[9] G. Z. Xiao and J. L. Massey, "A spectral characterization of correlation-immune combining functions," *IEEE Trans. Inform. Theory*, vol. 34, pp. 569–571, May 1988.
[10] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice Hall, 1975.
[11] M. Bellanger, *Digital Processing of Signals.* New York: Wiley, 1984.

On the Capacity of Generalized Write-Once Memory with State Transitions Described by an Arbitrary Directed Acyclic Graph

Fang-Wei Fu and A. J. Han Vinck, *Senior Member, IEEE*

***Abstract*—The generalized write-once memory introduced by Fiat and Shamir is a q-ary information storage medium. Each storage cell is expected to store one of q symbols, and the legal state transitions are described by an arbitrary directed acyclic graph. This memory model can be understood as a generalization of the binary write-once memory which was introduced by Rivest and Shamir. During the process of updating information, the contents of a cell can be changed from a 0-state to a 1-state but not *vice versa*. We study the problem of reusing a generalized write-once memory for T successive cycles (generations). We determine the zero-error capacity region and the maximum total number of information bits stored in the memory for T consecutive cycles for the situation where the encoder knows and the decoder does not know the previous state of the memory. These results extend the results of Wolf, Wyner, Ziv, and Körner for the binary write-once memory.**

***Index Terms*—Capacity, directed acyclic graph, information, WOM-codes, write-once memory.**

I. Introduction

A write-once memory (WOM) is a binary information storage medium. During the process of updating information, the contents of a cell can be changed from a 0-state to a 1-state but not *vice versa*. This class of WOM includes punch cards and digital optical discs in which binary data is represented by blanks (0's) and dots (1's). Due to the updating technology, the dots cannot be removed. Rivest and Shamir [1] showed that, if the encoder knows and the decoder does not know the previous state of the memory, WOM can be reused very efficiently by using the same code for every updating cycle. Wolf, Wyner, Ziv, and Körner [2] studied the WOM from an information-theoretical point of view. They determined the capacity region and the maximum total number of information bits stored in the memory for fixed T successive cycles by using arbitrary codes for every cycle. Cohen, Godlewski, and Merkx [3] presented a class of linear coset codes for the WOM. Cohen and Zëmor [4] presented a construction method for the error-correcting WOM codes.

Fiat and Shamir [5] studied the generalized write-once memory, which is a q-ary information storage medium. Each storage cell is expected to store one of q symbols, and the legal state transitions are described by an arbitrary directed acyclic graph. They extended the results of Rivest and Shamir [1] for the binary WOM to the generalized WOM, in the case when the encoder and decoder use the same code for every cycle. Heegard [6] investigated the noisy WOM and presented an inner bound for the ε-error capacity region. He showed that in some cases the inner bound is exactly the ε-error capacity region. Kuznetsov and Vinck [7] studied the general defective channel with informed encoder as a generalization of a memory with defects. They presented lower and upper bounds for the maximum transmission rate. As a corollary, they derived the capacities of the binary WOM and other constrained memories.

In this correspondence, we study the problem of reusing a generalized WOM for T successive cycles. We determine the zero-error capacity region and the maximum total number of information bits stored in the memory for T successive cycles for the situation where the encoder knows and the decoder does not know the previous state of the memory, and different codes are allowed to be used in every cycle. These results extend the results of Wolf, Wyner, Ziv, and Körner for the binary WOM to the generalized WOM.

Manuscript received July 21, 1997; revised June 10, 1998. This work was supported in part by the Chinese Education Ministry, the National Natural Science Foundation of China, and the University of Essen, and was done during October 1996–October 1997, while F.-W. Fu was visiting the Institute for Experimental Mathematics, University of Essen, 45326 Essen, Germany.

F.-W. Fu is with the Department of Mathematics, Nankai University, Tianjin 300071, China.

A. J. Han Vinck is with the Institute for Experimental Mathematics, University of Essen, 45326 Essen, Germany.

Communicated by K. Zeger, Editor at Large.

Publisher Item Identifier S 0018-9448(99)00084-X.

II. Definitions, Notations, and Model

In this section, we first give the mathematical model of the generalized WOM with notations as in [5]. Then we give some examples of the generalized WOM and conclude with the definitions of WOM codes, the capacity region, and the maximum total number of information bits stored in the memory for T successive cycles. Below we first introduce some concepts and notations of a directed graph.

A directed graph is a pair $(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertices and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. A directed edge from s to s' is denoted by $s \to s'$. A path from s to s' is a sequence of zero or more edges in $\mathcal{E}$ of the form $s = s_1 \to s_2 \to \cdots \to s_k = s'$. The notation $s \Longrightarrow s'$ is used to represent the fact that there exists a path from s to s'. A cycle is a nonempty path from a vertex to itself. A directed graph which does not contain cycles is a directed acyclic graph, abbreviated as DAG. A rooted DAG is a triple $(\mathcal{V}, \mathcal{E}, r)$ such that $(\mathcal{V}, \mathcal{E})$ is a DAG, the root $r \in \mathcal{V}$, and for any $s \in \mathcal{V}$ there is a path from r to s. In the sequel, we only consider rooted DAG's. We assume $\mathcal{V} = \{0, 1, \cdots, q-1\}\%$; and 0 is the root.

A generalized WOM is a q-ary information storage medium. Each storage cell is expected to store one of q symbols, and the legal state transitions are described by a rooted directed acyclic graph $(\mathcal{V}, \mathcal{E})$, abbreviated as $(\mathcal{V}, \mathcal{E})$-WOM. During the process of updating information we can update s to s' if and only if $s \Longrightarrow s'$. This memory model can be seen as a generalization of the binary write-once memory.

Remark: The writing (updating) constraints here are described by the paths (not edges) of a rooted DAG $(\mathcal{V}, \mathcal{E})$. We adopt this point of view from the original paper by Fiat and Shamir [5].

Example 1: $\mathcal{V} = \{0, 1\}$, $\mathcal{E} = \{0 \to 1\}$, the $(\mathcal{V}, \mathcal{E})$-WOM is a binary WOM. A storage cell in a "0" state may be left unchanged, or updated to the "1" state. A storage cell in a "1" state is then forever stuck at the "1" state.

