model of WOM with writing noise in fact exhibit the required degradation condition, and by that show that the scheme achieves the capacity of the considered model.

II. WOM Channel Model

A WOM is composed of n cells, where each cell has a state from some finite alphabet. In this paper we assume that the alphabet is $\{0, 1\}$. The reason for this is that polar codes are better understood with binary codeword symbols. However, there is no other fundamental difficulty to extend the results of the paper to non-binary alphabets (see, e.g., [1]). The main property of the WOM is that a cell at state 0 can change its state to 1, but once the cell state is 1, it cannot be changed anymore. The state of the cells is known to a user that wishes to store information on the memory. If some of the cells are in state 1, a code is required for the reliable storage of information, since not every sequence of n bits can be stored directly.

We consider a *stochastic* i.i.d. WOM model. Let S be a Bernoulli random variable that corresponds to the state of a

This work was supported in part by Intellectual Ventures, NSF grants 1218005, 1439465, 1440001, 1440014, and 1038578 and the US-Israel Binational Science Foundation (BSF) under Grant No. 2010075.

cell. Since the cells are i.i.d., we define the distribution of a cell state by

$$P(S = 1) = \beta, \quad (1)$$

for some $\beta \in [0, 1]$. We note that this model is not completely accurate for the application of rewriting in flash memories. In rewriting, the memory state is determined by the previous write to the memory, in which the codeword indices are not necessarily distributed independently. However, simulation results show that the i.i.d. approximation performs well in practice. To make the cell states precisely i.i.d., it was proposed in modifications M3 and M4 in [1, Section IV] to share a random permutation between the encoder and the decoder. However, it is not clear if this solution could be implemented efficiently in practice.

The focus of this work is on memories with *write errors*, where the corresponding bit channel is shown in Fig. 1. Here X and Y are Bernoulli random variables that correspond to the input and output of a single bit in the memory. If the cell state S is 1, the output Y will be 1 as well, regardless of the value of the input X, corresponding to a stuck bit. Otherwise, if the cell state S is 0, we assume that the memory exhibits a symmetric writing error with crossover probability α. This behavior is summarized by

$$P_{Y|XS}(1|x,s) = \begin{cases} \alpha & \text{if } (x,s) = (0,0) \\ 1-\alpha & \text{if } (x,s) = (1,0) \\ 1 & \text{if } s = 1. \end{cases} \quad (2)$$

Finally, in our model of study, we limit the number of cells that the encoder attempts to change from 0 to 1 in the memory. This limitation lends itself naturally to the application of WOM to settings in which multiple writes are considered. We express this limitation by a parameter ϵ that bounds the expectation of the input X given that $S = 0$ as follows

$$E(X|S = 0) \leq \epsilon. \quad (3)$$

The capacity of the WOM model is given in the following theorem.

Theorem 1. *[6, Theorem 4] The capacity of the memory described by (1), (2), and (3) is* $\mathcal{C} = (1-\beta)[h(\epsilon * \alpha) - h(\alpha)]$, *where* $\epsilon * \alpha \equiv \epsilon(1-\alpha) + (1-\epsilon)\alpha$.

The nested polar coding scheme of this paper achieves the capacity of Theorem 1. The presentation of the scheme requires some notation from polar coding.

III. Polar Coding Notation

For a positive integer n, let $[n] \equiv \{1, 2, \ldots, n\}$. Let $X_1^n = (X_1, X_2, \ldots, X_n)$, $Y_1^n = (Y_1, Y_2, \ldots, Y_n)$ and $S_1^n = (S_1, S_2, \ldots, S_n)$ be i.i.d. copies of X, Y and S, respectively. For integers $i < j$, let X_i^j represent the subvector $(X_i, X_{i+1}, \ldots, X_j)$ and for a set $\mathcal{A} \subset [n]$ let $X_{\mathcal{A}}$ represent the subvector $\{X_i\}_{i \in \mathcal{A}}$.

