Endurance-Limited Memories: Capacity and Codes

Yeow Meng Chee, Senior Member, IEEE, Michal Horovitz, Member, IEEE, Alexander Vardy, Fellow, IEEE, Van Khu Vu, and Eitan Yaakobi, Senior Member, IEEE

Abstract—Resistive memories, such as phase change memories and resistive random access memories have attracted significant attention in recent years due to their better scalability, speed, rewritability, and yet non-volatility. However, their limited endurance is still a major drawback that has to be improved before they can be widely adapted in large-scale systems. In this work, in order to reduce the wear out of the cells, we propose a new coding scheme, called endurance-limited memories (ELM) codes, that increases the endurance of these memories by limiting the number of cell programming operations. Namely, an \( \ell \)-change \( t \)-write ELM code is a coding scheme that allows to write \( t \) messages into some \( n \) binary cells while guaranteeing that each cell is programmed at most \( \ell \) times. In case \( \ell = 1 \), these codes coincide with the well-studied write-once memory (WOM) codes. We study some models of these codes which depend upon whether the encoder knows on each write the number of times each cell was programmed, knows only the memory state, or even does not know anything. For the decoder, we consider these similar three cases. We fully characterize the capacity regions and the maximum sum-rates of three models where the encoder knows on each write the number of times each cell was programmed. In particular, it is shown that in these models the maximum sum-rate is \( \log \sum_{i=0}^{\ell} \binom{n}{i} \). We also study and expose the capacity regions of the models where the decoder is informed with the number of times each cell was programmed. Finally we present the most practical model where the encoder read the memory before encoding new data and the decoder has no information about the previous states of the memory.

Index Terms—Endurance limited memories (ELM), rewriting codes, resistive memories.

I. INTRODUCTION

Emerging resistive memory technologies, such as resistive random access memories (ReRAM) and phase-change memories (PCM), have the potential to be the future’s universal memories. They combine several important attributes starting from the speed of SRAM, the density of DRAM, and the non-volatility of flash memories. However, they fall short in their write endurance, which significantly increases their bit error rate (BER). Hence, solving the limited endurance of these memories is crucial before they can be widely adapted in large-scale systems [9], [18], [28].

Resistive memories are non-volatile memories which are composed of cells. The information is stored in the cells by changing their resistance. They combine the following two properties of DRAM and flash memories. Similarly to flash memories and unlike DRAM they are non-volatile memories and thus they don’t require refresh operations. Furthermore, like DRAM and unlike flash memories they are rewritable without an erase operation. The main challenge that has remained to be solved in order to make these memories a legitimate candidate as a universal memory is their limited write endurance, which is the goal of this paper.

Endurance is defined as the number of set/reset cycles to switch the state of cells in ReRAM while it is still reliable. Owing to its importance, there are many researches which test and characterize the endurance of ReRAM in order to show a strong dependence on material of cells, cell size, [16], [26], and program operation [22]. To improve the endurance of ReRAM, recent research have focused on the structure and material of devices [16], [25] and programming schemes [6], [22]. In this work, we present a scheme to use rewriting code to improve the endurance lifetime of ReRAM.

Previous works have offered different solutions to combat the write endurance of resistive memories. In [12], the authors proposed to use locally repairable codes (LRC) in order to construct codes with small rewriting locality in order to mitigate both the problems of endurance and power consumption. In [27], the authors proposed mellow writes, a technique which is targeted to reduce the wear-out of the writes rather than reducing the number of writes. Lastly, several other works proposed coding schemes which correct stuck-at cells; see e.g. [14], [19], [24].