Example 2: $\mathcal{V} = \{0, 1, 2, 3\}$, $\mathcal{E} = \{0 \to 1, 0 \to 2, 0 \to 3\}$, during the updating process, a storage cell in a "0" state may be left unchanged or updated to the "1" or "2" or "3" state. A storage cell in an "i" $(i = 1, 2, 3)$ state is then forever stuck at the "i" state.

Example 3: $\mathcal{V} = \{0, 1, 2, 3\}$, $\mathcal{E} = \{0 \to 1, 1 \to 2, 2 \to 3\}$, during the updating process, a storage cell in a "0" state may be left unchanged or updated to the "1" or "2" or "3" state. A storage cell in a "1" state may be left unchanged, or updated to the "2" or "3" state. A storage cell in a "2" state may be left unchanged or updated to the "3" state. A storage cell in a "3" state is then forever stuck at the "3" state.

Assume that the $(\mathcal{V}, \mathcal{E})$-WOM consists of n storage cells. The initial state of all the storage cells is "0" (root). We want to reuse the $(\mathcal{V}, \mathcal{E})$-WOM for T successive cycles. We only consider the following case: The encoder knows and the decoder does not know the previous state of the memory. The encoder and decoder can use arbitrary codes for every cycle, and there are no decoding errors (zero-error case).

Notation: For the vectors

$$x^n = (x_1, x_2, \cdots, x_n) \in \mathcal{V}^n$$
$$y^n = (y_1, y_2, \cdots, y_n) \in \mathcal{V}^n$$

we denote $x^n \Longrightarrow y^n$ if and only if $x_i \Longrightarrow y_i$, $i = 1, 2, \cdots, n$.

Definition 1: An $(n, T, M_1, \cdots, M_T)$ code for the $(\mathcal{V}, \mathcal{E})$-WOM consists of T pairs of encoding and decoding functions $\{(f_t, g_t)\}_{t=1}^{T}$, where the message index sets $I_t = \{1, \cdots, M_t\}$, the encoding functions $f_t: I_t \times \mathcal{V}^n \longmapsto \mathcal{V}^n$, and the decoding functions $g_t: \mathcal{V}^n \longmapsto I_t$. These encoding and decoding functions satisfy: For any $m_1 \in I_1$, $m_2 \in I_2, \cdots, m_T \in I_T$, denote $y_0^n = (0, \cdots, 0) = \underline{0} \in \mathcal{V}^n$ and $y_t^n = f_t(m_t, y_{t-1}^n), t = 1, \cdots, T$. Then, $y_{t-1}^n \Longrightarrow y_t^n$ and $g_t(y_t^n) = m_t$, $t = 1, \cdots, T$.

Denote

$$R_t = (1/n) \log_2 M_t, \qquad t = 1, \cdots, T.$$

The T-tuple $(R_1, \cdots, R_T)$ is called the rate vector of this code. The closure of the set of all rate-vectors $\mathcal{A}_T(\mathcal{V}, \mathcal{E})$ is called the capacity region of the $(\mathcal{V}, \mathcal{E})$-WOM. The maximum total number of information bits stored in one storage cell of the $(\mathcal{V}, \mathcal{E})$-WOM during the T updating cycles is

$$C_T(\mathcal{V}, \mathcal{E}) = \max \left\{ \sum_{t=1}^{T} R_t | (R_1, R_2, \cdots, R_T) \in \mathcal{A}_T(\mathcal{V}, \mathcal{E}) \right\}.$$

III. Main Results

We present an information-theoretic single-letter characterization for the capacity region $\mathcal{A}_T(\mathcal{V}, \mathcal{E})$, and a formula for the maximum total number of information bits $C_T(\mathcal{V}, \mathcal{E})$ stored in one storage cell of the $(\mathcal{V}, \mathcal{E})$-WOM during the T updating cycles.

Let $H(\cdot)$ represent the entropy of a random variable or a probability vector; $H(\cdot|\cdot)$ represent the conditional entropy; $\Re_+$ denote the set of nonnegative real numbers; and

$$\hat{\mathcal{V}} = \{(v_1, v_2) \in \mathcal{V}^2 | v_1 \Longrightarrow v_2\}.$$

For the two random variables X, Y which take values in $\mathcal{V}$, we denote $X \Longrightarrow Y$ if for any $(x, y) \notin \hat{\mathcal{V}}$, $\Pr\{X = x, Y = y\} = 0$. Let the random variables $S_1, S_2, \cdots, S_T$ form a Markov chain which takes values in $\mathcal{V}$. We denote $S_1 \Longrightarrow S_2 \Longrightarrow \cdots \Longrightarrow S_T$, if for every $t = 2, 3, \cdots, T$, $S_{t-1} \Longrightarrow S_t$. Assume $\mathcal{B}_T(\mathcal{V}, \mathcal{E}) = \{(R_1, R_2, \cdots, R_T) \in \Re_+^T |$ there exist random variables $S_1, S_2, \cdots, S_T$, $S_1 \Longrightarrow S_2 \Longrightarrow \cdots \Longrightarrow S_T$, such that $R_1 \le H(S_1)$, $R_2 \le H(S_2|S_1), \cdots, R_T \le H(S_T|S_{T-1})\}$.

The set $\mathcal{R}_T(\mathcal{V}, \mathcal{E})$ is the closed set generated by $\mathcal{B}_T(\mathcal{V}, \mathcal{E})$.

Theorem 3.1: The zero-error capacity region of the $(\mathcal{V}, \mathcal{E})$-WOM is $\mathcal{A}_T(\mathcal{V}, \mathcal{E}) = \mathcal{R}_T(\mathcal{V}, \mathcal{E})$.

The proof of Theorem 3.1 will be given in Section IV.

Remark: Heegard [6] determined the ε-error capacity region for the deterministic WOM, which has the same representation form as Theorem 3.1 (zero-error capacity region). In [6], Heegard first derived an inner bound for the ε-error capacity region of the noisy WOM. Then he showed that this inner bound is tight for the deterministic WOM. It should be pointed out that the proof of the converse part of [6, Theorem 3] is not complete.

For a given rooted DAG $(\mathcal{V}, \mathcal{E})$, we define its transition matrix as $\boldsymbol{A} = (a_{ij})_{q \times q}$, where $a_{ij} = 1$ if there is a (zero or nonzero) path from vertex i to vertex j, and $a_{ij} = 0$, otherwise. We denote $\mathbf{1_m}$ as the all-one row vector of length m, and I_m as the unit matrix of order m. If Q is a matrix, we use Q^c to denote its transpose matrix.