Let n be a power of 2. Define a matrix $G_n = G^{\otimes \log_2 n}$ where $G = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ and $\otimes$ denotes the Kronecker power. A subset of the rows of G_n serves as a generator matrix in

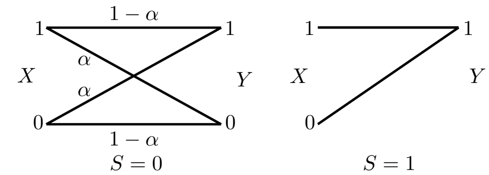

Fig. 1. A binary WOM with write errors.

linear polar coding schemes. Let U_1^n be the product $U_1^n = X_1^n G_n^{-1} = X_1^n G_n$. A subset of the coordinates of the vector U_1^n will contain the message to be stored in the memory.

Our analysis uses the Bhattacharyya parameters. For a conditional distribution $P_{Y|X}$, where X is a Bernoulli random variable, the Bhattacharyya parameter is defined by $Z_B(P_{Y|X}) \triangleq \sum_y \sqrt{P_{Y|X}(y|0) P_{Y|X}(y|1)}$. The Bhattacharyya parameter serves as an upper bound on the decoding error probability in polar codes for symmetric settings. Similarly, define the conditional Bhattacharyya parameter as $Z(X|Y) \triangleq 2\sum_y \sqrt{P_{X,Y}(0,y) P_{X,Y}(1,y)}$. The conditional Bhattacharyya parameter serves as a decoding error probability bound in polar codes for asymmetric settings. Note that the Bhattacharyya parameter and the conditional Bhattacharyya parameter are equal if X is distributed uniformly.

IV. Coding Scheme

We start by presenting a rough overview of our coding technique followed by a formal presentation. The achievability of Theorem 1 is shown in [4], [6] by random coding. The distribution by which the codewords are drawn in this achievability proof is called the capacity-achieving distribution, and it is used in our scheme. We denote this distribution by $P_{X|S}$. According to the proof of [6, Theorem 4], we have

$$P_{X|S}(1|0) = \epsilon, \qquad P_{X|S}(1|1) = \frac{\epsilon(1-\alpha)}{\epsilon * \alpha}. \quad (4)$$

Our construction is based on a combination of two methods: *non-linear* polar coding [7] and *nested* polar coding [10]. Non-linear polar codes allow to achieve the capacity of asymmetric channels and sources. In nested polar coding for Gelfand-Pinsker-type problems, the encoder first *compresses* the state using a lossy source code, and then transmits the compressed state together with the source message using a channel code. The consideration of the compressed state in the choice of codeword allows the conditional distribution of the codeword given the state to approximate the capacity-achieving distribution of Equation (4).

Let s_1^n, u_1^n, x_1^n and y_1^n be the realizations of the random variables S_1^n, U_1^n, X_1^n and Y_1^n, respectively. In a channel polar coding scheme, a vector u_1^n is used for representing the source message and a frozen vector. The channel input is the codeword $x_1^n = u_1^n G_n$. The coding scheme *polarizes* the conditional entropies $H(U_i|U_1^{i-1}, Y_1^n)$ for the different coordinates i of the vector U_1^n. In a nested polar coding scheme, the vector U_1^n is also polarized with respect to the conditional entropies $H(U_i|U_1^{i-1}, S_1^n)$. In our case the conditional entropies $H(U_i|U_1^{i-1}, S_1^n)$ are defined according to the capacity-achieving conditional distribution of Equation (4).

In addition, in non-linear polar codes, the vector U_1^n is also polarized with respect to the conditional entropies $H(U_i|U_1^{i-1})$. We take advantage of these three different polarizations of the vector U_1^n in our coding scheme. It is useful to define the polarized sets according to the conditional Bhattacharyya parameter, as follows:

$$A \equiv \{i : Z(U_i|U_1^{i-1}, Y_1^n) \leq 2^{-n^{1/2-\delta}}\},$$
$$B \equiv \{i : Z(U_i|U_1^{i-1}, S_1^n) \geq 1 - 2^{-n^{1/2-\delta}}\},$$
$$C \equiv \{i : Z(U_i|U_1^{i-1}) \leq 2^{-n^{1/2-\delta}}\},$$
$$D \equiv \{i : Z(U_i|U_1^{i-1}) \geq 1 - 2^{-n^{1/2-\delta}}\},$$

for some small $\delta > 0$. The sets A, B, C and D refer to sets of coordinates of the vector u_1^n. Also define the sets A^c, B^c, C^c and D^c to be the complements of A, B, C and D with respect to $[n]$. Intuitively, set A contains the coordinates of u_1^n that can be decoded reliably given y_1^n. Set B contains the coordinates that have small relation to s_1^n. Set C is composed of coordinates that do not contain much information about x_1^n, and set D is composed of coordinates that contain almost all the information about x_1^n. Note that those sets are not disjoint. The relation between the sets is depicted in Fig. 2.

