Write-Once-Memory Codes by Source Polarization

Xudong Ma
Email: xma@ieee.org

Abstract—We propose a new Write-Once-Memory (WOM) coding scheme based on source polarization. We prove that the proposed WOM codes are capacity-achieving. The WOM codes were first discussed in 1980s. However, constructing capacity-achieving WOM codes had remained an unsolved problem for more than three decades. Surprisingly, two constructions of capacity-achieving WOM codes were proposed recently [2] [3]. The coding scheme in [3] is based on algebraic approaches. The coding scheme in [2] is based on polar codes (channel polarization).

The current paper proposes a third approach of constructing capacity-achieving WOM codes by using source polarization. The source polarization was invented by Arikan for lossless source coding [4]. We show in this paper that source polarization may be used for constructing WOM codes as well. We prove that the thus constructed WOM codes are capacity-achieving. The WOM codes proposed in this paper have the advantages of low-complexity software and hardware implementations.

The paper is organized as follows. In Section III we will discuss the considered probability models for the WOM encoding problem. We will prove two lemmas, which are the consequences of Theorem 1 in [4]. In Section III we will present our invented WOM coding scheme. In Section IV we will show the invented coding scheme is capacity-achieving. Finally, some concluding remarks will be presented in Section V.

We use the following notation throughout this paper. We use $X_n$ to denote a sequence of symbols $X_n, X_{n+1}, \ldots, X_{N-1}, X_N$. We use upper-case letters to denote random variables and lower-case letters to denote the corresponding realizations. For example, $X_i$ is one random variable, and $x_i$ is one realization of the random variable $X_i$. For a binary random variable $X$ with probabilities $P(X = 0) = p$, $P(X = 1) = 1 - p$, we define the entropy function $H(p)$ as follows,

$$H(p) = p \log \left( \frac{1}{p} \right) + (1 - p) \log \left( \frac{1}{1 - p} \right) \quad (1)$$

II. PROBABILITY MODELS AND AUXILIARY LEMMAS

In this section, we will consider the following probability model of random variables $Y_1^N, X_1^N$ and $U_1^N$. The physical meanings of $Y_1^N, X_1^N$ are as follows. The random variables $Y_1^N$ are the states of the memory cells.
The random variables $X_i^N$ are the codeword bits. In this paper, we only consider binary cases, i.e., $Y_i^N, X_i^N$ are all binary random variables. However, it should be clear that the coding scheme and analysis can be easily generalized to the non-binary cases.

We assume that $Y_n$ are independent and identically distributed (iid) with the following probability distribution $P(\cdot)$,

$$P(Y_n) = \begin{cases} s, & \text{if } Y_n = 0 \\ 1-s, & \text{if } Y_n = 1 \end{cases}$$

where, $s$ is a real number, $0 < s < 1$. Let each $X_i$ be conditionally independent of the other $X_j$ given $Y_i$.

$$P(X_n|Y_n) = \begin{cases} 1, & \text{if } X_n = 0, Y_n = 0 \\ 0, & \text{if } X_n = 1, Y_n = 0 \\ t, & \text{if } X_n = 0, Y_n = 1 \\ 1-t, & \text{if } X_n = 1, Y_n = 1 \end{cases}$$

where, $t$ is a real number, $0 < t < 1$. Let us define a matrix $G_N$ in the same way as in [4].

$$G_N = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \otimes n B_N$$

where, $\otimes n$ denotes the $n$th Kronecker power and $B_N$ is the bit-reversal permutation. According to the definition, $G_N$ is an $N$ by $N$ square matrix. Define binary random variables $U_1, U_2, \ldots, U_N$ as follows.

$$U_i^N = X_i^N G_N$$

where, $X_i^N$ denotes the row vector $[X_1, X_2, \ldots, X_N]$ and $U_i^N$ denotes the row vector $[U_1, U_2, \ldots, U_N]$. Clearly, $U_i^N$ is uniquely determined by $X_i^N$.

