On the Capacity of Write-Once Memories

Michal Horovitz, *Student Member, IEEE*, and Eitan Yaakobi, *Member, IEEE*

***Abstract*— Write-once memory (WOM) is a storage device consisting of q-ary cells that can only increase their value. A WOM code is a coding scheme that allows writing multiple times to the memory without decreasing the levels of the cells. In the conventional model, it is assumed that the encoder can read the memory state before encoding, while the decoder reads only the memory state after encoding. However, there are three more models in this setup, which depend on whether the encoder and the decoder are *informed* or *uninformed* with the previous state of the memory. These four models were first introduced by Wolf *et al.*, where they extensively studied the WOM capacity in these models for the binary case. In the non-binary setup, only the model, in which the encoder is informed and the decoder is not, was studied by Fu and Vinck. In this paper, we first present constructions of WOM codes in the models where the encoder is uninformed with the memory state (that is, the encoder *cannot* read the memory prior to encoding). We then study the capacity regions and maximum sum-rates of non-binary WOM codes for all four models. We extend the results by Wolf *et al.* and show that the capacity regions for the models in which the encoder is informed and the decoder is informed or uninformed in both the ϵ-error and the zero-error cases are all identical. We also find the ϵ-error capacity region; in this case, the encoder is uninformed and the decoder is informed and show that, in contrary to the binary case, it is a proper subset of the capacity region in the first two models. Several more results on the maximum sum-rate are presented as well.**

***Index Terms*— Coding theory, flash memories, write-once memory (WOM)-codes, multi-level memories, capacity region, maximum sum-rate, z-channel, erasure channel, asymmetric errors.**

I. INTRODUCTION

WRITE-ONCE memory (*WOM*) is a storage medium consisting of cells that can only increase their level. WOM codes were first introduced by Rivest and Shamir [19] and were designed to record data more than once in a WOM. Examples of such storage media are punch cards and optical disks, and more recently flash memories. The goal in designing a WOM code is to maximize the total number of bits which are written to the memory in t writes, while the cells can only increase their level. The *rate* on each write is the ratio between the number of bits stored in the memory and the number of cells, and the *sum-rate* is the sum of all individual rates. The *capacity region* of the WOM is the set of all achievable rate tuples.

It is usually assumed that the encoder can read the current state of the cells before programing, while the decoder has access only to the state of the cells after programming but not before that. This is the most practical model in which the encoder reads the memory before encoding, and the decoder reads the memory only after the encoding ends. However, there are four models of WOM which depend on whether the encoder and decoder on each write are informed with the previous state of the memory before encoding [26].

The case where the encoder is informed with the previous state of the memory is called *Encoder Informed* (*EI*) and otherwise *Encoder Uninformed* (*EU*). The cases of *Decoder Informed* (*DI*) and *Decoder Uninformed* (*DU*) are defined similarly. For shorthand, we refer to these four models as follows: model 1 - EI:DI, model 2 - EI:DU, model 3 - EU:DI, and model 4 - EU:DU. Note that all these models can be seen as a special case of coding for communication over a discrete memoryless channel (*DMC*) with state, where the state is known/unknown to the encoder and decoder; for more details see [25].

The model which was mostly studied in the literature is model 2 due to its practical relevance; see e.g. [3], [4], [10], [11], [13], [25], [27]. From the information-theoretic point of view, model 1 is the easiest one, while the most difficult one is model 4 since in model 4, both the encoder and decoder cannot read the memory before a new message is programmed. However, model 4 has significant practical advantage as it provides fast programming by saving an additional read before a write. Additionally, models 3 and 4 can be used for the construction of RIO codes, which are designed for fast programming and reading in flash memories [20], [28].

The binary case of these four models was rigorously defined and studied by Wolf *et al.* in [26]. The authors studied the capacity regions and maximum sum-rates for the four models in this case. In particular, they calculated the ϵ-error and the zero-error capacity regions in models 1 and 2 and showed that they are all identical and thus the maximum sum-rates in these cases as well. They also showed that this is the capacity region for model 3 for the ϵ-error case. Note that the zero-error capacity region and the maximum sum-rate in model 3 are still unknown. The ϵ-error capacity region for model 4 was partially solved by an achievable region (i.e., it is not known whether it is a tight region), however, it was still possible to calculate its maximum sum-rate [26]. For example, for two-write binary WOM in model 4, the maximum sum-rate is 1.3881 and for three writes, it is 1.600.

Manuscript received June 2, 2016; revised March 6, 2017; accepted March 7, 2017. Date of publication March 29, 2017; date of current version July 12, 2017. This work was supported in part by the Israel Science Foundation under Grant 1624/14 and in part by the German-Israeli Foundation under Grant I-1356-407.6/2016. This paper was presented in the IEEE Information Theory Workshop [14] and in the IEEE International Symposium on Information Theory [15]. *(Corresponding author: Eitan Yaakobi.)*

The authors are with the Department of Computer Science, Technion, Israel Institute of Technology, Haifa 32000, Israel (e-mail: michalho@cs.technion.ac; yaakobi@cs.technion.ac.il).

Communicated by J. Chen, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2017.2689034

Much less is known in the non-binary case, where only the zero-error capacity region for model 2 was calculated by Fu and Vinck [8] and the maximum sum-rate was shown to be $\log \binom{q-1+t}{q-1}$.[1] In [10], Gabrys and Dolecek investigated the rates achieved in each write in capacity achieving WOM codes in model 2, and the distribution of symbols in such WOM codes. They also studied the fixed-rate case, where the rates in all writes are equal. Noisy WOM in model 2 was studied by Heegrad [13] and by Wang and Kim [25]. In [25], the authors investigated the capacity region and maximum sum-rate in this model, where in [13] a more general case was considered. In this paper we assume that the WOM is noiseless.

We study all four models. First, constructions for WOM codes are proposed for models 3 and 4, the models in which the encoder is uninformed. For model 4, where the decoder is uninformed with the previous state, we show how codes in the Z channel provide constructions of binary WOM codes. This result is extended for non-binary WOM codes in which codes correcting non-binary asymmetric errors are used. Similarly in model 3, erasure-correcting codes are invoked for such constructions, and in the non-binary setup we use codes for handling asymmetric erasures. For the non-binary case in these two models we use codes for the Manhattan distance.

We also study the capacity regions of non-binary WOM for models 1, 2, and 3, and generalize some of the results by Wolf *et al.* for the non-binary setup. We first show that the same property of binary WOM in models 1 and 2 holds for non-binary as well. That is, the ϵ-error and the zero-error capacity regions for models 1 and 2 are all the same. Then, we notice and show that in contrast to the binary case, model 3 for non-binary does not behave the same as models 1 and 2 and its ϵ-error capacity region is a proper subset of the capacity regions in models 1 and 2. Furthermore, its maximum sum-rate is also smaller than the one for models 1 and 2. We derive more results to get upper and lower bounds on the maximum sum-rates in models 3 and 4.

The rest of this paper is organized as follows. In Section II, we formally define all four possible models of WOM and discuss existing results. We follow by presenting constructions of WOM codes for model 4 in Section III and for model 3 in Section IV. In Section V, we prove that the capacity regions of non-binary WOM in the first two models, both for the ϵ-error and the zero-error cases, are equal. Thus these four regions are all equal to the capacity region which was studied in [8] for non-binary WOM in the zero-error case of model 2. In Section VI, we calculate the ϵ-error capacity region of non-binary WOM in model 3, which in Section VII, is proved to be a proper subset of the capacity region in models 1 and 2. Additionally, Section VII includes results regarding the maximum sum-rates in the four models. Lastly, Section VIII concludes the paper and lists some open problems.

II. Definitions and Preliminaries

In this section we formally define the models studied in the paper and review the related previous results. The memory consists of n q-ary cells, where initially all of them are in the zero state. On each write, it is only possible to increase the level of each cell. For a positive integer q, the set $\{0, \dots, q-1\}$ is denoted by $[q]$. A vector $\boldsymbol{c} \in [q]^n$ will be called also a *cell-state vector*. The vector $\boldsymbol{c} = \max\{\boldsymbol{c}_1, \boldsymbol{c}_2\}$ is given by $c_i = \max\{c_{1,i}, c_{2,i}\}$ for all $i \in [n]$. For two vectors $\boldsymbol{x}, \boldsymbol{y} \in [q]^n$, we say that $\boldsymbol{x} \le \boldsymbol{y}$ if for all $i \in [n]$, $x_i \le y_i$ and $\boldsymbol{x} < \boldsymbol{y}$ is defined similarly. For a vector of probabilities $\boldsymbol{p} = (p_1, \dots, p_n)$, $H(\boldsymbol{p})$ denotes the entropy function, $H(\boldsymbol{p}) = -\sum_{i=1}^{n} p_i \log p_i$.

There are four families of WOM codes which were defined first in [26]. We follow the definition presented in [15] for both the zero-error and the ϵ-error cases.

Definition 1: A q-ary t-write WOM code with error probability vector $\boldsymbol{p}_e = (p_{e_1}, \dots, p_{e_t})$, denoted by $[n, t; M_1, \dots, M_t]_q^{Ⓚ, \boldsymbol{p}_e}$, for $k \in \{1, 2, 3, 4\}$, is a coding scheme comprising of n q-ary cells and is defined by t encoding and decoding maps $\mathcal{E}_i, \mathcal{D}_i$. For the map $\mathcal{E}_i$, $Im(\mathcal{E}_i)$ is the image of the map. By definition $Im(\mathcal{E}_0) = \{(0, \dots, 0)\}$, and for i, $1 \le i \le t$, $Im^*(\mathcal{E}_i) = \{\max\{\boldsymbol{c}_1, \dots, \boldsymbol{c}_i\} : \boldsymbol{c}_j \in Im(\mathcal{E}_j), 1 \le j \le i\}$. For a message m we denote by $Ind_m(x)$ the indicator function, where $Ind_m(x) = 0$ if $m = x$, otherwise $Ind_m(x) = 1$. The encoding and decoding maps are defined as follows:

1) *If $k = 1$ (encoder and decoder informed - EI:DI) then for each i*

$$\mathcal{E}_i : [M_i] \times Im(\mathcal{E}_{i-1}) \to [q]^n,$$
$$\mathcal{D}_i : \{(\mathcal{E}_i(m, \boldsymbol{c}), \boldsymbol{c}) : m \in [M_i], \boldsymbol{c} \in Im(\mathcal{E}_{i-1})\} \to [M_i]$$

such that for all $(m, \boldsymbol{c}) \in [M_i] \times Im(\mathcal{E}_{i-1})$, $\mathcal{E}_i(m, \boldsymbol{c}) \ge \boldsymbol{c}$, and

$$\sum_{(m,\boldsymbol{c}) \in [M_i] \times Im(\mathcal{E}_{i-1})} Pr(m) Pr(\boldsymbol{c}) \cdot Ind_m(\mathcal{D}_i(\mathcal{E}_i(m, \boldsymbol{c}), \boldsymbol{c})) \le p_{e_i}.$$

2) *If $k = 2$ (encoder informed, decoder uninformed - EI:DU) then for each i*

$$\mathcal{E}_i : [M_i] \times Im(\mathcal{E}_{i-1}) \to [q]^n, \mathcal{D}_i : Im(\mathcal{E}_i) \to [M_i]$$

such that for all $(m, \boldsymbol{c}) \in [M_i] \times Im(\mathcal{E}_{i-1})$, $\mathcal{E}_i(m, \boldsymbol{c}) \ge \boldsymbol{c}$, and

$$\sum_{(m,\boldsymbol{c}) \in [M_i] \times Im(\mathcal{E}_{i-1})} Pr(m) Pr(\boldsymbol{c}) \cdot Ind_m(\mathcal{D}_i(\mathcal{E}_i(m, \boldsymbol{c}))) \le p_{e_i}.$$

3) *If $k = 3$ (encoder uninformed, decoder informed - EU:DI) then for each i*

$$\mathcal{E}_i : [M_i] \to [q]^n, \mathcal{D}_i : Im^*(\mathcal{E}_i) \times Im^*(\mathcal{E}_{i-1}) \to [M_i]$$

such that

$$\sum_{(m,\boldsymbol{c}) \in [M_i] \times Im^*(\mathcal{E}_{i-1})} Pr(m) Pr(\boldsymbol{c}) \cdot Ind_m(\mathcal{D}_i(\max\{\boldsymbol{c}, \mathcal{E}_i(m)\}, \boldsymbol{c})) \le p_{e_i}.$$

4) *If $k = 4$ (encoder and decoder uninformed - EU:DU) then for each i*

$$\mathcal{E}_i : [M_i] \to [q]^n, \mathcal{D}_i : Im^*(\mathcal{E}_i) \to [M_i]$$

[1] All logarithms in this paper are taken according to base 2.

such that

$$\sum_{(m,\boldsymbol{c})\in[M_i]\times Im^*(\mathcal{E}_{i-1})} Pr(m)Pr(\boldsymbol{c})\cdot Ind_m(\mathcal{D}_i(\max\{\boldsymbol{c},\mathcal{E}_i(m)\}))\le p_{e_i}.$$

If $p_{e_i}=0$ for all $1\le i\le t$, then the code is called a zero-error WOM code and is denoted by $[n,t;M_1,\dots,M_t]_q^{ⓚ,z}$.