The proposed coding scheme takes advantage of the polarized sets as follows. First, the encoder performs a lossy compression of the state s_1^n by the method of [7]. The compression is performed according to an information set $\mathcal{I}_S$, defined to be the set of coordinates that are *not* in the union $B \cup C$. The complement of the information set with respect to $[n]$ is denoted as $\mathcal{I}_S^c = B \cup C$. Each bit u_i for $i \in \mathcal{I}_S$ is set randomly to a value u with probability $P_{U_i|U_1^{i-1},S_1^n}(u|u_1^{i-1}, s_1^n)$. The coordinates in the set B are distributed almost uniformly given (U_1^{i-1}, S_1^n), and therefore they do not affect the joint distribution (X_1^n, S_1^n). For that reason we can place the source message in those coordinates. Since we also need the message to be decoded reliably given Y_1^n, we restrict it to the coordinates in the set A, which are almost deterministic given (U_1^{i-1}, Y_1^n). Taking both restrictions into account, we set the coordinates of u_1^n in the intersection $A \cap B$ to be equal to the source message.

The rest of the coordinates of u_1^n are set by the encoder according to a set of Boolean functions, as in [7]. Since this set of functions describes the code, it is known to both the encoder and the decoder. For a positive integer i, a Boolean function is denoted by $\lambda_i : \{0,1\}^{i-1} \to \{0,1\}$. Remember that the coordinates in the set $\mathcal{I}_S$ are determined according to the state s_1^n, and that the coordinates in $A \cap B$ are determined by the information message. The rest of the coordinates of u_1^n will be determined according to a set of Boolean functions $\lambda_{\mathcal{I}_S^c \setminus (A \cap B)} = \{\lambda_i\}_{\mathcal{I}_S^c \setminus (A \cap B)}$. The encoder sets a bit u_i according to the function $\lambda_i(u_1^{i-1})$.

Finally, remember that the set A contains the reliable coordinates of u_1^n given y_1^n. Therefore, the decoder can guess the coordinates in A, and will estimate the rest of the coordinates according to $\lambda_{\mathcal{I}_S^c \setminus A}$. A coordinate i in $\mathcal{I}_S^c$ can be decoded this way, since if $i \in A$, u_i can be estimated reliably, and if $i \in \mathcal{I}_S^c \setminus A$, then u_i can be recovered by λ_i. However, a coordinate $i \in \mathcal{I}_S$ can be decoded reliably only if $i \in A$, since we do not use Boolean functions for the set $\mathcal{I}_S$. Therefore, a reliable decoding requires $\mathcal{I}_S$ to be a subset of A. While we do not know if this fact always holds, we can show that the set difference $\mathcal{I}_S \setminus A$ is very small. Therefore, our strategy is to store the vector $u_{\mathcal{I}_S \setminus A}$ in additional cells. Since we show that the fraction $|\mathcal{I}_S \setminus A|/n$ approaches zero for large n, this strategy will not affect the asymptotic rate of the scheme or the expected weight of the codewords.

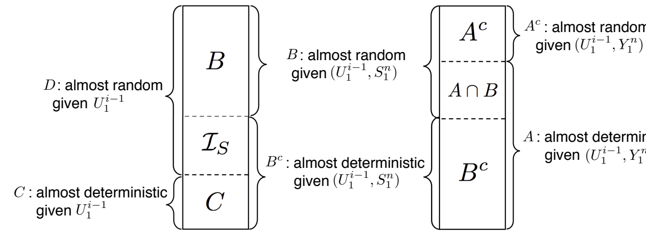

Fig. 2. Different polarizations of $U_1^n = X_1^n G_n$.

The additional cells should be coded as well to ensure their reliability. In this paper we assume that the additional cells are protected by a non-linear channel polar code, without utilizing the state information at the encoder. We do not discuss the details of the coding of the additional cells. However, we note that since the block length of the additional cells is much smaller than the main block length, the probability of decoding error at the additional cells is higher. In particular, the probability of decoding error at the additional cells is $2^{-\Omega(n^{1/2-\delta})}$ for any $\delta' > \delta$, while the decoding error probability in the main block of cells is $O(2^{-n^{1/2-\delta}})$, as in standard polar codes. The details of this issue are discussed in the longer version of this paper [2]. We now describe the coding scheme formally.