In order to combat the limited write endurance in resistive memories, this paper proposes to study the following new family of codes, called endurance-limited memory (ELM) codes. Assume there are \( n \) binary cells and \( t \) messages that are required to be stored in these cells sequentially. Assume also that each cell can be programmed at most \( \ell \geq 1 \) times. Then, we seek to find the set of achievable rates, i.e., the capacity region, and design code constructions for this model. Note that for \( \ell = 1 \), we get the classical problem of write-
once memory (WOM) codes [2], [10], [11], [17], [21], [23]. Besides that, if $t \leq \ell$, the coding scheme is trivial. Hence, in this work, we only focus on the cases where $t > \ell > 1$. Note that a trivial lower bound on the maximum sum-rate is $\min\{t, \ell\}$, which achieved by writing with rate 1 in the first $\min\{t, \ell\}$ writes and with rate 0 in the remaining writes.

Let us consider first the case where $\ell = 2$ and $t = 3$. A naive solution is to use a two-write WOM code for the first two writes and then write $n$ more bits on the third write. The maximum sum-rate using this solution will be $\log(3) + 1 = \log(3) + \log(3) + \log(3) + 1 = \log(3)^2 + 1$. The intuition behind this is as follows. Let $p_1$ be the probability to program a cell on the first write, so we assume that $p_1$ cells are programmed. Then, on the second and third writes we have a two-write WOM code problem for the $p_1$ programmed cells, and for the $(1 - p_1)n$ non-programmed cells, we can write twice on them so no coding is needed. The maximum sum-rate is achieved for $p_1 = 3/7$. However, it is still a challenging task to design specific code constructions that can approach sum-rate of $\log(3)$.

There are several models of ELM codes which can be studied. These models are distinguishable by the information that is available to the encoder and the decoder. In particular, for the encoder, we consider three cases which depend upon whether the encoder knows the number of times each cell was programmed, encoder informed all (EIA), only the current state of the cell, encoder informed partially (EIP), or no information about the cells state, encoder uninformed (EU). The decoder will also have three cases, corresponding to the same information that is available to the encoder. Thus, by considering all combinations of the above three cases for the encoder and the decoder, it is possible to define and study nine models, EX : DY, where $X \in \{IA, IP, U\}$.

For a positive integer $a$, the set $\{0, \ldots, a - 1\}$ is defined by $[a]$. Throughout this paper, we assume that the number of cells is $n$. We use the vector notation $c \in [2]^n$ to represent the cell-state vector of the $n$ memory cells, and the vector $v \in [\ell + 1]^n$, which will be called the cell-program-count vector, to represent the number of times each cell was programmed. Note that the state of a cell is the parity of the number of times it was programmed. Thus, if the encoder (or the decoder) knows the cell-program-count vector $v$, in particular it knows the cell-state vector $c$ as well. For a vector $v \in [\ell + 1]^n$, we denote by $(v)_2$ the length-$n$ binary vector which satisfies $(v)_2[k] = v_k (\bmod 2)$ for all $k \in [n]$, and we say that $(v)_2$ equals to $v$ modulo 2. The complement of a binary vector $c$ is denoted by $\overline{c}$. The all ones, zeros vector will be denoted by 1, 0, respectively. For two length-$n$ vectors $a$ and $b$, $a \oplus b$ is the vector obtained by pointwise addition. If $a$ and $b$ are binary vectors $a \oplus b$ is the vector obtained by pointwise addition modulo 2.

For a cell-program-count vector $v \in [\ell + 1]^n$ and a new cell-state vector $c \in [2]^n$ to be programmed to the cells, we define by $N(v, c) \in [\ell + 1]^n$ and $f(v, c) \in [2]^n$ the result of programming the new cell-state vector $c$. That is, $N(v, c)$ is the new cell-program-count vector after programming $c$, $f(v, c)$ is the new cell-state vector, and they are formally defined as follows. $N(v, c)_k = v_k$ if $c_k = v_k (\bmod 2)$, and otherwise $N(v, c)_k = \min\{\ell, v_k + 1\}$, where the index $k$ for a vector means its $k$-th element. Similarly, $f(v, c)_k = c_k$ if $v_k < \ell$, and otherwise $f(v, c)_k = v_k (\bmod 2)$. Note that $(N(v, c))_2 = f(v, c)$, i.e., $f(v, c)$ equals to $N(v, c)$ modulo 2. We are ready now to define all models studied in this paper.