If $q=2$ then the WOM code is called a binary WOM code, and q is usually omitted from the notation, otherwise it is called a non-binary WOM code. The *rate* of the i-th write is the ratio between the number of written bits and the total number of cells, i.e., $\mathcal{R}_i=\frac{\log M_i}{n}$, and the *sum-rate* of the WOM code is the sum of the individual rates on all writes, i.e., $\mathcal{R}_t^{sum}=\sum_{i=1}^t\mathcal{R}_i$. A rate tuple $\boldsymbol{r}=(\mathcal{R}_1,\dots,\mathcal{R}_t)$ is called *ϵ-error achievable* in model k, $k\in\{1,2,3,4\}$, if for all $\epsilon>0$ there exists an $[n,t;M_1,\dots,M_t]_q^{ⓚ,\boldsymbol{p}_e}$ WOM code with error probability vector $\boldsymbol{p}_e=(p_{e_1},\dots,p_{e_t})\le(\epsilon,\dots,\epsilon)$, such that $\frac{\log M_i}{n}\ge\mathcal{R}_i-\epsilon$. The rate tuple $\boldsymbol{r}$ will be called *zero-error achievable* if for all $1\le i\le t$, $p_{e_i}=0$. The *capacity region* of q-ary t-write WOM is the set of all the achievable rate tuples. Let $C_{q,t}^{ⓚ,\epsilon}$, $C_{q,t}^{ⓚ,z}$ denote the capacity region, and $\mathcal{R}_{q,t}^{ⓚ,\epsilon}$, $\mathcal{R}_{q,t}^{ⓚ,z}$ denote the maximum sum-rate of q-ary t-write WOM in model k for the ϵ-error, the zero-error case, respectively.

In [26], Wolf *et al.* studied the binary case, and proved that the capacity regions of binary t-write WOM in the first two models (EI models) are equal, both for the ϵ-error and the zero-error cases. They also showed that this region is the capacity region of model 3 for the ϵ-error case. That is, the following holds for all $t\ge1$:

$$C_{2,t}\triangleq C_{2,t}^{①,z}=C_{2,t}^{①,\epsilon}=C_{2,t}^{②,z}=C_{2,t}^{②,\epsilon}=C_{2,t}^{③,\epsilon},\qquad(1)$$

where

$$\begin{aligned}C_{2,t}=\Big\{(\mathcal{R}_1,\dots,\mathcal{R}_t)|\ &\mathcal{R}_1\le h(p_1),\\ &\mathcal{R}_2\le h(p_2)(1-p_1),\dots,\\ &\mathcal{R}_{t-1}\le h(p_{t-1})\prod_{i=1}^{t-2}(1-p_i),\\ &\mathcal{R}_t\le\prod_{i=1}^{t-1}(1-p_i),\\ &\text{where } 0\le p_1,\dots,p_{t-1}\le1/2\Big\}.\qquad(2)\end{aligned}$$

It is also known that the maximum sum-rate in all these cases is $\log(t+1)$. Note that the zero-error capacity region of model 3, i.e. $C_{2,t}^{③,z}$, is still unknown, even for the binary case.

In the non-binary case, only model 2 was studied by Fu and Vinck [8]. They calculated the zero-error capacity region $C_{q,t}^{②,z}$, and found that the maximum sum-rate in this model, denoted by $\mathcal{R}_{q,t}^{②,z}$, is $\log\binom{q-1+t}{q-1}$. In this paper we show that the capacity regions of models 1 and 2 for both the ϵ-error and the zero-error cases are all the same, and thus $\log\binom{q-1+t}{q-1}$ is the maximum sum-rate in these four cases. One can readily conclude that this is an upper bound on the sum-rate also for the other models. However we prove that in models 3 and 4 this bound is not tight.

In Section III and Section IV we show how codes in the *Z channel*, and the *binary erasure channel* (*BEC*) provide constructions of binary WOM codes in model 4 and model 3, respectively. The Z channel is a channel with binary inputs and outputs in which only a zero can change to a one and not vice versa. These are called *asymmetric errors*. The BEC is a channel with binary inputs and outputs in which a bit value can be erased.

For each channel (Z channel or BEC), we use the following two models: the *coding theory model*, which is used for zero-error WOM codes, and the *information theory model* which is applied for the ϵ-error case. The coding theory model mimics an adversarial channel, where the adversary knows the entire codeword prior to transmission and can corrupt up to $\lceil pn\rceil$ locations to each specific transmission, where $p\in[0,1]$. This is the worst-case noise model studied in coding theory. In the information theory model, the errors are generated in an independent identical distribution, with probability p. We call p the *channel error probability*. For more details about these two models, see for example [7].

We say that a length-n code $\mathcal{C}$ with M codewords is an $(n,M,\tau,p_e)_{BEC}$ *erasure-correcting code*, if it can correct at most any τ erasures with *decoding error probability* p_e. The decoding error probability, p_e, is defined as the probability of decoding incorrectly the transmitted message, where the erasure vector is chosen independently uniformly from the set of all the vectors of Hamming weight at most τ. Note that p_e can be defined as maximal or average decoding error probability where the maximum and average measures are computed over the set of the messages, since both of these problems have the same capacity in discrete memoryless channel (DMC) [6, pp. 194, 204]. If $p_e=0$ then we omit this parameter, and we have an $(n,M,\tau)_{BEC}$ *erasure-correcting code*, which is capable of correcting any τ erasures. An $(n,M,\tau,p_e)_Z$ *asymmetric-error-correcting code*, and $(n,M,\tau)_Z$ *asymmetric-error-correcting code* are defined similarly. Note, that p_e is usually defined to be the decoding error probability where each bit is corrupted independently uniformly with probability p. For $\tau=\lceil pn\rceil$, this definition is equivalent to our definition described above, with respect to the capacity results. We use our definition to simplify the construction and to clarify the connection between the two models of the used channel.

Let K be the Z channel or the BEC with channel error probability p. It is said that $\mathcal{R}$ is an achievable rate in the information theory model of channel K, if for each $\epsilon>0$ there exists an $(n,M,\tau,p_e)_K$ error-correcting code consisting of $M\ge2^{n(\mathcal{R}-\epsilon)}$ codewords of length n, which is capable of correcting $\tau=\lceil pn\rceil$ errors which occurred by K with decoding error probability $p_e<\epsilon$. Similarly, $\mathcal{R}$ is an achievable rate in the coding theory model of a channel K, if for each $\epsilon>0$ there exists an $(n,M,\tau)_K$ error-correcting code with the same parameters except for $p_e=0$.

III. Constructions for Model 4 - The EU:DU Model

In this section we study constructions of WOM codes in model 4. We first present some known results [26] about the capacity region and the maximum sum-rate of binary WOM in the ϵ-error case.

In subsection III-A we study the binary two-write case. We show a very simple construction for the zero-error case using only two cells which already achieves a significantly high sum-rate. Then, using codes for the Z channel we give a more general construction of binary WOM codes in this model. We use codes in the information theory model in the Z channel, in order to construct WOM codes for the ϵ-error case, which obtain each point in the achievable region presented in [26] and thus the maximum sum-rate $\mathcal{R}_{2,2}^{④,\epsilon}$ as well. By the same ideas, codes in the coding theory model are applied to construct WOM codes in the zero-error case. Based on the two-write constructions, we then show in subsection III-B a recursive construction of binary multiple-write WOM codes. The generalization for non-binary two-write WOM is presented in Appendix A.

The capacity region of a binary t-write WOM in model 4 for the ϵ-error case was studied in [26]. The following region was shown to be achievable

$$\begin{aligned}\widetilde{C}_t = \Big\{(\mathcal{R}_1, \ldots, \mathcal{R}_t) | \mathcal{R}_1 \le h(p_1),\\ \mathcal{R}_2 \le h(p_1p_2) - p_2h(p_1), \ldots,\\ \mathcal{R}_{t-1} \le h(\prod_{j=1}^{t-1} p_j) - p_{t-1}h(\prod_{j=1}^{t-2} p_j),\\ \mathcal{R}_t \le h(\prod_{j=1}^{t} p_j) - p_t h(\prod_{j=1}^{t-1} p_j),\\ \text{where } 0 \le p_1, \ldots, p_t \le 1\Big\}, \quad (3)\end{aligned}$$

that is, $\widetilde{C}_t \subseteq C_{2,t}^{④,\epsilon}$. However, it is not known if the converse holds as well, i.e., whether every achievable rate tuple belongs to this region. Here, $1 - p_i$ is the probability for programming a bit on the i-th write. For example, for $t = 2$ we get

$$\begin{aligned}\widetilde{C}_2 = \Big\{(\mathcal{R}_1, \mathcal{R}_2) \mid \mathcal{R}_1 \le h(p_1),\\ \mathcal{R}_2 \le h(p_1p_2) - p_2h(p_1),\\ \text{where } 0 \le p_1, p_2 \le 1\Big\}. \quad (4)\end{aligned}$$

Even though, this capacity region is not necessarily a tight region, the authors in [26] could still give an achievable upper bound on the sum-rate in this model. Specifically, the maximum sum-rate, denoted by P_t in [26], was shown to be given by

$$\mathcal{R}_{2,t}^{④,\epsilon} = \sup_{0 \le p_1, \ldots, p_t \le 1} \left\{ h(\prod_{j=1}^{t} p_j) + \sum_{i=2}^{t}((1-p_i)h(\prod_{j=1}^{i-1} p_j)) \right\}. \quad (5)$$

Furthermore it was shown that for all $t \ge 1$, $\mathcal{R}_{2,t}^{④,\epsilon} \le \frac{\pi^2}{6\ln 2} \approx 2.37$, and $\lim_{t\to\infty} \mathcal{R}_{2,t}^{④,\epsilon} = \frac{\pi^2}{6\ln 2} \approx 2.37$. Table I presents the values of $\mathcal{R}_{2,t}^{④,\epsilon}$ for $2 \le t \le 5$, along with the probabilities vectors $\boldsymbol{p} = (p_1, p_2, \ldots, p_t)$ that maximize this term.

Fig. 1 illustrates the capacity regions for binary two-write WOM in all the models. It demonstrates that $\widetilde{C}_2$ (see Equation (4)), which is an achievable region for binary WOM in model 4 for the ϵ-error case (inner line), is a proper subset of $C_{2,2}$ (see Equation (2)), the capacity region of binary WOM in all the other models, (outer curve).

TABLE I
MAXIMUM SUM-RATES IN MODEL 4

t	$\mathcal{R}_{2,t}^{④,\epsilon}$	$\boldsymbol{p} = (p_1, p_2, \ldots, p_t)$
2	1.3881	$(0.665, 0.4169)$
3	1.600	$(0.7475, 0.59, 0.395)$
4	1.7356	$(0.80, 0.69, 0.57, 0.39)$
5	1.9695	$(0.85, 0.75, 0.675, 0.60, 0.40)$

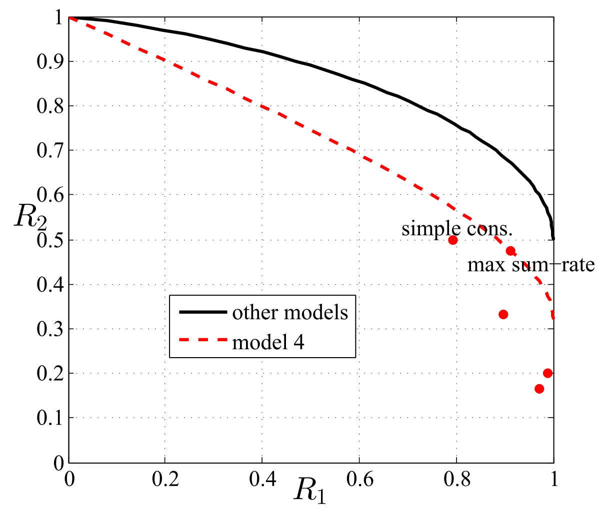

Fig. 1. A comparison between $C_{2,2}$-the capacity regions for binary two-write WOM in models 1, 2, and 3, and $\widetilde{C}_2$- an achievable region for binary two-write WOM in model 4. The points are rates of specific constructions of zero-error WOM codes in model 4.

A. Binary Two-Write WOM Codes

Let us start with a simple two-write WOM code construction.

Example 1: In this example, we show a construction of a binary $[2, 2; 3, 2]^{④,z}$ *WOM code. On the first write a ternary symbol is written according to the encoding map*

$$0 \to (0, 0), \quad 1 \to (0, 1), \quad 2 \to (1, 0).$$

Then, on the second write one more bit is written. If the bit value is zero the memory state does not change, that is the cells are programmed with $(0, 0)$. *Otherwise, the cells are programmed to the* $(1, 1)$ *state. The decoding on each write is clear from the encoding. The sum-rate of this construction is* $\frac{\log 3+1}{2} \approx 1.29$. *By concatenation, one can construct for each positive integer* n, *a* $[2n, 2; 3^n, 2^n]^{④,z}$ *WOM code with the same sum-rate.*

According to Table I, the maximum sum-rate of two-write WOM codes in the ϵ-error case is $\mathcal{R}_{2,2}^{④,\epsilon} = 1.388$, which is an upper bound for the zero-error case. Thus, this very simple example already achieves sum-rate which is fairly close to the upper bound. We show how to close on this gap for the ϵ-error case by using codes for the Z channel.

The construction of binary two-write WOM code we next propose is based on a reduction to the Z channel. We use two models of the channel, the coding theory model, which is invoked for constructing zero-error WOM codes, and the information theory model which is applied for the ϵ-error case, and achieves the capacity region $\widetilde{C}_2$ and thus the maximum sum-rate $\mathcal{R}_{2,2}^{④,\epsilon}$.

Recall that a length-n code $\mathcal{C}$ with M codewords is an $(n, M, \tau, p_e)_Z$ asymmetric-error-correcting code if it can correct at most τ asymmetric errors with decoding error probability p_e.