Construction 2.

Encoding

Input: *a message* $m_1^k \in \{0,1\}^k$ *and a state* $s_1^n \in \{0,1\}^n$.
Output: *a codeword* $x_1^n \in \{0,1\}^n$.

1) *Let* $\mathbf{u}_{A \cap B} = m_1^k$. *Then for* i *from* 1 *to* n, *if* $i \in \mathcal{I}_S$, *set*

$$u_i = \begin{cases} 0 & \text{with probability } P_{U_i|U_1^{i-1},S_1^n}(0|u_1^{i-1}, s_1^n) \\ 1 & \text{with probability } P_{U_i|U_1^{i-1},S_1^n}(1|u_1^{i-1}, s_1^n), \end{cases}$$

and if $i \in \mathcal{I}_S^c \setminus (A \cap B)$, *set* $u_i = \lambda_i(u_1^{i-1})$.
2) *The vector* $u_{\mathcal{I}_S \setminus A}$ *is stored in additional cells. Finally, store the codeword* $x_1^n = u_1^n G_n$.

Decoding

Input: *a noisy vector* y_1^n.
Output: *a message estimation* $\hat{m}_1^k$.
Estimate u_1^n *by* $\hat{u}_1^n(y_1^n, \lambda_{\mathcal{I}_S^c \setminus A})$ *as follows:*

1) *Recover the vector* $u_{\mathcal{I}_S \setminus A}$ *from the additional cells, and assign* $\hat{u}_{\mathcal{I}_S \setminus A} = u_{\mathcal{I}_S \setminus A}$.
2) *For* i *from* 1 *to* n, *set*

$$\hat{u}_i = \begin{cases} \arg\max_u P_{U_i|U_1^{i-1},Y_1^n}(u|\hat{u}_1^{i-1}, y_1^n) & \text{if } i \in A \\ \lambda_i(\hat{u}_1^{i-1}) & \text{if } i \in \mathcal{I}_S^c \setminus A. \end{cases}$$

3) *Return the estimated message* $\hat{m}_1^k = \hat{u}_{A \cap B}$.

Let n' be the number of additional cells needed to store the vector $u_{\mathcal{I}_S \setminus A}$. The main theorem of this paper addresses the properties of Construction 2.

Theorem 3. *Let the message m_1^k be distributed uniformly. Then for any constants $\delta' > \delta > 0$ and $\epsilon' > \epsilon > 0$, there exists a set of Boolean functions $\lambda_{\mathcal{I}_S \setminus (A \cap B)}$ for which Construction 2 satisfies the following:*

1) *The rate of the scheme is asymptotically optimal. Formally, $\lim_{n\to\infty} k/(n+n') = \mathcal{C}$.*
2) $E[X_i|S_i = 0] \le \epsilon'$.
3) *The decoding error probability is $2^{-\Omega(n^{1/2-\delta'})}$.*
4) *The encoding and decoding complexities are $O(n \log n)$.*

Due to space limitations, we skip the detailed proof of Theorem 3, and refer the reader to the longer version of this paper [2]. In this short version we only prove the rate optimality of the codes (claim 1 of Theorem 3).

V. Optimality of the Rate

To prove claim 1 of Theorem 3, we combine two separate claims. First, we show in Subsection V-A that the fraction of additional cells is negligible, and next, in Subsection V-B, we show that the rate of the main block is asymptotically optimal. Together the two results prove claim 1 of Theorem 3.