The configuration of the random variables $Y_i^N, X_i^N$ and $U_i^N$ is a special case of the scenarios in [4], where $Y_i^N$ is the side information, $X_i^N$ is the to be compressed bit string. The following lemmas are corollaries of Theorem 1 in [4].

**Lemma 2.1**: There exists a sequence of sets $F_N$, where, $N$ are integers and go to infinity, and each $F_N$ is a subset of $\{1, 2, \ldots, N\}$, such that the following hold.

- The cardinality $|F_N|$ of the set $F_N$ satisfies

$$\frac{|F_N|}{N} \geq (1-\epsilon_N)(1-s)H(t)$$

- For each $i \in F_N$, $H(U_i|Y_i^N, U_{i-1}^N) \geq 1 - \delta_N$
- $\epsilon_N > 0, \epsilon_N \to 0$ as $N \to \infty$, and
- $\delta_N > 0, \delta_N \to 0$ as $N \to \infty$.

**Lemma 2.2**: There exists a sequence of sets $F_N$, where, $N$ are integers and go to infinity, and each $F_N$ is a subset of $\{1, 2, \ldots, N\}$, such that the following hold.

- The cardinality $|F_N|$ of the set $F_N$ satisfies

$$\frac{|F_N|}{N} \geq (1-\epsilon_N)(1-s)H(t)$$

- For each $i \in F_N$,

$$\sum_{y_i^N, u_i^N} P(y_i^N, u_i^{i-1}) \left| P(U_i = 0|y_i^N, u_i^{i-1}) - \frac{1}{2} \right| \leq \delta_N$$

where, $\epsilon_N > 0, \epsilon_N \to 0$ as $N \to \infty$, and
- $\delta_N > 0, \delta_N \to 0$ as $N \to \infty$.

**Proof**: Prove by contradiction. Suppose not. Let $F_N$ denote a sequence of sets, such that, the cardinality $|F_N|$ of the set $F_N$ satisfies

$$\frac{|F_N|}{N} \geq (1-\epsilon_N)H(X_i|Y_i) = (1-\epsilon_N)(1-s)H(t)$$

where, $\epsilon_N > 0, \epsilon_N \to 0$ as $N \to \infty$. Suppose that for each such sequence $F_N$ and for each $N$, there exist at least one $i$, such that

$$\sum_{y_i^N, u_i^{i-1}} P(y_i^N, u_i^{i-1}) \left| P(U_i = 0|y_i^N, u_i^{i-1}) - \frac{1}{2} \right| = a_N, \quad y_i^N, u_i^{i-1}$$

and $a_N > 0, a_N$ is bounded away from 0. That is, there exists an $a > 0$, such that $a_N \geq a$ for all $N$.

Let us define one binary variable $b(y_i^N, u_i^{i-1})$ for each configuration $y_i^N, u_i^{i-1}$, such that

$$b(y_i^N, u_i^{i-1}) = \begin{cases} 0, & \text{if } P(U_i = 0|y_i^N, u_i^{i-1}) < 1/2 \\ 1, & \text{otherwise} \end{cases}$$

It can be easily checked that

$$\sum_{y_i^N, u_i^{i-1}} P(y_i^N, u_i^{i-1}) \left| P(U_i = 0|y_i^N, u_i^{i-1}) - \frac{1}{2} \right| = a_N$$

Therefore,

$$\sum_{y_i^N, u_i^{i-1}} P(y_i^N, u_i^{i-1}) \sum_{y_i^N, u_i^{i-1}} P(U_i = b(y_i^N, u_i^{i-1})|y_i^N, u_i^{i-1})$$

$$= \frac{1}{2} - a_N$$
Then, we have the bound in Eq. (18) where (a) follows from the fact that the entropy function $H(i) = H(1-t)$, (b) follows from the Jensen’s inequality because the entropy function $H(\cdot)$ is concave, and (c) follows from the fact that the entropy function $H(\cdot)$ is increasing in the interval $(0, 1/2)$. From the above inequality, we have $H(U_i|Y_i^N, U_i^{-1})$ is bounded away from 1. This statement contradicts Lemma 2.1. Hence, the supposition is false and the Lemma is true.