Remark: The transition matrix defined here is based on the path set (not based on the edge set $\mathcal{E}$). The matrix A described here is actually the incidence matrix of the "transitive closure" of the original DAG plus the identity matrix.

Theorem 3.2: The maximum total number of information bits stored in one storage cell of the $(\mathcal{V}, \mathcal{E})$-WOM during the T updating cycles is

$$C_T(\mathcal{V}, \mathcal{E}) = \log_2(\mathbf{1_q} \cdot \boldsymbol{A}^{T-1} \cdot \mathbf{1_q^c}).$$

The proof of Theorem 3.2 will be given in Section V.

Remark: For a binary WOM, the transition matrix is

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

and

$$(1, 1) \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{T-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = T + 1.$$

Hence, the maximum total number of information bits stored in one bit position of the binary WOM during the T updating cycles is $\log_2(T+1)$ (see Wolf, Wyner, Ziv, and Körner, [2]).

We consider the following two types of generalized WOM's.

1) $\mathcal{V} = \{0, 1, \cdots, q-1\}$, $\mathcal{E}_1 = \{0 \to 1, 0 \to 2, \cdots, 0 \to q-1\}$.
2) $\mathcal{V} = \{0, 1, \cdots, q-1\}$, $\mathcal{E}_2 = \{0 \to 1, 1 \to 2, \cdots, q-2 \to q-1\}$.

Corollary 3.1:

1) The capacity region of the $(\mathcal{V}, \mathcal{E}_1)$-WOM is $\mathcal{A}_T(\mathcal{V}, \mathcal{E}_1) = \{(R_1, R_2, \cdots, R_T) \in \Re_+^T |$ there exist probability vectors $\boldsymbol{p}^{(t)} = (p_0^{(t)}, p_1^{(t)}, \cdots, p_{q-1}^{(t)})$, $t = 1, 2, \cdots, T$, such that $R_1 \le H(\boldsymbol{p}^{(1)})$, $R_t \le (\Pi_{i=1}^{t-1} p_0^{(i)}) H(\boldsymbol{p}^{(t)})$, $t = 2, 3, \cdots, T\}$.
2) The maximum total number of information bits stored in one storage cell of the $(\mathcal{V}, \mathcal{E}_1)$-WOM during the T updating cycles is $C_T(\mathcal{V}, \mathcal{E}_1) = \log_2[1 + (q-1)T]$.

Proof:

1) Assume $S_1, S_2, \cdots, S_T$ are random variables which take values in $\mathcal{V}$, and $S_1 \Longrightarrow S_2 \Longrightarrow \cdots \Longrightarrow S_T$. Let

$$p_j^{(1)} = \Pr\{S_1 = j\}, \qquad j = 0, 1, \cdots, q-1$$
$$p_j^{(t)} = \Pr\{S_t = j | S_{t-1} = 0\},$$
$$j = 0, 1, \cdots, q-1, \;\; t = 2, 3, \cdots, T$$
$$\boldsymbol{p}^{(t)} = (p_0^{(t)}, p_1^{(t)}, \cdots, p_{q-1}^{(t)}), \qquad t = 1, 2, \cdots, T.$$

Then

$$H(S_1) = H(\boldsymbol{p}^{(1)});$$

$$H(S_t|S_{t-1}) = \sum_{j=0}^{q-1} \Pr\{S_{t-1} = j\} H(S_t|S_{t-1} = j).$$

Since

$$\Pr\{S_t = j|S_{t-1} = j\} = 1, \qquad j = 1, 2, \cdots, q-1,\ t = 2, 3, \cdots, T$$

we have

$$H(S_t|S_{t-1} = j) = 0, \qquad j = 1, 2, \cdots, q-1,\ t = 2, 3, \cdots, T.$$

This implies that

$$\begin{aligned} H(S_t|S_{t-1}) &= \Pr\{S_{t-1} = 0\} H(S_t|S_{t-1} = 0) \\ &= \Pr\{S_{t-1} = 0\} H(\boldsymbol{p}^{(t)}). \end{aligned}$$

From

$$\begin{aligned} \Pr\{S_{t-1} = 0\} &= \sum_{j=0}^{q-1} \Pr\{S_{t-2} = j\} \Pr\{S_{t-1} = 0|S_{t-2} = j\} \\ &= \Pr\{S_{t-2} = 0\} \Pr\{S_{t-1} = 0|S_{t-2} = 0\} \\ &= \Pr\{S_{t-2} = 0\} p_0^{(t-1)} = \cdots = \prod_{i=1}^{t-1} p_0^{(i)} \end{aligned}$$

it follows that

$$H(S_t|S_{t-1}) = \left(\prod_{i=1}^{t-1} p_0^{(i)}) H(\boldsymbol{p}^{(t)} \right).$$

The capacity region of the $(\mathcal{V}, \mathcal{E}_1)$-WOM is given by Theorem 3.1.

2) The transition matrix of the $(\mathcal{V}, \mathcal{E}_1)$-WOM is

$$\boldsymbol{A} = \begin{bmatrix} 1 & \boldsymbol{1_{q-1}} \\ 0 & I_{q-1} \end{bmatrix}.$$

Then

$$\boldsymbol{1_q} \cdot \boldsymbol{A}^{T-1} \cdot \boldsymbol{1_q^c} = 1 + (q-1)T.$$

The result follows from Theorem 3.2. □

Remark: If we take $q = 2$ in Corollary 3.1, then we get the results of [2] for binary WOM.

Corollary 3.2: The maximum total number of information bits stored in one storage cell of the $(\mathcal{V}, \mathcal{E}_2)$-WOM during the T updating cycles is

$$C_T(\mathcal{V}, \mathcal{E}_2) = \log_2 \boldsymbol{C_{q-1}^{T+q-1}} = \sum_{i=1}^{q-1} [\log_2(T+i) - \log_2 i]$$

where $\boldsymbol{C_m^n} = n!/m!(n-m)!$.