A. Fraction of Additional Cells is Negligible

The vector $u_{\mathcal{I}_S \setminus A}$ is stored on n' additional cells. To ensure a reliable storage, a polar coding scheme requires only a linear amount of redundancy in $|\mathcal{I}_S \setminus A|$. Therefore, to show that the additional cells do not affect the rate and codeword weight of the scheme, it is enough to show that $\lim_{n\to\infty} |\mathcal{I}_S \setminus A|/n = 0$. To show this, we use an idea from the proof of [10, Theorem 15]. Consider the set

$$\begin{aligned}\bar{\mathcal{I}}_S =& \{i : Z(U_i|U_1^{i-1}, S_1^n) \le 2^{-n^{1/2-\delta}} \\ & \text{and } Z(U_i|U_1^{i-1}) \ge 1 - 2^{-n^{1/2-\delta}}\}.\end{aligned}$$

By the definition of the sets above, $\bar{\mathcal{I}}_S \subseteq \mathcal{I}_S$. Therefore, by the distributivity of intersection over union, we have

$$|\mathcal{I}_S \setminus A| = |A^c \cap \mathcal{I}_S| = |A^c \cap \mathcal{I}_S \setminus \bar{\mathcal{I}}_S| + |A^c \cap \bar{\mathcal{I}}_S|.$$

To show that $\lim_{n\to\infty} |A^c \cap \mathcal{I}_S|/n = 0$, we start by showing that $\bar{\mathcal{I}}_S \subseteq A$, which implies that $|A^c \cap \bar{\mathcal{I}}_S| = 0$, and therefore that

$$|A^c \cap \mathcal{I}_S| = |A^c \cap (\mathcal{I}_S \setminus \bar{\mathcal{I}}_S)| \le |\mathcal{I}_S \setminus \bar{\mathcal{I}}_S|.$$

From [7, Theorem 1], we have $\lim_{n\to\infty} |\mathcal{I}_S \setminus \bar{\mathcal{I}}_S|/n = 0$.

Therefore, showing that $\bar{\mathcal{I}}_S \subseteq A$ proves that $\lim_{n\to\infty} |A^c \cap \mathcal{I}_S|/n = 0$. To show that $\bar{\mathcal{I}}_S \subseteq A$, it is enough to show that $Z(U_i|U_1^{i-1}, Y_1^n) \le Z(U_i|U_1^{i-1}, S_1^n)$ for all $i \in [n]$, by definitions of the sets $\bar{\mathcal{I}}_S$ and A. To show that $Z(U_i|U_1^{i-1}, Y_1^n) \le Z(U_i|U_1^{i-1}, S_1^n)$ for all $i \in [n]$, we use a sequence of reductions. First, we use the following theorem:

Theorem 4. *[7, Theorem 2] Let $\mathcal{Y}$ be the alphabet of Y and let $\tilde{\mathcal{Y}} = \{0,1\} \times \mathcal{Y}$ and $\tilde{Y} = (\tilde{X} \oplus X, Y)$ where (X, Y) is independent of $\tilde{X}$. Let $\tilde{U}_1^n \equiv \tilde{X}_1^n G_n$ and define*

$$\bar{W}_{Y,i}^{(n)}(\tilde{u}_1^{i-1}, \tilde{y}_1^n|\tilde{u}_i) = P_{\tilde{U}_1^{i-1}, \tilde{Y}_1^n|\tilde{U}_i}(\tilde{u}_1^{i-1}, \tilde{y}_1^n|\tilde{u}_i).$$

Then $Z(U_i|U_1^{i-1}, Y_1^n) = Z_B(\bar{W}_{Y,i}^{(n)})$.

Theorem 4 implies that $Z(U_i|U_1^{i-1}, Y_1^n) \le Z(U_i|U_1^{i-1}, S_1^n)$ if and only if $Z_B(\bar{W}_{Y,i}^{(n)}) \le Z_B(\bar{W}_{S,i}^{(n)})$ for all $i \in [n]$. To show the latter relation, we use the notion of stochastically degraded channels. A discrete memory-less channel (DMC) $W_1 : \{0,1\} \to \mathcal{Y}_1$ is stochastically degraded with respect to a DMC $W_2 : \{0,1\} \to \mathcal{Y}_2$, denoted as $W_1 \preceq W_2$, if there exists a DMC $W : \mathcal{Y}_2 \to \mathcal{Y}_1$ such that $W_1(y_1|x) = \sum_{y_2 \in \mathcal{Y}_2} W_2(y_2|x) W(y_1|y_2)$. We use the following result on the Bhattacharyya parameters of degraded channels:

Lemma 5. *[10, Lemma 21] (Degradation of $W_i^{(n)}$) Let $W : \{0,1\} \to \mathcal{Y}_1$ $W' : \{0,1\} \to \mathcal{Y}_2$ be two discrete memory-less channels (DMC) and let $W \preceq W'$. Then for all i, $W_i^{(n)} \preceq W'^{(n)}_i$ and $Z_B(W_i^{(n)}) \ge Z_B(W'^{(n)}_i)$.*

Lemma 5 reduces the proof of $Z_B(\bar{W}_{Y,i}^{(n)}) \le Z_B(\bar{W}_{S,i}^{(n)})$ for all $i \in [n]$ to showing that $P_{\tilde{S}|\tilde{X}} \preceq P_{\tilde{Y}|\tilde{X}}$. The last degradation relation is proven in the following lemma, which completes the proof that $\lim_{n\to\infty} |\mathcal{I}_S \setminus A|/n = 0$.