III. THE INVENTED CODING ALGORITHM

In this section, we will first present the encoding algorithm of the invented WOM codes. The encoding algorithm is a randomized algorithm. From Lemma 2.2, we can see that there exists a set $F_N$, such that for each $i \in F_N$, $P(U_i|y_i^N, u_i^{-1})$ is almost an uniform distribution. We call the set $F_N$ as the high-entropy set. Our invented coding algorithm uses the bits $U_i$ with $i \in F_N$ to record the to-be-recorded information. Let the cardinality $|F_N|$ of the set $F_N$ be $M$, then the number of the to-be-recorded bits is $M$. Let us denote the to-be-recorded bits by $v_1, v_2, \ldots, v_M$. The encoding algorithm is shown in Algorithm 1.

The encoding algorithm takes the inputs including the current memory states of the memory cells $y_i^N$, the high-entropy set $F_N$, the to-be-recorded message $v_i^M$, the parameter $s$ and $t$. The algorithm then determines $u_n$ one by one form $n = 1$ to $n = N$. If $n \in F_N$, then $u_n$ is set to one of the to-be-recorded information bits $v_n$. If $n \notin F_N$, then $u_n$ is randomly set to 1 with probability $P(U_n = 1|y_i^N, u_i^{-1})$, and $u_n$ is randomly set to 0 with probability $P(U_n = 0|y_i^N, u_i^{-1})$. After all the $u_n$ for $n = 1, \ldots, N$ have been determined, a vector $x_1^N$ is calculated as $x_1^N = u_1^N (G_N)^{-1}$. In other words, $u_1^N = x_1^NG_N$. The algorithm finally outputs $x_1^N$ as the codeword.

From the codeword $x_1^N$, the recorded information bits $v_i^M$ can be easily recovered as follows. Note that the auxiliary variables $u_i^N = x_i^NG_N$. And the recorded information bits $v_1, v_2, \ldots, v_M$ are the bits of $u_i^N$ at the positions in $F_N$. Thus, the proposed WOM codes have a rather simple decoding algorithm.

IV. CAPACITY ACHIEVING PROOF

In this section, we will prove that the proposed WOM codes are capacity achieving. We have defined an artificial probability distribution $P$ in Section III

$$P(Y_i) = \begin{cases} s, & \text{if } Y_i = 0 \\ 1-s, & \text{if } Y_i = 1 \end{cases} \text{ (19)}$$

Algorithm 1: Write-Once-Memory Codes by Source Polarization

1: The algorithm takes inputs

- the current memory states of the memory cells $y_i^N$
- the high-entropy set $F_N$
- the to-be-recorded message $v_i^M$
- parameter $s$
- parameter $t$

2: $n \leftarrow 1, m \leftarrow 1$

3: repeat

4: if $n \in F_N$ then

5: $u_n \leftarrow v_m$

6: $n \leftarrow n + 1$

7: $m \leftarrow m + 1$

8: else

9: Calculate $P(U_n|y_i^N, u_i^{-1})$

10: Randomly set $u_n$ according to the probability distribution $P(U_n|y_i^N, u_i^{-1})$. That is

$$u_n = \begin{cases} 0, & \text{with probability } P(U_n = 0|y_i^N, u_i^{-1}) \\ 1, & \text{with probability } P(U_n = 1|y_i^N, u_i^{-1}) \end{cases}$$

11: $n \leftarrow n + 1$

12: end if

13: until $n > N$

14: $x_1^N \leftarrow u_1^N (G_N)^{-1}$

15: The algorithm outputs $x_1^N$ as the WOM codeword

$$P(X_i|Y_i) = \begin{cases} 1, & \text{if } X_i = 0, Y_i = 0 \\ 0, & \text{if } X_i = 1, Y_i = 0 \\ t, & \text{if } X_i = 0, Y_i = 1 \\ 1-t, & \text{if } X_i = 1, Y_i = 1 \end{cases} \text{ (20)}$$