Theorem 1: Let $\mathcal{C}$ be an $(n, M, \tau, p_e)_Z$ asymmetric-error-correcting code. Then there exists an $[n, 2; M_1, M_2]^{④,p_e}$ WOM code with error probability vector $\boldsymbol{p}_e = (0, p_e)$, where $M_1 = \Sigma_{i=0}^{\tau}\binom{n}{i}$ and $M_2 = M$. If $p_e = 0$ then the constructed WOM code is a zero-error WOM code.

Proof: The proof will consist of describing the encoding and decoding maps of the WOM code. On the first write, M_1 messages are written by simply programming at most τ cells. We assume here, and later on, that there is a mapping between the set $[M_1]$ and the set of all binary vectors of Hamming weight at most τ. This mapping defines also the decoding of this write.

Let $\mathcal{E}, \mathcal{D}$ be the encoding, decoding map of the asymmetric-error-correcting code $\mathcal{C}$, respectively. The encoder on the second write receives a message $m \in [M_2]$ to be encoded to the memory and programs the cells with the vector $\mathcal{E}(m)$, given by applying the encoding map of $\mathcal{C}$. We denote the cell-state vector after the first write by $\boldsymbol{c}_1$. Since the vector $\boldsymbol{c}_1$ is already programmed to the memory, the cell-state vector on the second write becomes $\boldsymbol{c}_2 = \max\{\boldsymbol{c}_1, \mathcal{E}(m)\}$. The decoder on the second write applies the decoding map of $\mathcal{C}$ on $\boldsymbol{c}_2$. We have the following three observations:

1) $\mathcal{E}(m)$ is a codeword in $\mathcal{C}$,
2) $\boldsymbol{c}_2 \geq \mathcal{E}(m)$,
3) $d_H(\mathcal{E}(m), \boldsymbol{c}_2) \leq w_H(\boldsymbol{c}_1) \leq \tau$.

That is, the cell-state vector $\boldsymbol{c}_2$ is the outcome of at most τ asymmetric errors with respect to the codeword $\mathcal{E}(m)$. Since the code $\mathcal{C}$ is capable of correcting at most τ asymmetric errors with decoding error probability p_e, we have that $\mathcal{D}(\boldsymbol{c}_2) = m$ with probability at least $1 - p_e$, as required. □

Note that Example 1 is a special case of Theorem 1, in which the code $\mathcal{C}$ is a $(2, 2, 1)_Z$ asymmetric-error-correcting code. That is, the code is of length two, contains the two codewords $(0, 0)$ and $(1, 1)$ and it can correct a single asymmetric error.

The capacity of the Z channel in the information theory model was well studied in the literature before [22], [24]. Let α be the probability for occurrence of 1 in the codewords, and p be the crossover $0 \to 1$ probability, then the capacity was shown to be

$$h((1-\alpha)(1-p)) - (1-\alpha)h(p).$$

Thus, the capacity of the Z channel with the crossover $0 \to 1$ probability p equals to

$$\begin{aligned}\mathsf{cap}(Z) &= \max_{0\leq\alpha\leq 1}\{h((1-\alpha)(1-p)) - (1-\alpha)h(p)\}\\ &= \log\left(1 + (1-p)p^{p/(1-p)}\right).\end{aligned}$$

which is achieved for

$$\alpha = 1 - \frac{1}{(1-p)\left(1 + 2^{h(p)/(1-p)}\right)}.$$

In [26], Wolf *et al.* proved that it is possible to achieve all points in the region $\widetilde{C}_2$ by random coding. The next theorem proves this fact by using the construction from Theorem 1, and using capacity achieving codes in the Z channel in the information model. Theorems 1 and 2 provide more explicit constructions and important insight about the connection to the Z channel. Note that all the capacity achievable WOM codes are for the ϵ-error case.

Theorem 2: For any $\boldsymbol{r} \in \widetilde{C}_2$, $\boldsymbol{r}$ is ϵ-error achievable by the construction from Theorem 1.

Proof: We prove that for any $(\mathcal{R}_1, \mathcal{R}_2) \in \widetilde{C}_2$, and $\epsilon > 0$ there exists an $[n, 2; 2^{n\mathcal{R}'_1}, 2^{n\mathcal{R}'_2}]^{④,p_e}$ WOM code, constructed by Theorem 1, with error probability vector $\boldsymbol{p}_e = (0, p_e) \leq (\epsilon, \epsilon)$, and $\mathcal{R}'_1 \geq \mathcal{R}_1 - \epsilon$, $\mathcal{R}'_2 \geq \mathcal{R}_2 - \epsilon$.

Let $p_1, p_2 \in [0, 1]$ be such that $\mathcal{R}_1 \leq h(p_1)$ and $\mathcal{R}_2 \leq h(p_1 p_2) - p_2 h(p_1)$. Let $p = 1 - p_1$, $\alpha = 1 - p_2$, and $\epsilon > 0$. Based on the existence of capacity achieving codes for the Z channel, there exists an $(n, M, \tau, p_e)_Z$ asymmetric-error-correcting code $\mathcal{C}$, such that $\tau = \lceil pn \rceil$ if $p \in [0, 0.5]$[2] and otherwise $\tau = \lfloor pn \rfloor$, $p_e \leq \epsilon$, and

$$\frac{\log M}{n} \geq h((1-\alpha)(1-p)) - (1-\alpha)h(p) - \epsilon.$$

According to the construction from Theorem 1, there exists an $[n, 2; M_1, M_2]^{④,p_e}$ WOM code with error probability vector $\boldsymbol{p}_e = (0, p_e)$, where $M_1 = \sum_{i=0}^{\tau}\binom{n}{i}$ and $M_2 = M$. Based on Lemma 4.8 in [17]

$$\sum_{i=0}^{\tau}\binom{n}{i} \geq \frac{1}{n+1}2^{nh(\frac{\tau}{n})}.$$

Therefore, for n large enough, the rates of this WOM code satisfy

$$\begin{aligned}\mathcal{R}'_1 &= \frac{\log(\sum_{i=0}^{\tau}\binom{n}{i})}{n} \geq h\left(\frac{\tau}{n}\right) - \frac{\log(n+1)}{n}\\ &\geq h(p) - \frac{\log(n+1)}{n} \geq \mathcal{R}_1 - \epsilon,\end{aligned}$$

and

$$\begin{aligned}\mathcal{R}'_2 &= \frac{\log M_2}{n} \geq h((1-\alpha)(1-p)) - (1-\alpha)h(p) - \epsilon\\ &= h(p_1 p_2) - p_2 h(p_1) - \epsilon \geq \mathcal{R}_2 - \epsilon.\end{aligned}$$

□

As an immediate result from Theorem 2, we conclude that there exists a family of binary two-write WOM codes in model 4 for the ϵ-error case which achieve the maximum sum-rate $\mathcal{R}_{2,2}^{④,\epsilon}$.

Although the Z channel can provide us with capacity achieving codes for two writes for the ϵ-error case, yet it is not easy to find specific WOM codes with high sum-rates, mostly because the problem of finding such codes in the Z channel is still far to be solved.

By the same techniques, we can use codes for the coding theory model, which provide $(n, M, \tau)_Z$ asymmetric-error codes with decoding error probability zero, in order to construct zero-error WOM codes. However, the capacity in the

[2] If $\tau = 0.5$, we assume, without loss of generality, that n is even

TABLE II

AN $[n, 2; M_1, M_2]^{④,z}$ WOM CODE CONSTRUCTED BY AN $(n, M, \tau)_Z$ ASYMMETRIC-ERROR-CORRECTING CODE

$(n, M, \tau)_Z$	M_1	M_2	$\mathcal{R}_2^{sum}$ – sum-rate
$(5, 2, 4)_Z$	31	2	1.1908
$(6, 2, 4)_Z$	57	2	1.1388
$(6, 4, 3)_Z$	42	4	1.23205

coding theory model is unknown, and an upper-bound on the capacity is given by $\log\left(\frac{(\tau+1)2^n}{\sum_{j=0}^{\tau}\binom{n}{j}}\right)$ [1], [2]. Hence, if $\tau \approx pn$, and n goes to infinity, then the maximum sum-rate is upper-bounded by $1 - h(p)$. Thus, unfortunately, the maximum sum-rate of this construction for the zero-error case where $\tau \approx pn$, approaches 1 asymptotically.

We used some of the existing code constructions for asymmetric errors [12], and found WOM codes with the following parameters for the zero-error case, as described in Table II. The rates of these WOM codes are marked also in the plot of Fig. 1.

Note that the best sum-rate that we could find remains 1.29 which is achieved by the WOM code from Example 1 using two cells. The problem of closing this gap with specific WOM code constructions still remains an open problem. The extension of this construction to two-write non-binary WOM codes appears in Appendix A.

B. Binary t-Write WOM Codes

Given a binary two-write WOM code in model 4, we can construct binary t-write WOM codes for all t. We accomplish this goal by a recursive construction which is proved in the next theorem.

Theorem 3: If $\mathcal{C}_t$ is an $[n, t; M_1, M_2, \ldots, M_t]^{④,\boldsymbol{p}_e}$ WOM code, then there exists a $[2n, t+1; 3^n, M_1, M_2, \ldots, M_t]^{④,\boldsymbol{p}'_e}$ WOM code, $\mathcal{C}_{t+1}$, where $\boldsymbol{p}'_e = (0, p_{e_1}, \ldots, p_{e_t})$

Proof: The proof will consist of describing the encoding and decoding maps of the $(t+1)$-write WOM code $\mathcal{C}_{t+1}$. On the first write of $\mathcal{C}_{t+1}$, we invoke the first write of the two-write WOM code from Example 1. Thus, the $2n$ cells are divided into pairs, where each pair can be programmed to one of the following vectors $(0, 0)$, $(0, 1)$, or $(1, 0)$, resulting with 3^n messages.

For the next writes, each pair of two cells represents one logical cell, where the pair $(1, 1)$ represents a logical value one and the other three pairs represent a logical value zero. Thus, on the i-th write, $2 \le i \le t+1$, M_{i-1} messages can be encoded, decoded by using the encoder, decoder, of the $(i-1)$-st write of $\mathcal{C}_t$ over the n logical cells, respectively. □

By applying the construction of Theorem 3 recursively, and using a two-write WOM code with sum-rate $\mathcal{R}_2^{sum}$ as the parameter of the recursion, the sum-rate of the obtained t-write WOM code equals

$$\begin{aligned}\mathcal{R}_t^{sum} &= \log 3 \cdot \sum_{i=1}^{t-2}(2^{-i}) + 2^{2-t} \cdot \mathcal{R}_2^{sum} \\ &= \log 3 + 2^{2-t} \cdot (\mathcal{R}_2^{sum} - \log 3).\end{aligned}$$

TABLE III

A COMPARISON BETWEEN VALUES OF $\mathcal{R}_t^{sum}$ AND $\mathcal{R}_{2,t}^{④,\epsilon}$

t	$\mathcal{R}_2^{sum} = 1.29$	$\mathcal{R}_2^{sum} = 1.38$	$\mathcal{R}_{2,t}^{④,\epsilon}$
3	1.4387	1.4824	1.6000
4	1.5112	1.5337	1.7356
5	1.5480	1.5593	1.9695

The second and the third columns contain the values of $\mathcal{R}_t^{sum}$. The second column's values are achieved by zero-error codes using the binary two-write WOM code with sum-rate 1.29 from Example 1. The third column's values are attained for the ϵ-error case using capacity achieving binary two-write WOM codes.

Thus, the value $\mathcal{R}_t^{sum}$ approaches $\log 3$ as t gets large enough. In Table III, we present the results for the values of $\mathcal{R}_t^{sum}$ in case that $\mathcal{R}_2^{sum} = 1.29$ from Example 1, and for $\mathcal{R}_2^{sum} = \mathcal{R}_{2,2}^{④,\epsilon} = 1.388$, which is the maximum value for $\mathcal{R}_2^{sum}$ and can be asymptotically achieved by Theorem 2. These results are compared with the upper bound on the sum-rate, given by the value of $\mathcal{R}_{2,t}^{④,\epsilon}$.

IV. CONSTRUCTIONS FOR MODEL 3 - THE EU:DI MODEL

In this section we describe some constructions of WOM codes in model 3. First, we recall some known results [26] regarding the capacity region and the maximum sum-rate of binary WOM in the ϵ-error case in this model. Then, we study the binary two-write case. We show a very simple construction for the zero-error case. We also give a more general construction of binary WOM codes in this model, by using codes for the binary erasure channel (BEC).

As stated before in (1), the capacity regions of binary t-write WOM in model 1 and model 2 in both the ϵ-error and the zero-error cases, and the capacity region of binary t-write WOM in model 3 in the ϵ-error case are all identical [26]. This capacity region, denoted by $C_{2,t}$, is presented in (2), and the maximum sum-rate in these cases is $\log(t+1)$.

We start with a very simple construction for a zero-error WOM code using only three cells which is derived from a similar construction by Rivest and Shamir for model 2 [19].