Lemma 6.

$$P_{\tilde{S}|\tilde{X}} \preceq P_{\tilde{Y}|\tilde{X}}.$$

Proof: We need to show that there exists a DMC $W : \{0,1\}^2 \to \{0,1\}^2$ such that

$$P_{\tilde{S}|\tilde{X}}(\tilde{s}|\tilde{x}) = \sum_{\tilde{y} \in \{0,1\}^2} P_{\tilde{Y}|\tilde{X}}(\tilde{y}|\tilde{x}) W(\tilde{s}|\tilde{y}). \tag{5}$$

To define such channel W, we first claim that

$$P_{Y|X,S}(1|x,0) P_{X|S}(x|0) = (\epsilon * \alpha) P_{X|S}(x|1). \tag{6}$$

Equation (6) follows directly from Equation (4) since

$$\frac{P_{Y|X,S}(1|0,0) P_{X|S}(0|0)}{P_{X|S}(0|1)} = \frac{\alpha(1-\epsilon)}{\frac{\alpha(1-\epsilon)}{\epsilon*\alpha}} = \epsilon * \alpha,$$

$$\frac{P_{Y|X,S}(1|1,0) P_{X|S}(1|0)}{P_{X|S}(1|1)} = \frac{(1-\alpha)\epsilon}{\frac{(1-\alpha)\epsilon}{\epsilon*\alpha}} = \epsilon * \alpha.$$

Next, we claim that $\frac{P_{X,S}(x,1)}{P_{X,Y}(x,1)} = \frac{\beta}{(\epsilon*\alpha)(1-\beta)+\beta}$ for any $x \in \{0,1\}$, and therefore that $\frac{P_{X,S}(x,1)}{P_{X,Y}(x,1)} \in [0,1]$. This follows from

$$\begin{aligned}\frac{P_{X,S}(x,1)}{P_{X,Y}(x,1)} &\overset{(a)}{=} \frac{P_{X|S}(x|1) P_S(1)}{P_{Y,X|S}(1,x|0) P_S(0) + P_{Y,X|S}(1,x|1) P_S(1)} \\ &\overset{(b)}{=} P_{X|S}(x|1)\beta / [P_{Y|X,S}(1|x,0) P_{X|S}(x|0)(1-\beta) \\ &\quad + P_{Y|X,S}(1|x,1) P_{X|S}(x|1)\beta] \\ &\overset{(c)}{=} \frac{P_{X|S}(x|1)\beta}{(\epsilon*\alpha) P_{X|S}(x|1)(1-\beta) + P_{X|S}(x|1)\beta}\end{aligned}$$

$$=\frac{\beta}{(\epsilon * \alpha)(1-\beta)+\beta},$$

where (a) follows from the law of total probability, (b) follows from the definition of conditional probability, and (c) follows from Equations (2) and (6).

Denote the first coordinate of the random variable $\tilde{Y}$ by $\tilde{Y}_1 \equiv \tilde{X} \oplus X$, and the first coordinate of $\tilde{y}$ by y_1. The same notation is used also for $\tilde{S}$ and $\tilde{s}$. Since $\frac{P_{X,S}(x,1)}{P_{X,Y}(x,1)}$ is not a function of x and is in $[0,1]$, we can define W as following:

$$W(\tilde{s}|\tilde{y}) \triangleq \begin{cases} 1 & \text{if } \tilde{s}_1 = \tilde{y}_1 \text{ and } (s,y) = (0,0) \\ 1 - \frac{P_{X,S}(x,1)}{P_{X,Y}(x,1)} & \text{if } \tilde{s}_1 = \tilde{y}_1 \text{ and } (s,y) = (0,1) \\ \frac{P_{X,S}(x,1)}{P_{X,Y}(x,1)} & \text{if } \tilde{s}_1 = \tilde{y}_1 \text{ and } (s,y) = (1,1) \\ 0 & \text{otherwise.} \end{cases}$$