On the other hand, the randomized encoding algorithm in Section III may be considered as a random process. The random process induces a probability distribution $Q$. The probability distribution $Q$ is

$$Q(Y_n) = \begin{cases} s, & \text{if } Y_n = 0 \\ 1-s, & \text{if } Y_n = 1 \end{cases} \text{ (22)}$$

$$Q(U_i|Y_i^N, U_i^{-1}) = \begin{cases} 1/2, & \text{if } i \in F_N \\ P(U_i|Y_i^N, U_i^{-1}) & \text{otherwise} \end{cases} \text{ (23)}$$

$$Q(X_i^N|Y_i^N, U_i^N) = \begin{cases} 1, & \text{if } U_i^N = X_i^NG_N \\ 0, & \text{otherwise} \end{cases} \text{ (24)}$$

For example, $Q(U_n = 1)$ is the probability that $u_n$ is set to 1 by using the randomized encoding algorithm. Note
that we have assumed that the to-be-record bits \( v_m \) are random and equally probable.

We need the following telescoping expansion in our latter discussions.

**Lemma 4.1:**

\[
\prod_{n=1}^{N} A_n - \prod_{n=1}^{N} B_n = \sum_{i=1}^{N} \left( \prod_{n=1}^{i-1} A_n \right) (A_i - B_i) \left( \prod_{n=i+1}^{N} B_n \right)
\]

(25)

One important step of the capacity-achieving proof is the following bound on the total variation distance between the probability distributions \( P \) and \( Q \).

**Lemma 4.2:** The total variation distance between \( P \) and \( Q \) can be bounded as follows.

\[
\sum_{y_1^N,u_1^N} \left| P(y_1^N,u_1^N) - Q(y_1^N,u_1^N) \right| \leq 2N \delta_N
\]

(27)

where, \( \delta_N \to 0 \), as \( N \to \infty \).

**Proof:** We have the bound in Eq. 28 where

- (a) follows from the lemma 4.1,
- (b) follows from the definition of the probability distribution \( Q \);
- (c) follows from the fact the absolute value of a sum is always less than or equal to the sum of absolute values;
- (d) is resulting from a change of the order of summation;
- (e) is resulting from a change of the order of summation;
- (f) follows from the fact that the summation of probability \( Q \) is 1;
- (g) follows from the multiplication rule of probabilities and conditional probabilities; and
- (h) follows from Lemma 2.2.

The lemma is thus proven.

**Theorem 4.3:** For the random encoding algorithm \( \Pi \) assume the number of memory cells is \( N \). Then, the expected ratio of the number of memory cells turning from state 1 to state 0 to \( N \) goes to \((1 - s)H(t)\), as \( N \) goes to infinity. The expected ratio of recorded information bits to \( N \) goes to \((1 - s)H(t)\), as \( N \) goes to infinity. Thus, the proposed WOM codes are capacity-achieving.

**Proof:** Let us define the following indicator function \( I(x_n,y_n) \).

\[
I(x_n,y_n) = \begin{cases} 
1, & \text{if } x_n = 0, y_n = 1 \\
0, & \text{otherwise}
\end{cases}
\]

(29)

The expected number of memory cells turning from state 1 to state 0 is thus

\[
\mathbb{E}_Q \left[ \sum_{n=1}^{N} I(x_n,y_n) \right] = \sum_{y_1^N,u_1^N,x_1^N} Q(y_1^N,u_1^N,x_1^N) \sum_{n=1}^{N} I(x_n,y_n)
\]

(30)

On the other hand, we have the bound in Eq. 36 where (a) follows from that fact that the absolute value of a sum is always less than or equal to the sum of absolute values; (b) follows from the fact that the indicator function \( I(x_n,y_n) \leq 1 \); (c) follows from the fact that \( x_1^N \) is deterministic given \( y_1^N \) and \( u_1^N \); and (d) follows from Lemma 4.2. As a consequence, as \( N \) goes to infinity,

\[
\frac{1}{N} \mathbb{E}_Q \left[ \sum_{n=1}^{N} I(x_n,y_n) \right] \to \frac{1}{N} \mathbb{E}_P \left[ \sum_{n=1}^{N} I(x_n,y_n) \right]
\]

(31)