Definition 1: An $[n, t, \ell; M_1, \ldots, M_e]_{EX:DY,p_e}$-change t-write Endurance-Limited Memory (ELM) code with error-probability vector $p_e = (p_{e_1}, \ldots, p_{e_n})$, where $X, Y \in \{IA, IP, U\}$, is a coding scheme comprising of $n$ binary cells and is defined by $t$ encoding and decoding maps $(E_j, D_j)$ for $1 \leq j \leq t$. For the map $E_j, Im(E_j)$ is its image, where by definition $Im(E_0) = \{(0, \ldots, 0)\}$.

Furthermore, for $j \in [t + 1]$, let $N_j$ and $Im^*(E_j)$ be the sets of all state-program-count vectors, cell-state vectors which can be obtained after the first $j$ writes, respectively. Formally, for $1 \leq j$, $N_j = \{N(v, c) : c \in Im(E_j), v \in N_{j-1}\}$, where $N_0 = \{(0, \ldots, 0)\}$, and $Im^*(E_j) = \{w_2 : w \in N_j\}$. Note that for the EIA models $Im^*(E_j) = Im(E_j)$. The domain and the range of the encoding maps are defined as follows:

1. For the EIA models, $E_j : [M_j] \times N_{j-1} \rightarrow [2]^n$ such that for all $(m, v) \in [M_j] \times N_{j-1}$ it holds that $v + ((v)_2 \oplus E_j(m, v)) \in [\ell + 1]^n$.
2. For the EIP models, $E_j : [M_j] \times Im^*(E_{j-1}) \rightarrow [2]^n$.

For a message $m$ we denote by $I_m(x)$ the indicator function, where $I_m(x) = 0$ if $m = x$, and otherwise $I_m(x) = 1$. Additionally, for a message $m \in [M_j]$, $Pr(m)$ is the probability of programming a message $m$ on the $j$-th write, and for $v \in N_{j-1}$, $Pr(v)$ is the probability to have cell-program-count vector $v$ before the $j$-th write. The nine models are defined as follows. For all $1 \leq j \leq t$,
if \((X, Y) = (IA, IA)\) then
\[
D_j : \{(E_j(m, v), (m, v)) : m \in [M_j], v \in N_{j-1} \} \rightarrow [M_j],
\]
and
\[
\sum_{(m, v) \in [M_j] \times N_{j-1}} Pr(m)Pr(v) \cdot I_m(D_j(E_j(m, v), (m, v))) \leq p_{e_j},
\]
(2) if \((X, Y) = (IA, IP)\) then
\[
D_j : \{(E_j(m, v), (m, v)) : m \in [M_j], v \in N_{j-1} \} \rightarrow [M_j],
\]
and
\[
\sum_{(m, v) \in [M_j] \times N_{j-1}} Pr(m)Pr(v) \cdot I_m(D_j(E_j(m, v), (m, v))) \leq p_{e_j},
\]
if for all \(1 \leq j \leq t\), then the code is called a zero-error ELM code and is denoted by \([n, t, \ell; M_1, \ldots, M_t]^{EX:DY,z}\).