Example 2: In this example, we show a construction of a binary $[3, 2; 4, 4]^{③,z}$ WOM code, achieving sum-rate $4/3$. Table IV describes the encoders' maps. The encoding map of the first write defines also the decoding of this write. We denote the cell-state vector after the first, second write by $\boldsymbol{c}_1$,$\boldsymbol{c}_2$, respectively. Let $\boldsymbol{u}$ be the vector which was encoded in the second write. The decoder input is $\boldsymbol{c}_1$ and $\boldsymbol{c}_2 = \max\{\boldsymbol{c}_1, \boldsymbol{u}\}$. It is possible to verify the following equation.

$$\boldsymbol{u} = \begin{cases} (0,0,0) & \text{if } w_H(\boldsymbol{c}_2) \le 1 \\ \boldsymbol{c}_2 & \text{if } w_H(\boldsymbol{c}_2) = 2 \\ \boldsymbol{c}_2 + \boldsymbol{c}_1 & \text{if } w_H(\boldsymbol{c}_2) = 3 \end{cases}$$

Thus, the decoder can reveal $\boldsymbol{u}$, and the message of the second write is decoded by the second column in Table IV.

We next give a general construction for binary two-write WOM in this model. This construction is based on a reduction to the BEC. We construct ϵ-error WOM codes by codes in

TABLE IV
THE ENCODER'S MAPS OF EXAMPLE 2

Message	First write	Second write
0	000	000
1	001	110
2	010	101
3	100	011

the information theory model of the BEC, which achieve the capacity region $C_{2,2}^{③,\epsilon}$, and thus the maximum sum-rate $\log 3$ as well, while for the zero-error case, we invoke codes in the coding theory model.

The construction is specified directly by codes correcting erasures. Recall that a length-n code $\mathcal{C}$ with M codewords is an $(n, M, \tau, p_e)_{BEC}$ erasure-correcting code if it can correct at most τ erasures with decoding error probability p_e. In particular, codes with minimum Hamming distance $\tau + 1$ can correct τ erasures with $p_e = 0$.

Theorem 4: Let $\mathcal{C}$ be an $(n, M, \tau, p_e)_{BEC}$ erasure-correcting code. Then there exists an $[n, 2; M_1, M_2]^{③,\boldsymbol{p}_e}$ WOM code with decoding error probability $\boldsymbol{p}_e = (0, p_e)$, where $M_1 = \Sigma_{i=0}^{\tau}\binom{n}{i}$ and $M_2 = M$. If $p_e = 0$ then the constructed code is zero-error WOM code.

Proof: The proof will consist of describing the encoding and decoding maps of the WOM code. On the first write M_1 messages are written by simply programming at most τ cells. Let $\mathcal{E}, \mathcal{D}$ be the encoding, decoding maps of the erasure-correcting code $\mathcal{C}$, respectively. The encoder on the second write receives a message $m \in [M_2]$ to be encoded to the memory and programs the cells with the vector $\mathcal{E}(m)$. We denote the cell-state vector after the first, second write by $\boldsymbol{c}_1$,$\boldsymbol{c}_2$, respectively. Thus, $\boldsymbol{c}_2 = \max\{\boldsymbol{c}_1, \mathcal{E}(m)\}$. Let $S = \{i : c_{1,i} = 1\}$. We have the following four observations:

1) $\mathcal{E}(m)$ is a codeword in $\mathcal{C}$,
2) $\boldsymbol{c}_2 \geq \mathcal{E}(m)$,
3) $d_H(\mathcal{E}(m), \boldsymbol{c}_2) \leq w_H(\boldsymbol{c}_1) = |S| \leq \tau$,
4) the set S is known to the decoder.

That is, the cell-state vector $\boldsymbol{c}_2$ is the outcome of at most τ erasures in the codeword $\mathcal{E}(m)$ where the set of the erasures' locations is S. Since the code $\mathcal{C}$ is capable of correcting at most τ erasures, we have that $\mathcal{D}(\boldsymbol{c}_2) = m$ with probability at least $1 - p_e$, as required. □

Note that Example 2 is a special case of Theorem 4, in which the code $\mathcal{C}$ is a $(3, 4, 1)_{BEC}$ erasure-correcting code. That is, the code is of length three, contains the following four codewords $(0, 0, 0)$, $(1, 1, 0)$, $(1, 0, 1)$, $(0, 1, 1)$, and can correct a single erasure.

The next theorem proves that by capacity achieving codes in the BEC in the information model, it is possible to achieve all points in the region $C_{2,2} = C_{2,2}^{③,\epsilon}$ through the construction from Theorem 4 for the ϵ-error case. Recall that the capacity of the BEC in the information theory model is known to be $1 - p$, where p is the erasure probability [6, pp. 189].

Theorem 5: For any $\boldsymbol{r} \in C_{2,2}$, $\boldsymbol{r}$ is achievable by the construction form Theorem 4.

Proof: We prove that for any $(\mathcal{R}_1, \mathcal{R}_2) \in C_{2,2}$, and $\epsilon > 0$ there exists an $[n, 2; 2^{n\mathcal{R}'_1}, 2^{n\mathcal{R}'_2}]^{③,\boldsymbol{p}_e}$ WOM code, constructed by Theorem 4, with error probability vector $\boldsymbol{p}_e = (0, p_{e_2}) \leq (\epsilon, \epsilon)$, and $\mathcal{R}'_1 \geq \mathcal{R}_1 - \epsilon$, $\mathcal{R}'_2 \geq \mathcal{R}_2 - \epsilon$.

Let $p \in [0, 0.5]$ be such that $\mathcal{R}_1 \leq h(p)$ and $\mathcal{R}_2 \leq 1 - p$. By capacity achieving codes in the BEC in the information theory model, for p and $\epsilon > 0$, there exists an $(n, M, \tau, p_e)_{BEC}$ erasure-correcting code $\mathcal{C}$, such that $\tau = \lceil pn \rceil$, $p_e < \epsilon$, and

$$\frac{\log M}{n} \geq 1 - p - \epsilon.$$

According to the construction from Theorem 4, there exists an $[n, 2; M_1, M_2]^{③,\boldsymbol{p}_e}$ WOM code with error probability vector $\boldsymbol{p}_e = (0, p_e) \leq (\epsilon, \epsilon)$, where $M_1 = \sum_{i=0}^{\tau}\binom{n}{i}$ and $M_2 = M$. Using Lemma 4.8 in [17], as in Theorem 2, the rates of this WOM code satisfy

$$\mathcal{R}'_1 = \frac{\log(\sum_{i=0}^{\tau}\binom{n}{i})}{n} \geq \mathcal{R}_1 - \epsilon,$$

and $\mathcal{R}'_2 = \frac{\log M_2}{n} \geq 1 - p - \epsilon$ for n large enough. □

By the same techniques, we can use codes for the coding theory model, which provide $(n, M, \tau)_{BEC}$ erasure-correcting codes with decoding error probability zero, in order to construct zero-error WOM codes in model 3. However, the capacity in the coding theory model is unknown. The best known achievable scheme for the coding model corresponds to codes suggested by Gilbert and Varshamov [9], [23] which achieve a rate of $1 - h(p)$ where pn is the maximum number of erased indices. Thus, unfortunately, this lower-bound doesn't provide sum-rate greater than 1. On the other hand, the best known upper bound is the MRRW (McEliece-Rodemich-Rumsey-Welch) bound obtained as the solution of an LP [16]. This bound, called the second MRRW bound, strengthens the following three bounds: the Hamming (sphere-packing) $1 - h(\frac{p}{2})$, the Elias-Bassalygo and the first MRRW bound $h(\frac{1}{2} - \sqrt{p(1-p)})$ for $0 < p < \frac{1}{2}$. This upper-bound provides us a limit on the power of this construction for the zero-error case. In particular, according to this bound the maximum sum-rate in this construction is upper bounded by 1.18. Note that by the WOM code in Example 2, we have already managed to achieve sum-rate 1.33. The extension of this construction to non-binary WOM codes appears in Appendix B.

V. CAPACITY REGION OF MODELS 1 AND 2 – EI MODELS

In this section we follow the derivations from [8] and [26] to prove equality between the capacity regions of q-ary t-write WOM in models 1 and 2 (EI models), and additionally to show equality between the ϵ-error and the zero-error capacity regions in these models. That is, we extend the result for models 1 and 2 stated in (1), for the non-binary case and show that for all q and t,

$$C_{q,t}^{②,z} = C_{q,t}^{②,\epsilon} = C_{q,t}^{①,\epsilon} = C_{q,t}^{①,z}. \tag{6}$$

Clearly, $C_{q,t}^{②,z} \subseteq C_{q,t}^{②,\epsilon} \subseteq C_{q,t}^{①,\epsilon}$, and $C_{q,t}^{②,z} \subseteq C_{q,t}^{①,z} \subseteq C_{q,t}^{①,\epsilon}$. Thus, in order to establish the equalities in (6), it is enough to prove that $C_{q,t}^{①,\epsilon} \subseteq C_{q,t}^{②,z}$.

Let us introduce several more notations to be used in this section. For $1 \leq i \leq t$, and $0 \leq j_1 \leq j_2 \leq q-1$, let $p_{i,j_1 \to j_2}$ be the conditional probability of writing the symbol j_2 on the i-th write given that the cell is in state j_1, and $\boldsymbol{p}_{i,j}$ is the probability vector $\boldsymbol{p}_{i,j} = (p_{i,j\to j}, p_{i,j\to(j+1)}, \ldots, p_{i,j\to(q-1)})$.

If $j_1 > j_2$, then $p_{i,j_1\to j_2} = 0$, so $p_{i,(q-1)\to(q-1)} = 1$, and $\boldsymbol{p}_{i,q-1} = (p_{i,(q-1)\to(q-1)})$ is a probability vector of length 1. We let $Q_{i,j}$ denote the probability that a cell's state after i writes is j, and $\mathbf{Q}_i = (Q_{i,0}, Q_{i,1}, \ldots, Q_{i,q-1})$. Note that $\mathbf{Q}_i$ is a function of $\boldsymbol{p}_{1,j}, \boldsymbol{p}_{2,j}, \ldots, \boldsymbol{p}_{i,j}$, $j \in [q]$, and can be computed recursively for all $i \geq 1$ and $j \in [q]$ as follows:

$$Q_{0,j} = \begin{cases} 1, & \text{if } j = 0. \\ 0, & \text{else} \end{cases}$$
$$Q_{i,j} = \sum_{k=0}^{j} Q_{i-1,k}\, p_{i,k\to j}. \tag{7}$$

We define the following region $C_{q,t}$ in order to show that it is the capacity region of models 1 and 2. This region can be readily verified to be equivalent to the one presented by Fu and Vinck [8] for the zero-error case, though its representation was more general.

$$C_{q,t} = \Big\{(\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_t) | \forall 1 \leq i \leq t : \\ \mathcal{R}_i \leq \sum_{j=0}^{q-2} Q_{i-1,j} H(\boldsymbol{p}_{i,j}), \\ \forall 1 \leq i \leq t, j \in [q] : \\ \boldsymbol{p}_{i,j} \text{ is a probability vector,} \\ Q_{i,j} \text{ is defined in (7)}\Big\}. \tag{8}$$

Note that for $i = t$, the upper bound $\log(q-j)$ of $H(\boldsymbol{p}_{t,j})$ can be achieved. So, one can also write $\mathcal{R}_t \leq \sum_{j=0}^{q-2} Q_{t-1,j} \log(q-j)$.

For example, the capacity region of binary t-write WOM in models 1 and 2 is presented in (2), where for $1 \leq i \leq t$, p_i in (2) is the probability to write symbol 1 on the i-th write, which was denoted in $C_{q,t}$ by $p_{i,0\to1}$. Therefore $1 - p_i$ equals $p_{i,0\to0}$, and $Q_{i,0} = \prod_{j=1}^{i}(1-p_j)$.

In addition, the capacity region of 3-ary two-write WOM in models 1 and 2 is

$$C_{3,2} = \Big\{(\mathcal{R}_1, \mathcal{R}_2) | \mathcal{R}_1 \leq H(p_{1,0\to0}, p_{1,0\to1}, p_{1,0\to2}), \\ \mathcal{R}_2 \leq p_{1,0\to0} \cdot \log 3 \\ + p_{1,0\to1} \cdot \log 2, \\ \text{where } \boldsymbol{p}_{1,0} = (p_{1,0\to0}, p_{1,0\to1}, p_{1,0\to2}) \\ \text{is a probability vector}\Big\}.$$

Since Fu and Vinck [8] have already showed that $C_{q,t} = C_{q,t}^{②,z}$, to complete the proof of (6) we are only required to prove that $C_{q,t}^{①,\epsilon} \subseteq C_{q,t}$, i.e. the converse part. In this proof and in the rest of the paper, we denote by X_i the vector written on the i-th write to the memory, and by Y_i the cell-state vector after the i-th write. We let Y_0 be the zero vector.

Theorem 6 (Converse part): If there exists an $[n, t; M_1, \ldots, M_t]_q^{①,\boldsymbol{p}_e}$ *WOM code, where* $\boldsymbol{p}_e = (p_{e_1}, \ldots, p_{e_t})$*, then*

$$\left(\frac{\log M_1}{n} - \epsilon_1, \frac{\log M_2}{n} - \epsilon_2, \ldots, \frac{\log M_t}{n} - \epsilon_t\right) \in C_{q,t},$$

where $\epsilon_i = \frac{H(p_{e_i}) + p_{e_i}\log(M_i)}{n}$.