We show next that Eq. (5) holds for W defined above:

$$\begin{aligned}
&\sum_{\tilde{y}\in\{0,1\}^2} P_{\tilde{Y}|\tilde{X}}(\tilde{y}|\tilde{x})W(\tilde{s}|\tilde{y}) = \sum_{\tilde{y}\in\{0,1\}^2} P_{\tilde{Y}_1,Y|\tilde{X}}(\tilde{y}_1,y|\tilde{x})W(\tilde{s}|\tilde{y}) \\
&\overset{(d)}{=} \sum_{\tilde{y}\in\{0,1\}^2} P_{X,Y|\tilde{X}}(\tilde{y}_1 \oplus \tilde{x}, y|\tilde{x})W(\tilde{s}|\tilde{y}) \\
&\overset{(e)}{=} \sum_{\tilde{y}\in\{0,1\}^2} P_{X,Y}(\tilde{y}_1 \oplus \tilde{x}, y)W(\tilde{s}|\tilde{y}) \\
&= [P_{X,Y}(\tilde{s}_1 \oplus \tilde{x}, 0)W(\tilde{s}_1,0|\tilde{s}_1,0) \\
&\quad + P_{X,Y}(\tilde{s}_1 \oplus \tilde{x}, 1)W(\tilde{s}_1,0|\tilde{s}_1,1)]\mathbb{1}(s=0) \\
&\quad + P_{X,Y}(\tilde{s}_1 \oplus \tilde{x}, 1)W(\tilde{s}_1,1|\tilde{s}_1,1)\mathbb{1}(s=1) \\
&= \Bigg[P_{X,Y}(\tilde{s}_1 \oplus \tilde{x}, 0) \\
&\quad + P_{X,Y}(\tilde{s}_1 \oplus \tilde{x}, 1)\left(1 - \frac{P_{X,S}(\tilde{s}_1 \oplus \tilde{x}, 1)}{P_{X,Y}(\tilde{s}_1 \oplus \tilde{x}, 1)}\right)\Bigg] \cdot \mathbb{1}(s=0) \\
&\quad + P_{X,Y}(\tilde{s}_1 \oplus \tilde{x}, 1)\frac{P_{X,S}(\tilde{s}_1 \oplus \tilde{x}, 1)}{P_{X,Y}(\tilde{s}_1 \oplus \tilde{x}, 1)}\mathbb{1}(s=1) \\
&= [P_{X,Y}(\tilde{s}_1 \oplus \tilde{x}, 0) + P_{X,Y}(\tilde{s}_1 \oplus \tilde{x}, 1) \\
&\quad - P_{X,S}(\tilde{s}_1 \oplus \tilde{x}, 1)]\mathbb{1}(s=0) + P_{X,S}(\tilde{s}_1 \oplus \tilde{x}, 1)\mathbb{1}(s=1) \\
&\overset{(f)}{=} [P_X(\tilde{s}_1 \oplus \tilde{x}) - P_{X,S}(\tilde{s}_1 \oplus \tilde{x}, 1)]\mathbb{1}(s=0) \\
&\quad + P_{X,S}(\tilde{s}_1 \oplus \tilde{x}, 1)\mathbb{1}(s=1) \\
&\overset{(g)}{=} P_{X,S}(\tilde{s}_1 \oplus \tilde{x}, 0)\mathbb{1}(s=0) + P_{X,S}(\tilde{s}_1 \oplus \tilde{x}, 1)\mathbb{1}(s=1) \\
&= P_{X,S}(\tilde{s}_1 \oplus \tilde{x}, s) \overset{(h)}{=} P_{X,S|\tilde{X}}(\tilde{s}_1 \oplus \tilde{x}, s|\tilde{x}) \\
&= P_{\tilde{S}_1,S|\tilde{X}}(\tilde{s}_1, s|\tilde{x}) = P_{\tilde{S}|\tilde{X}}(\tilde{s}|\tilde{x}),
\end{aligned}$$

where (d) follows from the fact that $\tilde{Y}_1 = \tilde{X} \oplus X$, (e) follows from the independence of (X,Y) and $\tilde{X}$, (f) and (g) follow from the law of total probability, and (h) follows from the independence of (X,S) and $\tilde{X}$. So the channel W satisfies Equation (5) and thus the lemma holds. ■