On the other hand,

\[
\mathbb{E}_P \left[ \sum_{n=1}^{N} I(x_n,y_n) \right]
\]

(32)

\[
= \sum_{y_1^N,u_1^N,x_1^N} P(y_1^N,u_1^N,x_1^N) \sum_{n=1}^{N} I(x_n,y_n)
\]

(33)

\[
= \sum_{n=1}^{N} P(x_n = 0, y_n = 1)
\]

(34)

\[
= N(1 - s) t
\]

(35)

Therefore, the expected number of memory cells turning from state 1 to state 0 is \((1 - s) t N\) asymptotically. The
\[\sum_{y_1^N, u_1^N} |P(y_1^N, u_1^N) - Q(y_1^N, u_1^N)|\]
\[= \sum_{y_1^N, u_1^N} P(y_1^N) \left| \prod_{n=1}^{N} P(u_n| y_1^N, u_1^{n-1}) - \prod_{n=1}^{N} Q(u_n| y_1^N, u_1^{n-1}) \right|\]
\[\approx \sum_{y_1^N, u_1^N} P(y_1^N) \left[ \sum_{i=1}^{N} \left( \prod_{n=1}^{i-1} P(u_n| y_1^N, u_1^{n-1}) \right) \left( P(u_i| y_1^N, u_1^{i-1}) - Q(u_i| y_1^N, u_1^{i-1}) \right) \left( \prod_{n=i+1}^{N} Q(u_n| y_1^N, u_1^{n-1}) \right) \right]\]
\[\approx \sum_{y_1^N, u_1^N} P(y_1^N) \left[ \sum_{i \in F} \left( \prod_{n=1}^{i-1} P(u_n| y_1^N, u_1^{n-1}) \right) \left( P(u_i| y_1^N, u_1^{i-1}) - 1/2 \right) \left| \prod_{n=i+1}^{N} Q(u_n| y_1^N, u_1^{n-1}) \right| \right]\]
\[\leq \sum_{y_1^N, u_1^N} P(y_1^N) \left[ \sum_{i \in F} \left( \prod_{n=1}^{i-1} P(u_n| y_1^N, u_1^{n-1}) \right) \left( P(u_i| y_1^N, u_1^{i-1}) - 1/2 \right) \left| \prod_{n=i+1}^{N} Q(u_n| y_1^N, u_1^{n-1}) \right| \right]\]
\[\leq 2N\delta_N\] (28)

The number of recorded bits is the cardinality of the high-entropy set \(F_N\), which is \((1 - s)H(t)N\) asymptotically. It can be checked that this is exactly the capacity of 2-write binary WOM codes [1] [3]. It is not difficult to verify that the capacity region of \(t\)-write binary WOM codes can also be achieved by using the proposed coding scheme \(t\) times, for any integer \(t\). Thus the proposed coding scheme is capacity-achieving.

V. Conclusion

The paper presents one WOM coding scheme based on source polarization. We prove that the proposed coding scheme is capacity-achieving.

References

\[
\frac{1}{N} \left| \mathbb{E}_Q \left[ \sum_{n=1}^{N} I(x_n, y_n) \right] - \mathbb{E}_P \left[ \sum_{n=1}^{N} I(x_n, y_n) \right] \right| = \frac{1}{N} \left| \sum_{y_1^N, u_1^N, x_1^N} (Q(y_1^N, u_1^N, x_1^N) - P(y_1^N, u_1^N, x_1^N)) I(x_n, y_n) \right|
\]

\[
\leq \frac{1}{N} \sum_{y_1^N, u_1^N, x_1^N} |Q(y_1^N, u_1^N, x_1^N) - P(y_1^N, u_1^N, x_1^N)| I(x_n, y_n)
\]

\[
\leq \frac{1}{N} \sum_{y_1^N, u_1^N, x_1^N} |Q(y_1^N, u_1^N, x_1^N) - P(y_1^N, u_1^N, x_1^N)| \leq \frac{1}{N} \sum_{y_1^N, u_1^N} |Q(y_1^N, u_1^N) - P(y_1^N, u_1^N)| \leq 2\delta_N
\]

(36)