Proof: Let $S_1, \ldots, S_t$ be independent random variables, where S_i is uniformly distributed over the messages set $[M_i]$. The data processing yields the following Markov chain:

$$S_i|Y_{i-1} — X_i|Y_{i-1} — Y_i|Y_{i-1} — \hat{S}_i|Y_{i-1}$$

and therefore, $I(X_i; Y_i|Y_{i-1}) \geq I(S_i; \hat{S}_i|Y_{i-1})$. Additionally,

$$\begin{aligned} I(S_i; \hat{S}_i|Y_{i-1}) &= H(S_i|Y_{i-1}) - H(S_i|\hat{S}_i, Y_{i-1}) \\ &\geq H(S_i) - H(S_i|\hat{S}_i) \\ &\geq \log(M_i) - H(p_{e_i}) - p_{e_i}\log(M_i). \end{aligned}$$

The first inequality follows from the independence of Y_{i-1} and S_i which implies that $H(S_i|Y_{i-1}) = H(S_i)$, and from the fact that conditioning does not increase the entropy. The second inequality follows from Fano's inequality [6, pp. 38] $H(S_i|\hat{S}_i) \leq H(p_{e_i}) + p_{e_i}\log(M_i)$. Let L be an index random variable, which is uniformly distributed over the index set $[n]$. Since L is independent of all other random variables we get

$$\begin{aligned} &\frac{1}{n} I(X_i; Y_i|Y_{i-1}) \\ &\leq \frac{1}{n} H(Y_i|Y_{i-1}) \\ &\overset{(a)}{\leq} \frac{1}{n} \sum_{k=0}^{n-1} H(Y_{i,k}|Y_{i-1,k}) \\ &\overset{(b)}{=} H(Y_{i,L}|Y_{i-1,L}, L) \\ &\overset{(c)}{\leq} H(Y_{i,L}|Y_{i-1,L}) \\ &= \sum_{j=0}^{q-1} Pr(Y_{i-1,L} = j) H(Y_{i,L}|Y_{i-1,L} = j) \\ &\overset{(d)}{=} \sum_{j=0}^{q-2} Pr(Y_{i-1,L} = j) H(Y_{i,L}|Y_{i-1,L} = j), \end{aligned}$$

where steps (a) and (c) follow from the fact that entropy of a vector is not greater than the sum of the entropies of its components, and conditioning does not increase the entropy. Step (b) follows from the fact that

$$\begin{aligned} H(Y_{i,L}|Y_{i-1,L}, L) &= \sum_{k=0}^{n-1} Pr(L = k) H(Y_{i,k}|Y_{i-1,L}, L{=}k) \\ &= \frac{1}{n} \sum_{k=0}^{n-1} H(Y_{i,k}|Y_{i-1,k}), \end{aligned}$$

and step (d) follows from $H(Y_{i,L}|Y_{i-1,L} = q-1) = 0$.

Now, we set $p_{i,j_1\to j_2} = Pr(Y_{i,L} = j_2|Y_{i-1,L} = j_1)$ and thus conclude that

$$\begin{aligned} Q_{i,j} &\triangleq Pr(Y_{i,L} = j) \\ &= \sum_{k=0}^{j} Pr(Y_{i,L} = j, Y_{i-1,L} = k) \\ &= \sum_{k=0}^{j} Pr(Y_{i,L} = j|Y_{i-1,L} = k) Pr(Y_{i-1,L} = k) \\ &= \sum_{k=0}^{j} p_{i,k\to j} Q_{i-1,k}, \end{aligned}$$

and

$$\begin{aligned} \frac{\log(M_i)}{n} - \epsilon_i &\leq \frac{1}{n} I(X_i; Y_i|Y_{i-1}) \\ &\leq \sum_{j=0}^{q-2} Pr(Y_{i-1,L} = j) H(Y_{i,L}|Y_{i-1,L} = j) \\ &= \sum_{j=0}^{q-2} Q_{i-1,j} H\left(p_{i,j\to j}, \ldots p_{i,j\to(q-1)}\right), \end{aligned}$$

where $\epsilon_i = \frac{H(p_{e_i})+p_{e_i}\log(M_i)}{n}$, and the converse part is implied. $\square$

Later on, we denote the capacity region $C_{q,t}$ by $C_{q,t}^{⑫}$, and the maximum sum-rate of this region by $\mathcal{R}_{q,t}^{⑫}$. Recall that $\mathcal{R}_{q,t}^{⑫} = \log\binom{q-1+t}{q-1}$ [8].

VI. Capacity Region of Model 3 – EU:DI Model

In this section we study the ϵ-error capacity of model 3. The binary case of this model was proved in [26], and was shown to be the same as the capacity $C_{2,t}$. We observe in our analysis that this property no longer holds for the non-binary case. Note that the capacity region in the zero-error case in this model is still unknown even for the binary case.

The programming probabilities in this case are not defined as in models 1 and 2, simply because the encoder can no longer read the memory state prior to encoding on each write. We let $p_{i,j}$ be the probability of writing the symbol j on the i-th write. Denote the probability vector $\boldsymbol{p}_{i,j}$ for $1 \le i \le t$ and $j \in [q-1]$ to be

$$\boldsymbol{p}_{i,j} = \left(\Sigma_{k=0}^{j} p_{i,k}, p_{i,j+1}, \dots, p_{i,q-1}\right). \tag{9}$$

We let $Q_{i,j}$ be the probability of a cell's state to be in level j after the i-th write, and $\mathbf{Q}_i = (Q_{i,0}, Q_{i,1}, \dots, Q_{i,q-1})$. Note that $\mathbf{Q}_i$ is a function of $p_{i,j}$ which is calculated recursively for all $i \ge 1$ and $j \in [q]$ as follows:

$$Q_{0,j} = \begin{cases} 1, & \text{if } j = 0. \\ 0, & \text{else} \end{cases}$$
$$Q_{i,j} = Q_{i-1,j}(\sum_{k=0}^{j-1} p_{i,k}) + p_{i,j}(\sum_{k=0}^{j} Q_{i-1,k}). \tag{10}$$

We define the region $\widehat{C}_{q,t}$, and then prove that $\widehat{C}_{q,t} = C_{q,t}^{③,\epsilon}$,

$$\begin{aligned}\widehat{C}_{q,t} = \Big\{(\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_t) | \forall 1 \le i \le t :& \\ \mathcal{R}_i \le \sum_{j=0}^{q-2} Q_{i-1,j} H(\boldsymbol{p}_{i,j}),& \\ \forall 1 \le i \le t, j \in [q] :& \\ \boldsymbol{p}_{i,j} \text{ is a probability vector}& \\ \text{as defined in (9), and}& \\ Q_{i,j} \text{ is defined in (10)}\Big\}.&\end{aligned} \tag{11}$$

For example, the capacity region of 3-ary 2-write WOM is

$$\begin{aligned}\widehat{C}_{3,2} = \Big\{(\mathcal{R}_1, \mathcal{R}_2) |\ & \mathcal{R}_1 \le H(p_{1,0}, p_{1,1}, p_{1,2}), \\ & \mathcal{R}_2 \le p_{1,0} H(p_{2,0}, p_{2,1}, p_{2,2}) \\ & \quad + p_{1,1} H(p_{2,0} + p_{2,1}, p_{2,2}) \\ & \text{where } 0 \le p_{i,j}, \sum_{j=0}^{2} p_{i,j} = 1\Big\}.\end{aligned}$$

Theorem 7: $\widehat{C}_{q,t}$ is the capacity region of ϵ-error q-ary t-write WOM in model 3, i.e., $\widehat{C}_{q,t} = C_{q,t}^{③,\epsilon}$.

The proof of Theorem 7 consists of two parts, which are presented in the following two sub-sections.

A. Proof of Direct Part – Theorem 7

For the direct part, we prove that for each $\epsilon > 0$, and $(\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_t) \in \widehat{C}_{q,t}$ there exists an $[n, t; M_1, \dots, M_t]_{q,t}^{③,\boldsymbol{p}_e}$ WOM code, where $\boldsymbol{p}_e = (p_{e_1}, \dots, p_{e_t}) \le (\epsilon, \dots, \epsilon)$, and $\frac{\log M_i}{n} \ge \mathcal{R}_i - \epsilon$ for all $1 \le i \le t$. We use the well-known random channel-coding theorem [6, pp. 200]. We describe the encoding and decoding on each write.

On the i-th write, the encoder writes the symbol j with probability $p_{i,j}$, i.e., for $k \in [n]$: $Pr(X_{i,k} = j) = p_{i,j}$. It is readily verified that that for $k \in [n]$: $Pr(Y_{i,k} = j) = Q_{i,j}$, where $Q_{i,j}$ is defined in (10). The decoder on the i-th round knows both Y_{i-1}, Y_i. Thus, the i-th write presents a DMC with input X_i, output $Z_i = (Y_{i-1}, Y_i)$, and transition probabilities, $Pr(Y_{i,k}|X_{i,k})$, determined by the probability $Q_{i-1,j}$. Let $x_i = X_{i,k}$, $y_{i-1} = Y_{i-1,k}$, $y_i = Y_{i,k}$ for some index k. By the random coding theorem, for n large enough, we can have highly reliable transmission with $p_{e_i} \le \epsilon$ and provided rate $\mathcal{R}_i = I(x_i; z_i) - \epsilon$. That is, the following region is achievable

$$\Big\{(\mathcal{R}_1, \dots, \mathcal{R}_t) | \mathcal{R}_i \le I(x_i; z_i = (y_{i-1}, y_i))\Big\},$$

where

$$\begin{aligned} I(x_i; z_i) &= H(z_i = (y_{i-1}, y_i)) - H(z_i|x_i) \\ &= H(y_{i-1}) + H(y_i|y_{i-1}) - H(z_i = (y_{i-1}, y_i)|x_i) \\ &\overset{(a)}{=} H(\mathbf{Q}_{i-1}) + H(y_i|y_{i-1}) - H(\mathbf{Q}_{i-1}) \\ &= H(y_i|y_{i-1}) \\ &= \sum_{j=0}^{q-1} Pr(y_{i-1} = j) H(y_i|y_{i-1} = j) \\ &\overset{(b)}{=} \sum_{j=0}^{q-2} Pr(y_{i-1} = j) H(y_i|y_{i-1} = j) \\ &\overset{(c)}{=} \sum_{j=0}^{q-2} Q_{i-1,j} H(\boldsymbol{p}_{i,j}). \end{aligned}$$

Step (a) follows from $H((y_{i-1}, y_i)|x_i) = H(y_{i-1}|x_i)$ since y_i is a function of x_i, y_{i-1}, and $H(y_{i-1}|x_i) = H(y_{i-1})$ because y_{i-1} and x_i are independent. Also, $H(y_i|y_{i-1} = q-1) = 0$ implies (b), and (c) is provided by

$$Pr(y_i = \ell | y_{i-1} = j) = \begin{cases} \sum_{k=0}^{j} p_{i,k}, & \text{if } \ell = j \\ p_{i,\ell}, & \text{if } \ell > j \\ 0, & \text{else.} \end{cases}$$

Thus, this region is exactly $\widehat{C}_{q,t}$, and we have $\widehat{C}_{q,t} \subseteq C_{q,t}^{③,\epsilon}$.

B. Proof of Converse Part – Theorem 7

In this section we prove the converse part of the capacity region. We use the same techniques as in model 1. Thus, many details, which are identical to the proof of Theorem 6, are omitted.

Converse part of Theorem 7: If there exists an $[n, t; M_1, \dots, M_t]_q^{①,\boldsymbol{p}_e}$ WOM code, where $\boldsymbol{p}_e = (p_{e_1}, \dots, p_{e_t})$, then

$$\left(\frac{\log M_1}{n} - \epsilon_1, \frac{\log M_2}{n} - \epsilon_2, \dots, \frac{\log M_t}{n} - \epsilon_t\right) \in \widehat{C}_{q,t},$$

where $\epsilon_i = \frac{H(p_{e_i})+p_{e_i}\log(M_i)}{n}$.

Proof: Let $S_1, \ldots, S_t$ be independent random variables as defined in the proof of Theorem 6. We can follow the same steps as in this proof to get

$$\begin{aligned} I(X_i; Y_i|Y_{i-1}) &\geq I(S_i; \hat{S}_i|Y_{i-1}) \\ &\geq \log(M_i) - H(p_{e_i}) - p_{e_i}\log(M_i) \end{aligned}$$

and

$$\frac{1}{n}I(X_i; Y_i|Y_{i-1}) \leq \sum_{j=0}^{q-2} Pr(Y_{i-1,L}=j)H(Y_{i,L}|Y_{i-1,L}=j)$$

where L is an index random variable, which is uniformly distributed over the index set $[n]$, and is independent of all other random variables. The random variables $X_{i,L}$ and $Y_{i-1,L}$ are independent in model 3, and $Y_{i,L} = \max\{X_{i,L}, Y_{i-1,L}\}$. Therefore

$$Pr(Y_{i,L} = \ell|Y_{i-1,L} = j) = \begin{cases} \sum_{k=0}^{j} Pr(X_{i,L} = k), & \text{if } \ell = j \\ Pr(X_{i,L} = \ell), & \text{if } \ell > j \\ 0, & \text{else.} \end{cases}$$

Now, by choosing $p_{i,j} \triangleq Pr(X_{i,L} = j)$, we can conclude that

$$\begin{aligned} \frac{\log(M_i)}{n} - \epsilon_i &\leq \frac{1}{n}I(X_i; Y_i|Y_{i-1}) \\ &\leq \sum_{j=0}^{q-2} Pr(Y_{i-1,L} = j)H(Y_{i,L}|Y_{i-1,L} = j) \\ &= \sum_{j=0}^{q-2} Q_{i-1,j}H\left(\boldsymbol{p}_{i,j}\right), \end{aligned}$$

where $\boldsymbol{p}_{i,j} = (\sum_{k=0}^{j} p_{i,k}, p_{i,k+1}, \ldots, p_{i,q-1})$ defined in (9), and $\mathbf{Q}_{i,j}$ defined in (10), and $\epsilon_i = \frac{H(p_{e_i})+p_{e_i}\log(M_i)}{n}$. Thus, the converse part is implied. $\square$

VII. Comparison Between the Capacities of the Models

The main goal of this section is to compare between the capacity regions of models 1 and 2 (EI models), and model 3 (EU:DI model). We prove that for all $q > 2$ and $t \geq 2$, the ϵ-error capacity region of q-ary t-write WOM in model 3 is a subset of the interior of the capacity region of q-ary t-write WOM in models 1 and 2. This implies that $C_{q,t}^{③,\epsilon} \subsetneq C_{q,t}^{⑫}$ and $\mathcal{R}_{q,t}^{③,\epsilon} < \mathcal{R}_{q,t}^{⑫}$, however, we will see that the difference between $\mathcal{R}_{q,t}^{③,\epsilon}$ and $\mathcal{R}_{q,t}^{⑫}$ is upper bounded by a constant which depends only on q but not on t. We also examine the probabilities which attain the maximum sum-rate. These results are demonstrated in Fig. 2 and Fig. 3. In addition, finding the capacity region of model 4 is an open question, even for the binary case. However, in [26] it was proved that the maximum sum-rate of binary t-write WOM in model-4 is finite where $t \to \infty$, while in the other models it goes to infinity. We generalize this result for all $q > 2$.