Proof: The transition matrix of the $(\mathcal{V}, \mathcal{E}_2)$-WOM is $\boldsymbol{U_q} = (u_{ij})_{q \times q}$, where $u_{ij} = 1$, if $j \geq i$; $u_{ij} = 0$, if $j < i$. Here

$$\boldsymbol{U_q} = \begin{bmatrix} \boldsymbol{U_{q-1}} & \boldsymbol{1_{q-1}^c} \\ \boldsymbol{0_{q-1}} & 1 \end{bmatrix}$$

where $\boldsymbol{0_{q-1}}$ is the zero vector of length $q-1$.

Therefore,

$$\boldsymbol{1_q} \cdot \boldsymbol{U_q^{t-1}} \cdot \boldsymbol{1_q^c} = 1 + \sum_{j=0}^{t-1} \boldsymbol{1_{q-1}} \cdot \boldsymbol{U_{q-1}^j} \cdot \boldsymbol{1_{q-1}^c}.$$

Denote $\beta_{t,q} = \boldsymbol{1_q} \cdot \boldsymbol{U_q^{t-1}} \cdot \boldsymbol{1_q^c}$, $t = 1, 2, \cdots$. Then

$$\beta_{t,q} = 1 + \sum_{j=1}^{t} \beta_{j,q-1}, \qquad t \geq 1, q \geq 2.$$

We show that $\beta_{t,q} = \boldsymbol{C_{q-1}^{t+q-1}}$ by induction on q. From Corollary 3.1 we know that $\beta_{t,2} = t+1, t \geq 1$. This implies that the assertion is true for $q = 2$. Assume the assertion is true for $q = k$, then

$$\beta_{t,k+1} = 1 + \sum_{j=1}^{t} \beta_{j,k} = 1 + \sum_{j=1}^{t} \boldsymbol{C_{k-1}^{j+k-1}} = \boldsymbol{C_k^{t+k}}.$$

This implies that the assertion is true for $q = k+1$. Therefore, by induction, we know the assertion is true. Hence

$$C_T(\mathcal{V}, \mathcal{E}_2) = \log_2 \boldsymbol{C_{q-1}^{T+q-1}}.$$ □

Remark: The capacity region $\mathcal{A}_T(\mathcal{V}, \mathcal{E}_2)$ is given by Theorem 3.1. Unlike Corollary 3.1, here the capacity region has no more explicit description.

IV. Proof of Theorem 3.1

In this section, we generalize the methods of Wolf, Wyner, Ziv, and Körner [2] for proving the coding theorems of binary WOM to prove Theorem 3.1. The proof of Theorem 3.1 is divided into two parts: 1) proof of the direct part of Theorem 3.1, and 2) proof of the converse part of Theorem 3.1. In order to prove Theorem 3.1, we introduce some properties of typical sequences.

A. Properties of Typical Sequences

In this subsection, we only list those properties of typical sequences which we need for establishing our results. For details of typical sequences and the proofs of these properties, we refer to the book written by Csiszár and Körner [8]. If A is a finite set, $|A|$ denotes the cardinality of A in this correspondence.

Let $\mathcal{X}, \mathcal{Y}$ be two finite sets, and let $\mathcal{P}(\mathcal{X})$ be the set of probability distributions on $\mathcal{X}$. For $x^n \in \mathcal{X}^n$, let $N(a|x^n)$ be the number of occurrences of $a \in \mathcal{X}$ in x^n. The type of x^n is the distribution P_{x^n} on $\mathcal{X}$, defined by

$$P_{x^n}(a) = \frac{1}{n} N(a|x^n), \qquad a \in \mathcal{X}.$$

Let $\mathcal{T}^n(\mathcal{X})$ be the set of types. For $P \in \mathcal{T}^n(\mathcal{X})$, we denote

$$T_P^n = \{x^n \in \mathcal{X}^n | P_{x^n} = P\}.$$

Assume X is a random variable, which takes values in $\mathcal{X}$. If its probability distribution $P_X \in \mathcal{T}^n(\mathcal{X})$, we denote $T_{P_X}^n = T_X^n$. For a pair of sequences

$$x^n = (x_1, \cdots, x_n) \in \mathcal{X}^n$$

and

$$y^n = (y_1, \cdots, y_n) \in \mathcal{Y}^n$$

let $N(a, b|x^n, y^n)$ be the number of occurrences of $(a, b) \in \mathcal{X} \times \mathcal{Y}$ in $\{(x_i, y_i)\}_{i=1}^n$. The joint type of x^n and y^n is the distribution P_{x^n, y^n} on $\mathcal{X} \times \mathcal{Y}$, defined by

$$P_{x^n, y^n}(a, b) = \frac{1}{n} N(a, b|x^n, y^n), \qquad a \in \mathcal{X}, b \in \mathcal{Y}.$$

In the same way, we can define $\mathcal{T}^n(\mathcal{X} \times \mathcal{Y})$ and T_P^n, $P \in \mathcal{T}^n(\mathcal{X} \times \mathcal{Y})$. Suppose X, Y are two random variables, which take values in $\mathcal{X}$ and $\mathcal{Y}$, respectively. If the joint probability distribution $P_{X,Y} \in \mathcal{T}^n(\mathcal{X} \times \mathcal{Y})$, then the marginal distribution $P_X \in \mathcal{T}^n(\mathcal{X})$ and $P_Y \in \mathcal{T}^n(\mathcal{Y})$. We denote $T_{P_{X,Y}}^n = T_{X,Y}^n$. For $x^n \in T_X^n$, we denote

$$T_{Y|X}^n(x^n) = \{y^n \in \mathcal{Y}^n | P_{x^n, y^n} = P_{X,Y}\}.$$

Property 4.1: For any $P \in \mathcal{P}(\mathcal{X})$, there exists a sequence of types $P_n \in \mathcal{T}^n(\mathcal{X})$, such that

$$\max_{x \in \mathcal{X}} |P_n(x) - P(x)| \to 0, \qquad n \to \infty.$$

Property 4.2:

$$|\mathcal{T}^n(\mathcal{X})| \le (n+1)^{|\mathcal{X}|}.$$

Property 4.3:

$$(n+1)^{-|\mathcal{X}|} 2^{nH(X)} \le |T_X^n| \le 2^{nH(X)}.$$

Property 4.4: For any $x^n \in T_X^n$, we have

$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|} 2^{nH(Y|X)} \le |T_{Y|X}^n(x^n)| \le 2^{nH(Y|X)}.$$

B. Proof of the Direct Part of Theorem 3.1

The Direct Part of Theorem 3.1: Assume $S_1, S_2, \cdots, S_T$ are random variables, which take values in $\mathcal{V}$, and $S_1 \Longrightarrow S_2 \Longrightarrow \cdots \Longrightarrow S_T$. For any fixed $\delta > 0$ (very small), set

$$M_1 = 2^{n[H(S_1) - \delta]}$$
$$M_t = 2^{n[H(S_t|S_{t-1}) - \delta]}, \qquad t = 2, 3, \cdots, T.$$

For n sufficiently large, there exists an $(n, T, M_1, \cdots, M_T)$ code for the $(\mathcal{V}, \mathcal{E})$-WOM.