B. The Rate of the Main Block

We need to show that $\lim_{n\to\infty} k/n = (1-\beta)[H(\alpha * \epsilon) - H(\alpha)]$. By [7, Equations (38) and (39)], we have $\lim_{n\to\infty} |B|/n = H(X|S)$, and $\lim_{n\to\infty} |A^c|/n = H(X|Y)$. So we get

$$\begin{aligned}\lim_{n\to\infty} k/n = \lim_{n\to\infty} |A \cap B|/n \geq \lim_{n\to\infty} (|B| - |A^c|) \\ = H(X|S) - H(X|Y).\end{aligned}$$

Heegard showed in [6] that $H(X|S) - H(X|Y) = (1-\beta)[H(\alpha * \epsilon) - H(\alpha)]$. Therefore the proof of claim 1 in Theorem 3 is completed.

VI. Discussion

We presented a capacity-achieving coding scheme for noisy write-once memories. The model of noisy WOM in this paper differs from the model in [9] by the fact that here cells with state $s = 1$ are not prone to errors. For the error model in [9], we were not able to prove a channel degradation analogous to Lemma 6 and hence the results obtained in [9] are sub-optimal. A recent chaining method described in [12] implies efficient capacity-achieving coding schemes without the degradation requirement. Thus, the chaining method could be useful for the error model of [9]. However, the use of the chaining method comes at the price of a lower code rate in the finite blocklength regime.

References

[1] D. Burshtein and A. Strugatski, "Polar write once memory codes," *IEEE Trans. Inf. Theor.*, vol. 59, no. 8, pp. 5088–5101, Aug. 2013.

[2] E. En Gad, Y. Li, J. Kliewer, M. Langberg, A. Jiang, and J. Bruck, "Polar coding for noisy write-once memories," *Manuscript available at http://paradise.caltech.edu/papers/etr124.pdf*, 2014.

[3] A. Gabizon and R. Shaltiel, "Invertible zero-error dispersers and defective memory with stuck-at errors," in *APPROX-RANDOM*, 2012, pp. 553–564.

[4] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems of Control Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[5] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for broadcast channels," in *Proc. IEEE Int. Symp. on Inf. Theor. (ISIT),*, July 2013, pp. 1127–1131.

[6] C. Heegard, "On the capacity of permanent memory," *IEEE Trans. Inf. Theor.*, vol. 31, no. 1, pp. 34–42, January 1985.

[7] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Trans. Inf. Theor.*, vol. 59, no. 12, pp. 7829–7838, 2013.

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

[9] A. Jiang, Y. Li, E. En Gad, M. Langberg, and J. Bruck, "Joint rewriting and error correction in write-once memories," in *Proc. IEEE Int. Symp. Inf. Theor. (ISIT)*, 2013, pp. 1067–1071.

[10] S. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inf. Theor.*, vol. 56, no. 4, pp. 1751–1768, April 2010.

[11] A. Kusnetsov and B. S. Tsybakov, "Coding in a memory with defective cells,," *translated from Problemy Peredachi Informatsii*, vol. 10, no. 2, pp. 52–60, 1974.

[12] M. Mondelli, S. H. Hassani, I. Sason, and R. Urbanke, "Achieving Marton's region for broadcast channels using polar codes," *Available at http://arxiv.org/abs/1401.6060*, January 2014.

[13] R. Rivest and A. Shamir, "How to reuse a write-once memory," *Information and Control*, vol. 55, no. 1-3, pp. 1–19, 1982.

[14] A. Shpilka, "Capacity achieving multiwrite WOM codes," *IEEE Trans. Inf. Theor.*, vol. PP, no. 99, pp. 1–1, 2013.

[15] E. Yaakobi, P. Siegel, A. Vardy, and J. Wolf, "Multiple error-correcting wom-codes," *IEEE Trans. Inf. Theor.*, vol. 58, no. 4, pp. 2220–2230, April 2012.

[16] G. Zemor and G. D. Cohen, "Error-correcting wom-codes," *IEEE Trans. Inf. Theor.*, vol. 37, no. 3, pp. 730–734, May 1991.