Even though the next lemma is a straightforward property we use its proof in deriving the other results in this section.

Lemma 1: If $\boldsymbol{r} \in \widehat{C}_{q,t}$ then $\boldsymbol{r} \in C_{q,t}$, i.e., $\widehat{C}_{q,t} \subseteq C_{q,t}$.

Proof: The claim is derived simply by

$$\widehat{C}_{q,t} = C_{q,t}^{③,\epsilon} \subseteq C_{q,t}^{①,\epsilon} = C_{q,t}.$$

However we present another proof, which examines the probabilities for which $\boldsymbol{r}$ is attained in $C_{q,t}$.

Let $\boldsymbol{r} = (\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_t) \in \widehat{C}_{q,t}$ where $\boldsymbol{r}$ is achieved by probabilities $p_{i,j}$, and define

$$\widehat{p}_{i,j_1\to j_2} = \begin{cases} \sum_{k=0}^{j_2} p_{i,k}, & \text{if } j_2 = j_1 \\ p_{i,j_2}, & \text{if } j_2 > j_1 \\ 0, & \text{else.} \end{cases}$$

Thus, one can readily verify that $\boldsymbol{r}$ is achieved in $C_{q,t}$ for probabilities $\widehat{p}_{i,j_1\to j_2}$. It is possible to show that the probabilities to have symbol j in a cell after the i-th write, denoted by $Q_{i,j}$, are equal in both regions, $C_{q,t}$ and $\widehat{C}_{q,t}$. $\square$

Lemma 2: If $\boldsymbol{r} = (\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_t) \in \widehat{C}_{q,t}$ and $\mathbf{0} \neq (\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_{i-1})$, then there exists $\boldsymbol{r}' = (\mathcal{R}'_1, \mathcal{R}'_2, \ldots, \mathcal{R}'_t) \in \widehat{C}_{q,t}$ such that $\boldsymbol{r} \leq \boldsymbol{r}'$ and $\boldsymbol{r}'$ is attained by $p'_{i,j}$, $1 \leq i \leq t$, $j \in [q]$, with $Q'_{i-1,0}, Q'_{i-1,1} > 0$ where $Q'_{i-1,0}, Q'_{i-1,1}$ are defined in (10).

Proof: Denote by $p_{i,j}$, $1 \leq i \leq t$, $j \in [q]$, the parameters for which $\boldsymbol{r}$ is attained $\widehat{C}_{q,t}$. We prove that the exist a set of parameters $p'_{i,j}$, $1 \leq i \leq t$, $j \in [q]$, which correspond to a rate vector $\boldsymbol{r}'$ such that $\boldsymbol{r}' \geq \boldsymbol{r}$. Let j be the smallest number such that $\mathcal{R}_j > 0$. Note that $j < i$ since $\mathbf{0} \neq (\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_{i-1})$. We can choose $p'_{1,0} = p'_{2,0} = \ldots = p'_{j-1,0} = 1$ provides $\mathcal{R}'_1 = \mathcal{R}'_2 = \ldots = \mathcal{R}'_{j-1} = 0$. Note the entropy of a vector is determined by the values of its components disregarding their order. Thus, on the j-th write, we can choose the probability vector $(p_{j,0}, p_{j,1}, \ldots, p_{j,q-1})$ in non-increasing order which provides $\mathcal{R}'_j \geq \mathcal{R}_j$, and $p'_{j,0} \geq p'_{j,1} > 0$ since $\mathcal{R}_j > 0$. For the next writes, let $\ell > j$, if $p_{\ell,0} = 0$, then let k be the smallest number in $[q]$ such that $p_{\ell,k} > 0$. If $k = 1$, then $p'_{\ell,0} = p'_{\ell,1} = \frac{p_{\ell,1}}{2}$ else, $k > 1$, then $p'_{\ell,0} = p'_{\ell,1} = p'_{\ell,k} = \frac{p_{\ell,1}}{3}$. It can be easily verified the $\mathcal{R}'_\ell \geq \mathcal{R}_\ell$ for all $\ell > j$, and $Q'_{i',0}, Q'_{i',1} > 0$ for all $i' \geq j$, and in particular $Q'_{i-1,0}, Q'_{i-1,1} > 0$ as required. $\square$

In the next three lemmas, we will show that $\widehat{C}_{q,t}$ is a subset of the interior of $C_{q,t}$.

Lemma 3: If $\boldsymbol{r} = (\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_t) \in \widehat{C}_{q,t}$ and $\mathbf{0} \neq (\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_{t-1})$, then there exists an $\epsilon > 0$ such that $\boldsymbol{r}_{t,+\epsilon} \triangleq (\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_t + \epsilon) \in C_{q,t}$.

Proof: If there exists an $\epsilon > 0$ such that $\boldsymbol{r}_{t,+\epsilon} \in \widehat{C}_{q,t}$, then by Lemma 1, the claim is obviously true. Otherwise, by Lemma 2, there exists $\boldsymbol{r}' = (\mathcal{R}'_1, \mathcal{R}'_2, \ldots, \mathcal{R}'_t) \in \widehat{C}_{q,t}$ such that $\boldsymbol{r} \leq \boldsymbol{r}'$ and $\boldsymbol{r}'$ is attained by $p_{i,j}$, $1 \leq i \leq t$, $j \in [q]$, where $Q_{t-1,0}, Q_{t-1,1} > 0$ are defined in (10). We can assume that $\mathcal{R}'_t = \mathcal{R}_t$, and $\forall\epsilon > 0 : \ \boldsymbol{r}'_{t,+\epsilon} \notin \widehat{C}_{q,t}$ (else, there exists an $\epsilon > 0$ such that $\boldsymbol{r}_{t,+\epsilon} \in \widehat{C}_{q,t}$, and this is the first case). Thus, we can conclude that $\mathcal{R}'_t$ is equal to

$$\max_{\substack{(p'_{t,0},\ldots,p'_{t,q-1}) \\ \text{prob. vector}}} \left\{ \sum_{j=0}^{q-2} Q_{t-1,j}H\left((\sum_{k=0}^{j} p'_{t,k}), p'_{t,j+1}, \ldots, p'_{t,q-1} \right) \right\},$$

where $p'_{t,j}$ is a probability to write symbol j on the t-th write, and $Q_{t-1,j}$ is defined in (10). By the proof of Lemma 1, $\boldsymbol{r}' \in \widehat{C}_{q,t}$ implies $\boldsymbol{r}' \in C_{q,t}$, where $Q_{t-1,j}$, the probabilities to have symbol j after the $(t-1)$-th write, are equals in both regions, $C_{q,t}$ and $\widehat{C}_{q,t}$. Let us define $\widehat{\mathcal{R}}_t \triangleq \sum_{j=0}^{q-2} Q_{t-1,j}\log(q-j)$. Clearly $(\mathcal{R}'_1, \mathcal{R}'_2, \ldots, \widehat{\mathcal{R}}_t) \in C_{q,t}$, and we prove now that $\mathcal{R}'_t < \widehat{\mathcal{R}}_t$. Note that for each $\boldsymbol{x}$, a probability vector of length m,

$$H(x_0, x_1, \ldots, x_{m-1}) \leq \log(m),$$

with equality if and only if $x_i = \frac{1}{m}$ for all $i \in [m]$. Additionally, $\mathcal{R}'_t = \widehat{\mathcal{R}}_t$ if and only if for all $j \in [q-2]$

$$Q_{t-1,j} H\left((\sum_{k=0}^{j} p_{t,k}), p_{t,j+1}, \ldots, p_{t,q-1}\right) = Q_{t-1,j} \log(q-j).$$

Recall that $Q_{t-1,0}, Q_{t-1,1} > 0$. Thus, from $j = 0$ we derive $p_{t,k} = \frac{1}{q}$ for all $k \in [q]$, and similarly from $j = 1$ we have $p_{t,k} = \frac{1}{q-1}$ for all $2 \leq k \in [q]$, and $p_{t,0} + p_{t,1} = \frac{1}{q-1}$. These two conditions cannot hold simultaneously. Thus, we conclude that $\mathcal{R}_t = \mathcal{R}'_t < \widehat{\mathcal{R}}_t$.

Therefore, one can readily verify that there exists $\epsilon > 0$ such that $\boldsymbol{r}_{t,+\epsilon} \in C_{q,t}$ □

We can extend Lemma 3 so it is possible to increase each coordinate in a rate tuple.

Lemma 4: If $\boldsymbol{r} = (\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_i, \ldots, \mathcal{R}_t) \in \widehat{C}_{q,t}$ and $\boldsymbol{0} \neq (\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_{i-1})$, then there exists an $\epsilon > 0$ such that $\boldsymbol{r}_{i,+\epsilon} \triangleq (\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_i + \epsilon, \ldots, \mathcal{R}_t) \in C_{q,t}$.

Proof: If there exists $\epsilon > 0$ such that $\boldsymbol{r}_{i,+\epsilon} \in \widehat{C}_{q,t}$, then by Lemma 1, the claim is obviously true. Otherwise we prove it by induction of i, where the base case is $i = t$ which was proved in Lemma 3. By induction hypothesis, for all $\ell \in \{i+1, i+2, \ldots, t\}$ there exists $\epsilon_\ell > 0$ such that $\boldsymbol{r}_{\ell,+\epsilon_\ell} \in C_{q,t}$. Thus, by time-sharing technique, for $\epsilon' = \min\{\epsilon_\ell/(t-i)\}_{\ell=i+1}^{t}$ we have $\boldsymbol{r}_1 = (\mathcal{R}_1, \ldots, \mathcal{R}_i, \mathcal{R}_{i+1} + \epsilon', \mathcal{R}_{i+2} + \epsilon', \ldots, \mathcal{R}_t + \epsilon') \in C_{q,t}$. By Lemma 2, there exists $\boldsymbol{r}' = (\mathcal{R}'_1, \mathcal{R}'_2, \ldots, \mathcal{R}'_t) \in \widehat{C}_{q,t}$ such that $\boldsymbol{r} \leq \boldsymbol{r}'$ attained by $p_{i,j}$, $1 \leq i \leq t$, $j \in [q]$, where $Q_{i-1,0}, Q_{i-1,1} > 0$. By the same techniques as in the previous lemma, we can prove that there exists $\epsilon'' > 0$ such that $\boldsymbol{r}_2 = (\mathcal{R}'_1, \mathcal{R}'_2, \ldots, \mathcal{R}'_{i-1}, \mathcal{R}'_i + \epsilon'', 0, \ldots, 0) \in C_{q,t}$. Using $\boldsymbol{r}_1, \boldsymbol{r}_2 \in C_{q,t}$ and time sharing method, we can conclude that there exists an $\epsilon > 0$, such that $\boldsymbol{r} \leq \boldsymbol{r}^* = (\mathcal{R}^*_1, \mathcal{R}^*_2, \ldots, \mathcal{R}^*_{i-1}, \mathcal{R}'_i + \epsilon, \mathcal{R}^*_{i+1}, \ldots, \mathcal{R}^*_t) \in C_{t,q}$, and therefore $\boldsymbol{r}_{i,+\epsilon} \in C_{t,q}$. □

Based on Lemma 4 by using time-sharing technique, we conclude the following lemma.

Lemma 5: If $(\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_t) \in \widehat{C}_{q,t}$ then there exists an $\epsilon > 0$ such that $(\mathcal{R}_1 + \epsilon, \mathcal{R}_2 + \epsilon, \ldots, \mathcal{R}_t + \epsilon) \in C_{q,t}$.

Finally, the next corollary summarizes the discussion above.

Corollary 1: For all $q > 2$ and $t \geq 2$, $C_{q,t}^{③,\epsilon} \subsetneq C_{q,t}^{⑫}$ and $\mathcal{R}_{q,t}^{③,\epsilon} < \mathcal{R}_{q,t}^{⑫}$.

To illustrate the results in Corollary 1, we show in Fig. 2, the capacity regions of $C_{3,2}, \widehat{C}_{3,2}$ which compare between models 1 and 2 and model 3.

Fu and Vinck [8] found the probabilities which attain the maximum sum-rate for models 1 and 2. Even though we cannot carry the same analysis for model 3, we can still have the following result for the probabilities that achieve the maximum sum-rate in this case.