Remark: $M_1, \cdots, M_T$ are assumed to be positive integers. This assumption does not affect the asymptotic behavior of the code.

Proof: For simplification, we only prove the direct part of Theorem 3.1 for $T = 2$. It is easy to extend our proof for general T. Note that we can define the encoding and decoding function for the tth cycle in the same way as for the second cycle.

By Property 4.1, we can assume $P_{S_1,S_2} \in \mathcal{T}^n(\mathcal{V} \times \mathcal{V})$. This assumption does not affect the asymptotic behavior of the following code construction.

Suppose $\{F_m\}_{m=1}^{M_2}$ is a partition of $T_{S_2}^n$, i.e., $F_m \bigcap F_{m'} = \emptyset$, $m \ne m'$; and $\bigcup_{m=1}^{M_2} F_m = T_{S_2}^n$. There exists an (n, T, M_1, M_2) code for the $(\mathcal{V}, \mathcal{E})$-WOM, if this partition satisfies the following property.

Property ♣: For any $u \in T_{S_1}^n$ and any $m_2 \in I_2 = \{1, \cdots, M_2\}$, there exists a vector $x \in F_{m_2}$, such that $u \Longrightarrow x$.

The encoding and decoding functions can be defined as follows.

The first cycle: By Property 4.3, we can choose M_1 different elements $v_1, v_2, \cdots, v_{M_1} \in T_{S_1}^n$, encoding function $f_1(i) = v_i$, decoding function $g_1(v_i) = i$.

The second cycle: We know that for any $u \in T_{S_1}^n$, and any $j \in I_2 = \{1, \cdots, M_2\}$, there exists a vector $x_j(u) \in F_j$, such that $u \Longrightarrow x_j(u)$. We can define the encoding function $f_2(j, u) = x_j(u)$, and the decoding function $g_2(x) = j$, if $x \in F_j$.

In what follows, we show that there exists such a partition of $T_{S_2}^n$, by using a random coding method. With every $b \in T_{S_2}^n$, we connect a random index r_b which is uniformly distributed over the message set $I_2 = \{1, \cdots, M_2\}$, and all these random indices are independent. Define

$$F_m = \{b \in T_{S_2}^n | r_b = m\}, \qquad m = 1, 2, \cdots, M_2.$$

Then $\{F_m\}_{m=1}^{M_2}$ forms a random partition of $T_{S_2}^n$. For a fixed $u \in T_{S_1}^n$, set

$$G(u) = \{x \in T_{S_2}^n | u \Longrightarrow x\}, \qquad k_u = |G(u)|.$$

Since $S_1 \Longrightarrow S_2$, then $T_{S_2|S_1}^n(u) \subseteq G(u)$. By Property 4.4, we have

$$k_u = |G(u)| \ge |T_{S_2|S_1}^n(u)| \ge (n+1)^{-q^2} 2^{nH(S_2|S_1)}.$$

Fix $m \in I_2 = \{1, \cdots, M_2\}$, and $u \in T_{S_1}^n$, then

$$\begin{aligned} \Pr\{F_m \cap G(u) = \emptyset\} &= \Pr\{\text{for every } b \in G(u), r_b \ne m\} \\ &= \left[1 - \frac{1}{M_2}\right]^{k_u} \\ &\le \exp\{-k_u/M_2\} \\ &\le \exp\{-(n+1)^{-q^2} 2^{nH(S_2|S_1)}/M_2\} \\ &= \exp\{-(n+1)^{-q^2} 2^{n\delta}\} \to 0, \qquad n \to \infty. \end{aligned}$$

Therefore,

$$\begin{aligned} &\Pr\{\exists m \in I_2 \text{ and } u \in T_{S_1}^n, \text{s.t. } F_m \cap G(u) = \emptyset\} \\ &\quad \le M_2 \cdot |T_{S_1}^n| \cdot \exp\{-(n+1)^{-q^2} 2^{n\delta}\} \\ &\quad \le q^{2n} \cdot \exp\{-(n+1)^{-q^2} 2^{n\delta}\} \to 0, \qquad n \to \infty. \end{aligned}$$

This implies that when n is sufficiently large, there exists a partition of $T_{S_2}^n$, which satisfies **Property ♣**. This completes the proof of the direct part of Theorem 3.1. □

C. Proof of the Converse Part of Theorem 3.1

The Converse Part of Theorem 3.1: If there exists an $(n, T, M_1, \cdots, M_T)$ code for the $(\mathcal{V}, \mathcal{E})$-WOM, then the rate vector

$$((1//n)\log_2 M_1, (1/n)\log_2 M_2, \cdots, (1/n)\log_2 M_T) \in \mathcal{R}_T(\mathcal{V}, \mathcal{E}).$$

Proof: Let $W_1, W_2, \cdots, W_T$ be independent random variables, uniformly distributed over the message set $I_t = \{1, \cdots, M_t\}$, $t = 1, 2, \cdots, T$, respectively. Let $\{(f_t, g_t)\}_{t=1}^T$ be the T pairs of encoding and decoding functions for the $(n, T, M_1, \cdots, M_T)$ code. Denote $Y_0^n = \underline{0}$ (zero vector),

$$Y_t^n = (Y_{t,1}, Y_{t,2}, \cdots, Y_{t,n}) = f_t(W_t, Y_{t-1}^n), \qquad t = 1, \cdots, T.$$