The rate on the \(j\)-th write of an \([n, t, \ell; M_1, \ldots, M_t]^{EX:DY,p}\) ELM code, \(X, Y \in \{IA, IP, U\}\), is defined as \(R_j = \log M_j\), and the sum-rate is the sum of the individual rates on all writes, \(R_{sum} = \sum_{j=1}^{t} R_j\). A rate tuple \(R = (R_1, \ldots, R_t)\) is called \(\epsilon\)-error achievable in model \(EX:Y, D\), if for all \(\epsilon > 0\) there exists an \([n, t, \ell; M_1, \ldots, M_t]^{EX:DY,p}\) ELM code with error-probability vector \(p_e = (p_{e_1}, \ldots, p_{e_t}) < (\epsilon, \ldots, \epsilon)\) such that \(\log M_j \geq R_j - \epsilon\). The rate tuple \(R\) will be called zero-error achievable if for all \(1 \leq j \leq t\), \(p_{e_j} = 0\). The \(\epsilon\)-error capacity region of the \(EX:Y, D\) model is the set of all \(\epsilon\)-error achievable rate tuples, that is,
\[
C_{t, \ell, \epsilon}^{EX:DY,z} = \{(R_1, \ldots, R_t) | (R_1, \ldots, R_t) \text{ is } \epsilon\text{-error achievable} \}
\]
and the \(\epsilon\)-error maximum sum-rate will be denoted by \(R_{t, \ell, \epsilon}^{EX:DY,z}\). The zero-error capacity region \(C_{t, \ell}^{EX:DY,z}\) and the zero-error maximum sum-rate \(R_{t, \ell}^{EX:DY,z}\) are defined similarly. We say that \(R \lessgtr R'\) for \(R = (R_1, \ldots, R_t)\) and \(R' = (R'_1, \ldots, R'_t)\) if \(R_j \leq R'_j\) for all \(1 \leq j \leq t\) and \(R < R'\) if \(R \lessgtr R'\) and \(R \neq R'\).

According to these definitions it is easy to verify the following relations. For \(g \in \{z, \epsilon\}\) and \(X, Y \in \{IA, IP, U\}\) it holds that
\[
C_{t, \ell, \epsilon}^{EU:Y, D} = C_{t, \ell}^{EIP:Y, D} \subseteq C_{t, \ell}^{EIA:Y, D},
\]
and
\[
C_{t, \ell}^{EX:Y, D} \subseteq C_{t, \ell}^{EX:Y, D, t}, \quad C_{t, \ell}^{EX:Y, D} \subseteq C_{t, \ell}^{EX:Y, D, t},
\]
Similar connections hold for the maximum sum-rates.

Note that if \(\ell \geq t\) then all problems are trivial since it is possible to program all cells on each write, so the capacity region in all models is \([0, 1]^t\) and the maximum sum-rate is \(t\). For \(\ell = 1\), we get the classical and well-studied WOM codes [2], [10], [11], [17], [21], [23]. In this case, we also notice that the IA and IP models are the same for both the encoder and the decoder. The capacity region and the maximum sum-rate in most of these cases are known; see e.g. [10], [11], [17], [23]. In the rest of this paper, and unless stated otherwise, we assume that \(1 \leq \ell < t\).

A. Related Work

The EIA models of ELM codes studied in this paper are strongly related to non-binary WOM codes and their modified versions studied in [3], [8], [11], [13]. In these EIA models, we can treat every cell as an \((\ell + 1)\)-ary cell, where it is only possible to increase its level by one on each write, while its maximum level is \(\ell\). In non-binary WOM codes,
each cell has $q$ levels and its level can not be decreased \[8, 11\]. In a write $\ell$-step-up memory \[3\], a special version of the non-binary WOM code, each cell has $q$ levels and each time we write a cell, its level can be increased by at most some value $\ell$. Recently, Kobayashi et al. \[13\] also studied a modified version of non-binary WOM codes, called write constrained memories, where there is a cost on each state transition. Yet, these codes are not identical and we can not apply previous results to solve the models of ELM codes. In fact, our results on EIA models are useful to obtain an explicit formula of the capacity region of write $\ell$-step-up memory codes. We also note that some models of ELM codes, such as the EIP:DU model, are much different from previous models and are difficult to solve.

Moreover, our proposed ELM coding scheme is also related to the cooling code which is used to control the peak temperature of an interconnect \[1\], \[2\]. In \[1\], cooling codes are proposed to avoid all hottest wires and in \[2\], cooling codes are shown to be equivalent to two-write binary WOM codes. Later in this work, we will use cooling codes as two-write binary WOM codes in order to construct our ELM codes. Now, we discuss the ability of our ELM codes to control the peak temperature of an interconnection. In fact, using our coding scheme in this work, we can control the maximal number of switches in each wire. We note that the temperature of each wire is closely related to the number of switches in each wire. And thus, we can control the peak temperature of each wire.