Lemma 6: Assume that $\boldsymbol{r} \in \widehat{C}_{q,t}$ achieves the maximum sum-rate. Then, there exist probabilities $p_{i,j}$ for $1 \leq i \leq t$, $j \in [q]$ which correspond to the rate tuple $\boldsymbol{r}$ and $p_{t,0} = p_{t,1}$.

Proof: according to Lemma 2, if $\boldsymbol{r} \in \widehat{C}_{q,t}$ then there exists $\boldsymbol{r}' = (\mathcal{R}'_1, \mathcal{R}'_2, \ldots, \mathcal{R}'_t) \in \widehat{C}_{q,t}$ such that $\boldsymbol{r} \leq \boldsymbol{r}'$ and $\boldsymbol{r}'$ attained by $p_{i,j}$, $1 \leq i \leq t$, $j \in [q]$, where $Q_{t-1,0}, Q_{t-1,1} > 0$. Note that $\boldsymbol{r} = \boldsymbol{r}'$ since $\boldsymbol{r}$ is maximum sum-rate point, and

$$\mathcal{R}_t = \left\{\sum_{j=0}^{q-2} Q_{t-1,j} H\left((\sum_{k=0}^{j} p_{t,k}), p_{t,j+1}, \ldots, p_{t,q-1}\right)\right\},$$

All the summands in $\mathcal{R}_t$ equation except for the first, use only the value $p_{t,0} + p_{t,1}$, while the first summand achieves maximum for $p_{t,0} = p_{t,1}$. Since $\boldsymbol{r}$ is a maximum sum-rate point, we can conclude that $p_{t,0} = p_{t,1}$. □

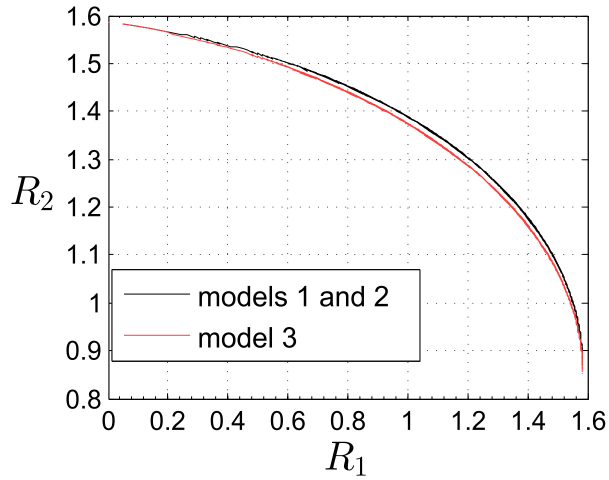

Fig. 2. A comparison between the capacity regions $C_{3,2}$ (models 1 and 2—the outer line) and $\widehat{C}_{3,2}$ (model 3—the inner line).

Even though we do not know the exact maximum sum-rate for non-binary WOM in models 3 and 4, it is still possible to derive a lower and upper bound in order to have better estimations on these values. The following result is proved by similar techniques from [11].

Lemma 7: For all $q \geq 2, t \geq 1, k \in \{3, 4\}, x \in \{z, \epsilon\}$:

$$(q-2)\mathcal{R}_{2,\lfloor\frac{t}{q-1}\rfloor}^{ⓚ,x} + \mathcal{R}_{2,t-(q-2)\lfloor\frac{t}{q-1}\rfloor}^{ⓚ,x} \leq \mathcal{R}_{q,t}^{ⓚ,x} \leq (q-1)\mathcal{R}_{2,t}^{ⓚ,x}.$$

Proof: The left inequality proved by using a construction described in [11]. Once we have binary WOM codes in model k, C_1 for $t_1 = \lfloor\frac{t}{q-1}\rfloor$ writes and C_2 for $t_2 = t - (q-2)\lfloor\frac{t}{q-1}\rfloor$ writes, we can construction a q-ary t-write WOM code by using the q levels "layer by layer". The t rounds are divided to $q-1$ stages where each stage (except the last) contains t_1 writes. On the first stage, only $0, 1$ symbols are used according to C_1, on the next stage only the $1, 2$ levels are used according to C_1 by mapping $0 \mapsto 1, 1 \mapsto 2$, and so on. The last binary WOM code that will be used is C_2 contains $t - (q-2)\lfloor\frac{t}{q-1}\rfloor$ writes.

The proof of the right inequality consists of a reduction. Given C, an $[n, t; M_1, \ldots, M_t]_q^{ⓚ,\boldsymbol{p}_e}$ $(1 \leq k \leq 4)$ WOM code, with sum-rate $\mathcal{R}_{q,t}^{sum} = \frac{\sum_{i=1}^{t}\log(M_i)}{n}$, we can construct C', an $[n(q-1), t; M_1, \ldots, M_t]_2^{ⓚ,\boldsymbol{p}_e}$ WOM code, with sum-rate $\mathcal{R}_{2,t}^{sum} = \frac{\sum_{i=1}^{t}\log(M_i)}{n(q-1)} = \frac{\mathcal{R}_{q,t}^{sum}}{q-1}$. Thus, we have

$$\mathcal{R}_{q,t}^{sum} = (q-1)\mathcal{R}_{2,t}^{sum} \leq (q-1)\mathcal{R}_{2,t}^{ⓚ,x},$$

where $x = z$ if $\boldsymbol{p}_e = \boldsymbol{0}$, otherwise $x = \epsilon$.

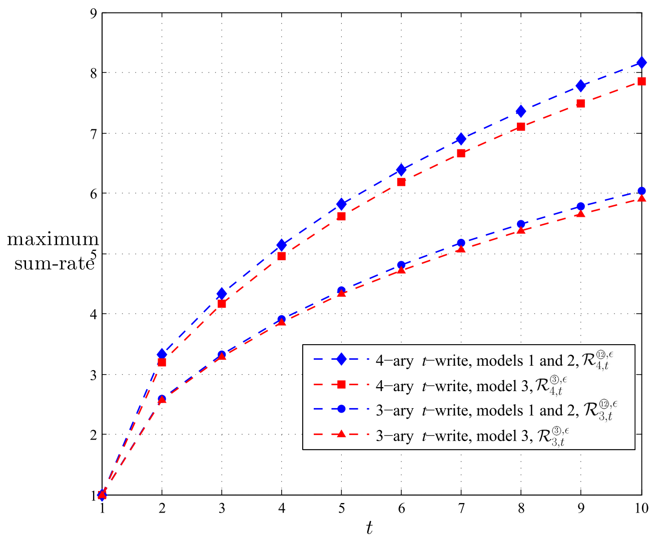

Fig. 3. A comparison between, $\mathcal{R}^{⑫}_{q,t}$, the maximum sum-rates in models 1 and 2, and $\mathcal{R}^{③,\epsilon}_{q,t}$, the maximum sum-rates in model 3, for $q \in \{3,4\}$.

Now we describe the encoding and decoding maps C'. Let $f : [q] \to \{0,1\}^{q-1}$ be the mapping, where $j \to 1^j 0^{q-1-j}$, and for $v \in [q]^n$, $f(v)$ is the concatenating the values $f(v_1), f(v_2), \ldots, f(v_n)$. Denote by $\mathcal{V}$ the image of f, i.e., $\mathcal{V} = \{0^{q-1-j}1^j : j \in [q]\}$. Note that $f : [q] \to \mathcal{V}$ is bijective, and so invertible.

Let $\mathcal{E}_i$ and $\mathcal{D}_i$ be the encoding and decoding maps of C of the i-th write, $1 \le i \le t$. The encoder on the i-th write of C' receives a message $m \in [M_i]$ to be encoded to the memory and programs the cells with the vector $f(\mathcal{E}_i(m))$, given by applying f on the result of the encoding map $\mathcal{E}_i$. We denote the cell-state vector after the i-write by $\boldsymbol{c}_i$. It is easy to verify that $m = \mathcal{D}_i(f^{-1}(\boldsymbol{c}_i))$ with probability at least $1 - p_{e_i}$. □

By Corollary 1 we have that $\mathcal{R}^{③,\epsilon}_{q,t} < \mathcal{R}^{⑫}_{q,t}$ for each $q > 2$ and $t > 1$. However, in the sequel, we state that the difference between these maximum sum-rates is bounded by a constant.

Lemma 8: For all q and t the following holds

$$\mathcal{D}_{q,t} \triangleq \mathcal{R}^{⑫}_{q,t} - \mathcal{R}^{③,\epsilon}_{q,t} \le (q-1)\log(q-1),$$

that is the difference between the maximum sum-rates of an ϵ-error q-ary t-write WOM in model 3 and a q-ary t-write WOM in models 1 and 2 is bounded by $\mathcal{D}_{q,t} \le (q-1)\log(q-1)$ for each t.

Proof: On one hand

$$\mathcal{R}^{⑫}_{q,t} = \log\binom{q-1+t}{q-1} \le (q-1)\log(t+1),$$

and on the other hand, by Lemma 7,

$$\mathcal{R}^{③}_{q,t} \ge (q-1)\log\left(\left\lfloor \frac{t}{q-1} \right\rfloor + 1\right).$$

Thus, we have

$$\begin{aligned}\mathcal{D}_{q,t} &\le (q-1)(\log(t+1) - \log(\lfloor \frac{t}{q-1} \rfloor + 1)) \\ &\le (q-1)\log(q-1).\end{aligned}$$

□

In Fig. 3, we compare between the maximum sum-rates in these models for $q = 3, 4$.

As a consequence of Lemma 7 we conclude with the following corollary, which illustrates the huge difference between the maximum sum-rates of model 4 and the others.

Corollary 2: For all $q \ge 2$, $\mathcal{R}^{④,\epsilon}_{q,t} \le (q-1) \cdot 2.37$ and therefore $\lim_{t\to\infty} \mathcal{R}^{④,\epsilon}_{q,t} \le (q-1) \cdot 2.37 < \infty$.

Proof: By Lemma 7, $\mathcal{R}^{④,\epsilon}_{q,t} \le (q-1)\mathcal{R}^{④,\epsilon}_{2,t}$. The maximum sum-rate, $\mathcal{R}^{④,\epsilon}_{2,t}$, is given by (5) [26], where for all $t \ge 1$, $\mathcal{R}^{④,\epsilon}_{2,t} \le \frac{\pi^2}{6\ln 2} \approx 2.37$, and $\lim_{t\to\infty} \mathcal{R}^{④,\epsilon}_{2,t} = \frac{\pi^2}{6\ln 2} \approx 2.37$. Thus, we have $\forall q :\ \mathcal{R}^{④}_{q,t} \le (q-1)\mathcal{R}^{④,\epsilon}_{2,t} \le (q-1) \cdot 2.37$. □

VIII. Conclusion and Open Problems

In this paper, we studied constructions and the capacity region of write-once memories. We first presented constructions of WOM codes for models 3 (EU:DI) and 4 (EU:DU). We then studied the capacity region and maximum sum-rate of non-binary WOM for all four models both for the zero-error and the ϵ-error cases. While the results in the paper expand the state of the art knowledge on write-once memories, there are still several interesting problems which are left open. Some of them are summarized as follows:

1) Calculating the zero-error capacity region and maximum sum-rate in model 3 (the EU:DI model) for all $q \ge 2$ and $t \ge 2$.
2) Calculating the zero-error and the ϵ-error capacity regions and maximum sum-rate in these two cases, in model 4 (the EU:DU model) for all $q \ge 2$ and $t \ge 2$. Note that only the ϵ-error maximum sum-rate for binary t-write WOM is known [26].
3) Find zero-error and ϵ-error WOM code constructions for models 3 and 4, both for binary and non-binary. In particular, it is left open to improve the zero-error construction from Example 1 for model 4 with sum-rate 1.29 and the zero-error construction from Example 2 for model 3 with sum-rate 1.33.

Appendix A
Non-Binary Two-Write WOM in Model 4

In order to construct non-binary WOM codes in model 4, we follow an analogy to the one from Theorem 1 which uses codes in the Z channel for the binary setup. More specifically, we consider here non-binary asymmetric errors, which can only increase the level of each cell. That is, if a q-ary cell is stored with the value $b \in [q]$ and a *q-ary asymmetric error* has occurred, then only the values $\{b+1, \ldots, q-1\}$ can be received. Note that the Z channel is a special case of this model for $q = 2$. Given two vectors $\boldsymbol{u}, \boldsymbol{v} \in [q]^n$, the *Manhattan distance* between $\boldsymbol{u}$ and $\boldsymbol{v}$, $d_M(\boldsymbol{u}, \boldsymbol{v})$, is defined to be $d_M(\boldsymbol{u}, \boldsymbol{v}) = \Sigma_{i=0}^{n-1}|u_i - v_i|$. The *Manhattan weight* of $\boldsymbol{u}$, $w_M(\boldsymbol{u})$ is defined to be $w_M(\boldsymbol{u}) = d_M(\boldsymbol{u}, \boldsymbol{0}) = \Sigma_{i=0}^{n-1} u_i$.

Let $\boldsymbol{u} \in [q]^n$ be a stored cell-state vector and $\boldsymbol{u}+\boldsymbol{e} \in [q]^n$ be the received vector where $\boldsymbol{e} \ge \boldsymbol{0}$ is the error vector. We say that a length-n code $\mathcal{C}$ over q-ary symbols with M codewords is an $(n, M, \tau)_q$ *asymmetric-error-correcting code* if it can correct any asymmetric error vector with Manhattan weight at most τ. We denote the number of q-ary length-n vectors of Manhattan weight at most τ by $B_q(n,\tau)$. Using the inclusion-exclusion principle, we conclude

$$B_q(n,\tau) = \sum_{k=0}^{\min\{n,\lfloor \frac{\tau}{q} \rfloor\}} (-1)^k \binom{n+\tau-kq}{k,\, n-k,\, \tau-kq}.$$

Note that a length-n code over q-ary symbols with M codewords and minimum Manhattan distance d is an $(n, M, \lfloor \frac{d-1}{2} \rfloor)_q$ asymmetric-error-correcting code.