It follows that $Y_{t-1}^n \Longrightarrow Y_t^n$, and $g_t(Y_t^n) = W_t, Y_{t-1,i} \Longrightarrow Y_{t,i}$, $t = 1, 2, \cdots, T$, $i = 1, 2, \cdots, n$. Because W_t and Y_{t-1}^n are independent, we have

$$\begin{aligned} H(W_t) &= H(W_t|Y_{t-1}^n) = H(W_t, Y_{t-1}^n|Y_{t-1}^n) \\ &\overset{1)}{\ge} H(Y_t^n|Y_{t-1}^n) \overset{2)}{\ge} H(W_t|Y_{t-1}^n) = H(W_t) \end{aligned}$$

where 1) follows from the fact that $Y_t^n = f_t(W_t, Y_{t-1}^n)$; 2) follows from the fact that $g_t(Y_t^n) = W_t$. Therefore,

$$H(W_t) = H(Y_t^n|Y_{t-1}^n), \qquad t = 1, 2, \cdots, T.$$

Let L be an index random variable, which uniformly distributes over the index set $\{1, 2, \cdots, n\}$, and L is independent of all other random variables. Then

$$\begin{aligned} \frac{1}{n}\log_2 M_t = \frac{1}{n}H(W_t) &= \frac{1}{n}H(Y_t^n|Y_{t-1}^n) \\ &\le \frac{1}{n} \sum_{i=1}^n H(Y_{t,i}|Y_{t-1,i}) \\ &\overset{1)}{=} H(Y_{t,L}|Y_{t-1,L}, L) \\ &\le H(Y_{t,L}|Y_{t-1,L}) \end{aligned}$$

where 1) follows from the fact that

$$\begin{aligned} H(Y_{t,L}|Y_{t-1,L}, L) &= \sum_{i=1}^n \Pr\{L = i\} H(Y_{t,L}|Y_{t-1,L}, L = i) \\ &= \frac{1}{n} \sum_{i=1}^n H(Y_{t,i}|Y_{t-1,i}). \end{aligned}$$

For every $t = 1, 2, \cdots, T$, $Y_{t-1,L} \Longrightarrow Y_{t,L}$, but random variables $Y_{1,L}, Y_{2,L}, \cdots, Y_{T,L}$ may not form a Markov chain. We take a new set of random variables $S_1, S_2, \cdots, S_T$, which take values in $\mathcal{V}$, and the joint probability distribution is defined by

$$\begin{aligned}\Pr\{S_1 = j_1, S_2 = j_2, \cdots, S_T = j_T\} &= \Pr\{Y_{1,L} = j_1\} \Pr\{Y_{2,L} = j_2 | Y_{1,L} = j_1\} \\ &\quad \cdots \Pr\{Y_{T,L} = j_T | Y_{T-1,L} = j_{T-1}\}.\end{aligned}$$

It is not difficult to see that for every $t = 1, 2, \cdots, T$, the random variables (S_{t-1}, S_t) and $(Y_{t-1,L}, Y_{t,L})$ have the same probability distribution. Therefore, $S_1 \Longrightarrow S_2 \Longrightarrow \cdots \Longrightarrow S_T$, and

$$H(S_1) = H(Y_{1,L}),$$
$$H(S_t|S_{t-1}) = H(Y_{t,L}|Y_{t-1,L}), \qquad t = 2, 3, \cdots, T.$$

Hence

$$\frac{1}{n}\log_2 M_t \le H(S_t|S_{t-1}), \qquad t = 1, 2, \cdots, T$$

where $S_0 = 0$. This implies that

$$\left(\frac{1}{n}\log_2 M_1, \frac{1}{n}\log_2 M_2, \cdots, \frac{1}{n}\log_2 M_T\right) \in \mathcal{R}_T(\mathcal{V}, \mathcal{E}). \qquad \square$$

V. Proof of Theorem 3.2:

A vector $\boldsymbol{x} = (x_0, x_1, \cdots, x_{q-1})$ is called positive if $x_0, x_1, \cdots, x_{q-1} > 0$. We denote

$$\log_2 \boldsymbol{x} = (\log_2 x_0, \log_2 x_1, \cdots, \log_2 x_{q-1}).$$

A vector $\boldsymbol{p} = (p_0, p_1, \cdots, p_{q-1})$ is called a probability vector if $p_0, p_1, \cdots, p_{q-1} \ge 0$ and $\Sigma_{i=1}^{q-1} p_i = 1$. It follows from the *Log sum inequality* (see [10, p. 29]) that

Lemma 5.1: For a positive constant vector

$$\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \cdots, \lambda_{q-1})$$

and a probability vector

$$\boldsymbol{p} = (p_0, p_1, \cdots, p_{q-1})$$

$$H(\boldsymbol{p}) + [\log_2 \boldsymbol{\lambda}] \cdot \boldsymbol{p}^c \le \log_2 \left[\sum_{i=0}^{q-1} \lambda_i\right] = \log_2(\boldsymbol{\lambda} \cdot \mathbf{1}_q^c).$$

Equality holds if and only if

$$p_i = \lambda_i \Big/ \left[\sum_{j=0}^{q-1} \lambda_j\right], \qquad i = 0, 1, \cdots, q-1.$$

For a DAG $(\mathcal{V}, \mathcal{E})$ with 0 as the root, we denote

$$\mathcal{V}_i = \{j \in \mathcal{V} | i \Longrightarrow j\}, \qquad i = 0, 1, \cdots, q-1$$

where $\mathcal{V}_0 = \mathcal{V}$. The transition matrix of the DAG $(\mathcal{V}, \mathcal{E})$ is defined as $\boldsymbol{A} = (a_{ij})_{q \times q}$, where $a_{ij} = 1$, $j \in \mathcal{V}_i$, and $a_{ij} = 0$, $j \notin \mathcal{V}_i$. Let

$$\boldsymbol{a}_i = (a_{i,0}, a_{i,1}, \cdots, a_{i,q-1}), \qquad i = 0, 1, \cdots, q-1.$$

For the random variables $S_1, S_2, \cdots, S_T$, which take values in $\mathcal{V}$, and $S_1 \Longrightarrow S_2 \Longrightarrow \cdots \Longrightarrow S_T$, we define

$$d_i^{(t)} = \Pr\{S_t = i\}, \qquad i = 0, 1, \cdots, q-1;\ t = 1, 2, \cdots, T$$
$$\boldsymbol{d}^{(t)} = (d_0^{(t)}, d_1^{(t)}, \cdots, d_{q-1}^{(t)}), \qquad t = 1, 2, \cdots, T$$

where $\boldsymbol{d}^{(t)}$, $t = 1, 2, \cdots, T$ are probability vectors. Assume

$$b_{i,j}^{(t)} = \Pr\{S_t = j | S_{t-1} = i\}, \qquad i, j = 0, 1, \cdots, q-1;\ t = 2, 3, \cdots, T$$

where $b_{i,j}^{(t)} = 0$, $j \notin \mathcal{V}_i$. We know that $\boldsymbol{b}_i^{(t)} = (b_{i,0}^{(t)}, b_{i,1}^{(t)}, \cdots, b_{i,q-1}^{(t)})$, $i = 0, 1, \cdots, q-1$; $t = 2, 3, \cdots, T$, are probability vectors, $\boldsymbol{B}^{(t)} = (b_{i,j}^{(t)})_{q \times q}$, $t = 2, 3, \cdots, T$, are stochastic matrices, and $\boldsymbol{d}^{(t-1)} \cdot \boldsymbol{B}^{(t)} = \boldsymbol{d}^{(t)}$. Then,

$$H(S_1) = H(\boldsymbol{d}^{(1)})$$
$$H(S_t|S_{t-1}) = \sum_{i=0}^{q-1} d_i^{(t-1)} H(\boldsymbol{b}_i^{(t)}), \qquad t = 2, 3, \cdots, T.$$

Suppose

$$\boldsymbol{A}^t \cdot \mathbf{1}_q^c = (\theta_0^{(t)}, \theta_1^{(t)}, \cdots, \theta_{q-1}^{(t)})^c, \qquad t = 0, 1, 2, \cdots, T,$$

where $\boldsymbol{A}^0 = I_q$ (unit matrix of order q).

Lemma 5.2: For any $m = 1, 2, \cdots, T$ we have

$$\sum_{l=1}^{m} H(S_{T+1-l}|S_{T-l}) \le \boldsymbol{d}^{(T-m)} \cdot \log_2(\boldsymbol{A}^m \cdot \mathbf{1}_q^c).$$

Equality holds if and only if for every $l = 1, 2, \cdots, m$

$$b_{i,j}^{(T+1-l)} = \begin{cases} \theta_j^{(l-1)} \Big/ \sum_{r \in \mathcal{V}_i} \theta_r^{(l-1)}, & \text{if } j \in \mathcal{V}_i \\ 0, & \text{if } j \notin \mathcal{V}_i. \end{cases} \qquad (\spadesuit)$$

Proof: We prove this lemma by using induction on m.

Case $m = 1$: From $b_{i,j}^{(T)} = 0$, $j \notin \mathcal{V}_i$, we have

$$H(\boldsymbol{b}_i^{(T)}) \le \log_2 |\mathcal{V}_i| = \log_2(\boldsymbol{a}_i \cdot \mathbf{1}_q^c).$$

Therefore,

$$H(S_T|S_{T-1}) = \sum_{i=0}^{q-1} d_i^{(T-1)} H(\boldsymbol{b}_i^{(T)}) \le \boldsymbol{d}^{(T-1)} \cdot \log_2(\boldsymbol{A} \cdot \mathbf{1}_q^c).$$

Equality holds if and only if $b_{i,j}^{(T)} = 1/|\mathcal{V}_i|$, if $j \in \mathcal{V}_i$, and $b_{i,j}^{(T)} = 0$, if $j \notin \mathcal{V}_i$. This implies that the assertion is true for $m = 1$. Assume that the assertion is true for $m = k$. Below we prove the assertion is true for $m = k + 1$.

$$\begin{aligned}&\sum_{l=1}^{k+1} H(S_{T+1-l}|S_{T-l}) \\ &= H(S_{T-k}|S_{T-k-1}) + \sum_{l=1}^{k} H(S_{T+1-l}|S_{T-l}) \\ &\le H(S_{T-k}|S_{T-k-1}) + \boldsymbol{d}^{(T-k)} \cdot \log_2(\boldsymbol{A}^k \cdot \mathbf{1}_q^c) \\ &= \sum_{i=0}^{q-1} d_i^{(T-k-1)} H(\boldsymbol{b}_i^{(T-k)}) + \boldsymbol{d}^{(T-k-1)} \cdot \boldsymbol{B}^{(T-k)} \log_2(\boldsymbol{A}^k \cdot \mathbf{1}_q^c) \\ &= \sum_{i=0}^{q-1} d_i^{(T-k-1)} [H(\boldsymbol{b}_i^{(T-k)}) + \boldsymbol{b}_i^{(T-k)} \cdot \log_2(\boldsymbol{A}^k \cdot \mathbf{1}_q^c)].\end{aligned}$$

By Lemma 5.1 and the fact $b_{i,j}^{(T-k)} = 0$, $j \notin \mathcal{V}_i$, we have

$$\begin{aligned}H(\boldsymbol{b}_i^{(T-k)}) + \boldsymbol{b}_i^{(T-k)} \cdot \log_2(\boldsymbol{A}^k \cdot \mathbf{1}_q^c) &\le \log_2\left(\sum_{r \in \mathcal{V}_i} \theta_r^{(k)}\right) \\ &= \log_2(\boldsymbol{a}_i \cdot \boldsymbol{A}^k \cdot \mathbf{1}_q^c).\end{aligned}$$

Hence

$$\sum_{l=1}^{k+1} H(S_{T+1-l}|S_{T-l}) \le \boldsymbol{d}^{T-k-1} \cdot \log_2(\boldsymbol{A}^{k+1} \cdot \mathbf{1}_q^c).$$

Equality holds if and only if for every $l = 1, 2, \cdots, k+1$, $(\spadesuit)$ holds. This implies that the assertion is true for $m = k+1$. This completes the proof of the lemma. $\square$

Proof of Theorem 3.2: From Lemma 5.2, we have

$$\sum_{t=1}^{T-1} H(S_{t+1}|S_t) \le \boldsymbol{d}^{(1)} \cdot \log_2(\boldsymbol{A}^{T-1} \cdot \mathbf{1}_q^c).$$

By Lemma 5.1 and $H(S_1) = H(\boldsymbol{d}^{(1)})$, we have

$$\begin{aligned} &H(S_1) + H(S_2|S_1) + \cdots + H(S_T|S_{T-1}) \\ &\quad \le H(\boldsymbol{d}^{(1)}) + \boldsymbol{d}^{(1)} \cdot \log_2(\boldsymbol{A}^{T-1} \cdot \mathbf{1}_q^c) \\ &\quad \le \log_2(\mathbf{1}_q \cdot \boldsymbol{A}^{T-1} \cdot \mathbf{1}_q^c), \end{aligned}$$

where equality holds if and only if for every $l = 1, 2, \cdots, T-1$, (♠) holds. This implies that $C_T(\mathcal{V}, \mathcal{E}) = \log_2(\mathbf{1}_q \cdot \boldsymbol{A}^{T-1} \cdot \mathbf{1}_q^c)$. □

Below we present a combinatorial proof for the converse part of Theorem 3.2. The main idea of this proof comes from Simonyi and Tardos [9], who presented a combinatorial proof of the converse coding theorem for write-unidirectional memory.