The Lee metric is more investigated than the Manhattan metric, and may be used here. Given two vectors $\boldsymbol{u}, \boldsymbol{v} \in [q]^n$, the Lee distance between $\boldsymbol{u}$ and $\boldsymbol{v}$, $d_L(\boldsymbol{u}, \boldsymbol{v})$, is defined to be $d_L(\boldsymbol{u}, \boldsymbol{v}) = \Sigma_{i=0}^{n-1} \min\{|u_i - v_i|, q - |u_i - v_i|\}$. Note that for each two vectors $\boldsymbol{u}, \boldsymbol{v} \in [q]^n$, it holds that $d_L(\boldsymbol{u}, \boldsymbol{v}) \le d_M(\boldsymbol{u}, \boldsymbol{v})$. Therefore, a code $\mathcal{C}$ with minimum Lee distance d, has minimum Manhattan distance d', where $d' \ge d$.

Theorem 8: Let $\mathcal{C}$ be an $(n, M, \tau)_q$ asymmetric-error-correcting code. Then there exists an $[n, 2; M_1, M_2]_q^{④,z}$ WOM code, where $M_1 = B_q(n,\tau)$ and $M_2 = M$.

Proof: The proof will consist of describing the encoding and decoding maps of the WOM code. On the first write M_1 messages can be written by simply programming vectors with Manhattan weight at most τ.

Let $\mathcal{E}$, $\mathcal{D}$ be the encoding, decoding map of the error-correcting code $\mathcal{C}$, respectively. Then, the encoder of the second write receives a message $m \in [M_2]$ to be encoded to the memory and programs the cells by applying the encoding map $\mathcal{E}(m)$ of $\mathcal{C}$. Let $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ be the memory cell-state vectors after the first and the second write, respectively. Denote by $\boldsymbol{e}$ the error vector $\boldsymbol{c}_2 - \mathcal{E}(m)$. Note that $\boldsymbol{c}_2 = \max\{\boldsymbol{c}_1, \mathcal{E}(m)\}$, and $\boldsymbol{e} \le \boldsymbol{c}_1$ We have the following three observations:

1) $\mathcal{E}(m)$ is a codeword in $\mathcal{C}$,
2) $\boldsymbol{c}_2 \ge \mathcal{E}(m)$,
3) $d_M(\mathcal{E}(m), \boldsymbol{c}_2) = w_M(\boldsymbol{e}) \le w_M(\boldsymbol{c}_1) \le \tau$.

That is, the cell-state vector $\boldsymbol{c}_2$ is the outcome of the error vector $\boldsymbol{e}$. Since the code $\mathcal{C}$ is capable of correcting an asymmetric error vector with Manhattan weight at most τ, we have that $\mathcal{D}(\boldsymbol{c}_2) = m$, as required. □

Even though the result from Theorem 8 provides us with a specific construction of non-binary WOM codes in model 4, it is not clear what the asymptotic result of this construction is. That is, given n large enough what the best choice to choose the value of τ is. Furthermore, codes in the Manhattan distance are not easy to construct, and our attempts to find efficient codes in the Lee metric for this construction were unsuccessful. We also note that we are not limited to using this type of asymmetric non-binary error-correcting codes. We could use codes which correct limited magnitude errors [5] as long as the programmed cell-state vectors on the first write have the same limited magnitude.

Appendix B
Non-Binary Two-Write WOM in Model 3

In order to construct non-binary WOM codes in model 3, we follow an analogy to the one from Theorem 4 which uses codes in the erasure channel for the binary setup. The techniques are similar to those described in model 4, Appendix A. More specifically, we consider here non-binary erasures. That is, if a q-ary cell is stored with the value $b \in [q]$ and an *asymmetric erasure* has occurred, then a value from the set $\{b, \dots, q-1\}$ can be received with the erasure mark. Thus, if the value $a \in [q]$ is received with the erasure mark, then the correct value is in $[a+1]$. Note that the binary erasure channel is a special case of this model for $q = 2$.

Let $\boldsymbol{u} \in [q]^n$ be a stored cell-state vector and $\boldsymbol{v} \in [q]^n$ be the received vector with erasures locations $S \subseteq [n]$. Denote by $\boldsymbol{e} \in [q]^n$ the erasure vector, i.e., $\boldsymbol{e}_i = \boldsymbol{v}_i$ if $i \in S$, and otherwise $\boldsymbol{e}_i = 0$. We say that a length-n code $\mathcal{C}$ over q-ary symbols with M codewords is an $(n, M, \tau)_q$ *asymmetric-erasure-correcting code* if it can correct any asymmetric erasure vector with Manhattan weight at most τ. Note that a length-n code over q-ary symbols with M codewords and minimum Manhattan distance d is an $(n, M, d-1)_q$ asymmetric-erasure-correcting code.

Theorem 9: Let $\mathcal{C}$ be an $(n, M, \tau)_q$ asymmetric-erasure-correcting code. Then there exists an $[n, 2; M_1, M_2]_q^{③,z}$ WOM code, where $M_1 = B_q(n,\tau)$ and $M_2 = M$.

Proof: The proof will consist of describing the encoding and decoding maps of the WOM code. On the first write M_1 messages can be written by simply programming vectors with Manhattan weight at most τ. Let $\mathcal{E}, \mathcal{D}$ be the encoding, decoding map of the code $\mathcal{C}$, respectively. Then, the encoder of the second write receives a message $m \in [M_2]$ to be encoded to the memory and programs the cells by applying the encoding map $\mathcal{E}(m)$ of $\mathcal{C}$. Let $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ be the memory cell-state vector after the first, second write, respectively. We have that $\boldsymbol{c}_2 = \max\{\boldsymbol{c}_1, \mathcal{E}(m)\}$. The input to the decoder on the second write is $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$. Let S be the set of the erasures' locations, i.e., $S = \{i : c_{1,i} = c_{2,i} > 0\}$, and $\boldsymbol{e}$ be the erasure vector. We have the following four observations.

1) $\mathcal{E}(m)$ is a codeword in $\mathcal{C}$,
2) $\boldsymbol{c}_2 \ge \mathcal{E}(m)$,
3) $w_M(\boldsymbol{e}) = \sum_{i\in S} c_{2,i} = \sum_{i \in S} c_{1,i} \le w_M(\boldsymbol{c}_1) \le \tau$,
4) the set S is known to the decoder.

That is, the cell-state vector $\boldsymbol{c}_2$ is the outcome of an erasure vector with Manhattan weight at most τ where the erasures locations are known to the decoder. Since the code $\mathcal{C}$ is capable to correct an erasure vector of Manhattan weight τ, there exists exactly one codeword $\boldsymbol{c} \in \mathcal{C}$ which match to $\boldsymbol{c}_2$ on the indices in $[n] \setminus S$ and $\boldsymbol{c} \le \boldsymbol{c}_2$. Thus, we have that $\mathcal{D}(\boldsymbol{c}_2) = m$, as required. □

The disadvantages of the construction in Theorem 9 are essentially the same as described in Appendix A for model 4. Recall that a length-n code over q-ary symbols with M codewords and Manhattan distance d, is an $(n, M, \lfloor \frac{d-1}{2} \rfloor)_q$ asymmetric-error-correcting code, and an $(n, M, d-1)_q$ asymmetric-erasure-correcting code. Thus, given such a code, the construction in Theorem 9 for model 3, yields a better sum-rate than the construction from Theorem 8 for model 4. Nevertheless, we couldn't find efficient codes also for this model, by the same reasons.

ACKNOWLEDGMENT

The authors thank the anonymous reviewer and the associate editor Prof. Jun Chen for their valuable comments and suggestions, which have contributed for the clarity of the paper and its presentation.

REFERENCES

[1] L. A. Bassalygo, "New upper bounds for error-correcting codes," (in Russian), *Problemy Peredachi Informacij*, vol. 1 no. 4, pp. 41–44, 1965, (trans: Problems of Information Transmission vol. 1 no. 4, pp. 32–35).

[2] J. M. Borden, "Bounds and constructions for error correcting/detecting codes on the Z-channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Feb. 1981, pp. 94–95.

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

[4] G. Cohen, P. Godlewski, and F. Merkx, "Linear binary code for write-once memories," *IEEE Trans. Inf. Theory*, vol. 32, no. 5, pp. 697–700, Sep. 1986.

[5] Y. Cassuto, M. Schwartz, V. Bohossian, and J. Bruck, "Codes for asymmetric limited-magnitude errors with application to multilevel flash memories," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1582–1595, Apr. 2010.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. Hoboken, NJ, USA: Wiley, 1991.

[7] B. K. Dey, S. Jaggi, M. Langberg, and A. D. Sarwate, "Upper bounds on the capacity of binary channels with causal adversaries," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3753–3763, Jun. 2013.

[8] F. W. Fu and A. J. H. Vinck, "On the capacity of generalized write-once memory with state transitions described by an arbitrary directed acyclic graph," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 308–313, Jan. 1999.

[9] E. N. Gilbert, "A comparison of signalling alphabets," *Bell Syst. Tech. J.*, vol. 31, no. 3, pp. 504–522, 1952.

[10] R. Gabrys and L. Dolecek, "Constructions of nonbinary WOM codes for multilevel flash memories," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1905–1919, Apr. 2015.

[11] R. Gabrys, E. Yaakobi, L. Dolecek, P. H. Siegel, A. Vardy, and J. K. Wolf, "Non-binary WOM-codes for multilevel flash memories," in *Proc. IEEE Inf. Theory Workshop*, Paraty, Brazil, Oct. 2011, pp. 40–44.

[12] C. Helgesen, "Asymmetric error-correcting codes with minimal block length," unpublished manuscript, Bergen, 1984.

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

[14] M. Horovitz and E. Yaakobi, "WOM codes with uninformed encoder," in *Proc. IEEE Inf. Theory Workshop*, Jerusalem, Israel, Apr./May 2015, pp. 1–5.

[15] M. Horovitz and E. Yaakobi, "On the capacity of non-binary write-once memory," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 945–949.

[16] R. McEliece, E. Rodemich, H. Rumsey, and L. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities," *IEEE Trans. Inf. Theory*, vol. 23, no. 2, pp. 157–166, Mar. 1997.

[17] J. H. van Lint, *Introduction to Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[18] T. R. N. Rao and A. S. Chawla, "Asymmetric error codes for some LSI semiconductor memories," in *Proc. Ann. Southeastern Symp. Syst. Theory*, 1975, pp. 170–171.

[19] R. L. Rivest and A. Shamir, "How to reuse a write-once memory," *Inform. Control.*, vol. 55, nos. 1–3, pp. 1–19, Oct./Dec. 1982.

[20] E. Sharon and I. Alrod, "Coding scheme for optimizing random I/O performance," in *Non-Volatile Memories Workshop*, San Diego, CA, USA, Mar. 2013.

[21] A. Shpilka, "Capacity-achieving multiwrite WOM codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1481–1487, Mar. 2014.

[22] L. G. Tallini, S. Al-Bassam, and B. Bose, "On the capacity and codes for the Z-channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Lausanne, Switzerland, Jul. 2002, p. 422.

[23] R. R. Varshamov, "Estimate of the number of signals in error correcting codes," *Doklady Akademii Nauk SSSR*, vol. 117, no. 5, pp. 739–741, 1957.

[24] S. Verdu, "Channel capacity," in *The Electrical Engineering Handbook*. Boca Raton, FL, USA: CRC Press, 1997, ch. 73.5, pp. 1671–1678.

[25] L. Wang and Y. H. Kim, "Sum-capacity of multiple-write noisy memory," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Petersburg, Russia, Jul./Aug. 2011, pp. 2527–2531.

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

[27] E. Yaakobi, S. Kayser, P. H. Siegel, A. Vardy, and J. K. Wolf, "Codes for write-once memories," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5985–5999, Sep. 2012.

[28] E. Yaakobi and R. Motwani, "Construction of random input-output codes with moderate block lengths," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1819–1828, May 2016.

Michal Horovitz was born in Israel in 1987. She received the B.Sc. degree from the Open University of Israel, Ra'anana, Israel, in 2009, from the department of Mathematics and from the department of Computer Science. She is currently a Ph.D. student in the Computer Science Department at the Technion—Israel Institute of Technology, Haifa, Israel. Her research interests include coding theory with applications to non-volatile memories, information theory, and combinatorics.

Eitan Yaakobi (S'07–M'12) is an Assistant Professor at the Computer Science Department at the Technion—Israel Institute of Technology. He received the B.A. degrees in computer science and mathematics, and the M.Sc. degree in computer science from the Technion—Israel Institute of Technology, Haifa, Israel, in 2005 and 2007, respectively, and the Ph.D. degree in electrical engineering from the University of California, San Diego, in 2011. Between 2011-2013, he was a postdoctoral researcher in the department of Electrical Engineering at the California Institute of Technology. His research interests include information and coding theory with applications to nonvolatile memories, associative memories, data storage and retrieval, and voting theory. He received the Marconi Society Young Scholar in 2009 and the Intel Ph.D. Fellowship in 2010-2011.