Property: For an $(n, T, M_1, \cdots, M_T)$ code of $(\mathcal{V}, \mathcal{E})$-WOM, we have

$$\prod_{i=1}^{T} M_i \le [\mathbf{1}_q \cdot \boldsymbol{A}^{T-1} \cdot \mathbf{1}_q^c]^n.$$

Proof: Assume

$$\hat{\mathcal{V}}(T) = \{(v_1, v_2, \cdots, v_T) \in \mathcal{V}^T | v_1 \Longrightarrow v_2 \Longrightarrow \cdots \Longrightarrow v_T\}.$$

It is easy to verify that

$$|\hat{\mathcal{V}}(T)| = \mathbf{1}_q \cdot \boldsymbol{A}^{T-1} \cdot \mathbf{1}_q^c.$$

Suppose $\{(f_t, g_t)\}_{t=1}^T$ are T pairs of encoding and decoding functions of an $(n, T, M_1, \cdots, M_T)$ code for the $(\mathcal{V}, \mathcal{E})$-WOM, see Definition 1 in Section II. Denote

$$\begin{aligned} \mathcal{Q}(n,T) = \{&(y_1^n, y_2^n, \cdots, y_T^n) | y_i^n \in \mathcal{V}^n, \text{ and there exist} \\ &m_1 \in I_1, m_2 \in I_2, \cdots, m_T \in I_T, \text{ such that} \\ &y_0^n = \underline{\mathbf{0}}, y_t^n = f_t(m_t, y_{t-1}^n),\ t = 1, \cdots, T\}. \end{aligned}$$

From Definition 1 of the $(\mathcal{V}, \mathcal{E})$-WOM code we know that there is a one-to-one correspondence between two sets $I_1 \times I_2 \times \cdots \times I_T$ and $\mathcal{Q}(n, T)$. Hence

$$\prod_{i=1}^{T} M_i = |\mathcal{Q}(n,T)|.$$

Since for every $(y_1^n, y_2^n, \cdots, y_T^n) \in \mathcal{Q}(n,T)$, we have $y_1^n \Longrightarrow y_2^n \Longrightarrow \cdots \Longrightarrow y_T^n$. Then $(y_1^n, y_2^n, \cdots, y_T^n) \in [\hat{\mathcal{V}}(T)]^n$. Therefore, $\mathcal{Q}(n,T) \subseteq [\hat{\mathcal{V}}(T)]^n$, and

$$\prod_{i=1}^{T} M_i = |\mathcal{Q}(n,T)| \le |\hat{\mathcal{V}}(T)|^n = [\mathbf{1}_q \cdot \boldsymbol{A}^{T-1} \cdot \mathbf{1}_q^c]^n. \quad \square$$

Remark: The proof above is only valid for zero-error codes. It is not difficult to see from the proofs of Theorems 3.1 and 3.2 in Section IV and V that Theorems 3.1 and 3.2 are still true for ε-error codes.

VI. Conclusions

In this correspondence, we study the problem of how to reuse a generalized WOM for T successive cycles. When the encoder knows and the decoder does not know the previous state of the memory and different codes are allowed to be used in every cycle, we determine the zero-error capacity region and the maximum total number of information bits stored in the memory for T successive cycles. These results extend the results of Wolf, Wyner, Ziv, and Körner for the binary WOM to the generalized WOM. By considering the previous state of the memory as side-information available to the encoder and decoder, Wolf, Wyner, Ziv, and Körner [2] studied the binary WOM in the following cases:

(E_+, D_+) Both the encoder and decoder know the previous state of the memory.

(E_+, D_-) The encoder knows and the decoder does not know the previous state of the memory.

(E_-, D_+) The encoder does not know and the decoder knows the previous state of the memory.

(E_-, D_-) Both the encoder and decoder do not know the previous state of the memory.

Using zero-error and ε-error as performance criteria, they investigated the problem of determining the capacity region and the maximum total number of information bits stored in the memory for T successive cycles.

The generalized WOM can be studied in the same way as that considered by Wolf, Wyner, Ziv, and Körner [2]. Here we study the generalized WOM only in the case (E_+, D_-) with zero-error codes, because this case is the most interesting and natural one. It should be pointed out that some other results for the binary WOM in [2] can also be established for the generalized WOM. There are still some unsolved difficult problems for the generalized WOM, for example, in the cases (E_-, D_-) and (E_-, D_+) with ε-error codes, the explicit formula for the maximum total number of information bits stored in the memory for T successive cycles is not known to us.

References

[1] R. L. Rivest and A. Shamir, "How to reuse a write-once memory," *Inform. Contr.*, vol. 55, no. 1, pp. 1–19, 1982.

[2] J. K. Wolf, A. D. Wyner, J. Ziv, and J. Körner, "Coding for a write-once memory," *AT&T Bell Labs. Tech. J.*, vol. 63, no. 6, pp. 1089–1112, 1984.

[3] G. D. Cohen, P. Godlewski, and F. Merkx, "Linear block codes for write-once memories," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 697–700, Sept. 1986.

[4] G. Zëmor, and G. D. Cohen, "Error-correcting WOM-codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 730–734, May 1991.

[5] A. Fiat and A. Shamir, "Generalized " write-once" memories," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 470–480, May 1984.

[6] C. Heegard, "On the capacity of permanent memory," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 34–42, Jan. 1985.

[7] A. V. Kuznetsov and A. J. H. Vinck, "On the general defective channel with informed encoder and capacities of some constrained memories," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1866–1871, Nov. 1994.

[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[9] G. Simonyi, "On write-unidirectional memories," *IEEE Trans. Inform. Theory*, vol. 35, pp. 663–667, May 1989.

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