Recently, the EIA:DIA model of ELM is shown to be useful for two-dimensional (2D) weight-constrained code scheme \[15\]. In a 2D weight-constrained code, each codeword is an array of size $m \times n$ where the number of 1 symbols in each row and each column is limited by $pm$ and $qm$, respectively. The encoder and decoder know the values of all the $mn$ bits. We can view this 2D weight-constrained code as we write $m$ messages in ReRAM where each cell can be switched at most $qm$ times and both encoder and decoder know all previous messages.

### B. Our Contribution

In this work, we propose a novel scheme of rewriting code to improve the endurance lifetime of ReRAM, called endurance limited memories (ELM) code. In $\ell$-change $t$-write ELM codes, each cell can be programmed at most $\ell$ times during the writing of $t$ messages. In the case that $\ell$ is much smaller than $t$, we can significantly improve the lifetime of the memories. Depending upon whether the encoder (E) and decoder (D) know the number of times each cell is programmed (IA), the current state of each cell (IP), or have no information on the state of each cell (U), we present and investigate all nine models of ELM codes. We then present several constructions of ELM codes for the EIA models. To construct these codes, we need some new ideas even though we also use several special families of WOM codes as components of our ELM codes. In Section IV, we use a known technique in information theory to obtain the capacity region of the EIP:DIA model. The capacity region of all EIA models is compared to the EIP:DU model in Section V. Using the same technique carefully, we can also find the capacity region of the EU:DIA model in Section VI. Finally, in Section VII, we study the most practical model for ReRAM, the EIP:DU model. Although several good bounds on the maximum sum-rate are presented, these bounds are not tight and finding an exact formula of the maximum sum-rate of the EIP:DU model is still an open problem. To achieve some good constructive lower bounds, we provide several constructions of EIP:DU ELM codes for the zero-error case. These results are novel and we need different original methods to achieve them.

### III. The EIA Models

In this section, we explore the capacity region and the maximum sum-rate of the EIA models for both the $\epsilon$-error and the zero-error cases. We also propose capacity achieving codes for these cases.

For each $j \in \{t + 1\}$, we let $c_j$ denote the binary length-$n$ vector which represents the cell-state vector after the $j$-th write, where $c_0 = 0$. Recall that in the EIA models, on the $j$-th write the encoder knows the number of times each cell was programmed before the current write. That is, the encoder receives as an input a length-$n$ cell-program-count vector $N(c_{j-1}) \in \{t + 1\}^n$ that represents the number of times each cell was programmed so far. Next, for all $t$ and $\ell$, we define the region $C_{t, \ell}$ and in Theorem 2, we prove that this is the capacity region of all the EIA models.

For $1 \leq j \leq t$ and $i \in \{\ell + 1\}$, $i \leq j$, let $p_{j,i} \in [0, 0.5]$, be the probability to program a cell on the $j$-th write, given that this cell has been already programmed $i$ times. We define $p_{j,0} = p_{j,\ell} = 0$ for $1 \leq j \leq t$, and let $Q_{j,i}$ be the probability that a cell has been programmed exactly $i$ times on the first $j$ writes. Formally, $Q_{j,i}$ is defined recursively by using $p_{j,i}$ and $p_{j,i-1}$ as follows:

\[
Q_{j,i} = \begin{cases} 
Q_{j-1,i}(1 - p_{j,i}) + Q_{j-1,i-1}p_{j,i-1}, & \text{if } i > 0, \\
Q_{j-1,i}(1 - p_{j,i}), & \text{if } i = 0,
\end{cases}
\]

where for $j = 0$ we set $Q_{0,0} = 1$ otherwise $Q_{0,i} = 0$. The rates region $C_{t, \ell}$ is defined as follows:

\[
C_{t, \ell} = \left\{ \left( R_1, \ldots, R_t \right) | \forall 1 \leq j \leq t : R_j \leq \sum_{i=0}^{\min(\ell,j) - 1} Q_{j-1,i}h(p_{j,i}) \right\},
\]

where $h(x) = \frac{x}{1-x}$ for $0 < x < 1$. The rates $R_j$ are defined as follows:

\[
R_j = \left\{ \left( R_1, \ldots, R_t \right) | \forall 1 \leq j \leq t : R_j \leq \sum_{i=0}^{\min(\ell,j) - 1} Q_{j-1,i}h(p_{j,i}) \right\}.
\]
where in this paper both \( h(x) \) and \( H(X) \) represent the binary entropy function for \( 0 \leq x \leq 1 \) and a random variable \( X \), respectively.

Note that for \( \ell = 1 \), it is possible to verify that we get the capacity region of WOM [10], [17], [23]. It is also readily verified that the maximum sum-rate is achieved with \( p_{ij} = 0.5 \) for all \( t - j + 1 \geq \ell - i \), since \( t - j + 1 \) is the number of the remaining writes, and \( \ell - i \) is the number of times the cell can be programmed. Thus, if \( t - j + 1 \neq \ell - i \) then we can program the cell with probability 0.5 to obtain the maximum rate. The next theorem proves that for \( 2 \leq \ell \leq t - 1 \), \( \mathcal{C}_{t,\ell} \) is the capacity region of the \( \ell \)-change t-write ELM in all EIA models, and thus we denote this capacity by \( \mathcal{C}_{t,\ell}^{EIA} \), and the maximum sum-rate by \( R_{t,\ell}^{EIA} \).

**Theorem 2:** The rates region \( \mathcal{C}_{t,\ell} \) is the capacity region of the \( \ell \)-change t-write ELM in all EIA models for both \( \epsilon \)-error and zero-error cases. That is, for all \( g \in \{ z, \epsilon \} \) and \( Y \in \{ IA, IP, U \}, C_{t,\ell}^{EIA:DU,z} \subseteq C_{t,\ell} \subseteq C_{t,\ell}^{EIA:DI,z} \), and \( C_{t,\ell}^{EIA:DU,\epsilon} \subseteq C_{t,\ell}^{EIA:DI,\epsilon} \).

Proof: Recall that by the definitions of the models
\[
C_{t,\ell}^{EIA:DU,z} \subseteq C_{t,\ell}^{EIA:DI,z} \subseteq C_{t,\ell} \subseteq C_{t,\ell}^{EIA:DI,\epsilon},
\]
and
\[
C_{t,\ell}^{EIA:DU,\epsilon} \subseteq C_{t,\ell}^{EIA:DI,\epsilon} \subseteq C_{t,\ell}.
\]

The rest of the proof consists of two parts. The first part, called the direct part, proves that \( \mathcal{C}_{t,\ell} \subseteq C_{t,\ell}^{EIA:DU,\epsilon} \), and in the second, called the converse part, we prove that \( C_{t,\ell}^{EIA:DU,\epsilon} \subseteq \mathcal{C}_{t,\ell} \). The direct part is proved in Subsection III-A for the zero-error case of the EIA:DU model, while the converse part is proved in Subsection III-B for the \( \epsilon \)-error case of the EIA:DI model.

Next, we seek to present the capacity region of the EIA models in a recursive form. While we see this representation of the capacity region more intuitive, it will also help us in finding the maximum sum-rate of this model. For all \( t \geq 1 \) and \( \ell \geq 1 \), let \( \mathcal{C}_{t,\ell} \) be the following region which is defined recursively as follows. For \( t > \ell \geq 1 \)
\[
\mathcal{C}_{t,\ell} = \{(R_1, \ldots, R_t) | R_1 \leq h(p), p \in [0, 0.5], \text{ for } 2 \leq j \leq t, R_j \leq pR_j' + (1 - p)R_j'', (R_2', \ldots, R_t') \in \mathcal{C}_{t-1,\ell-1} \text{ and } (R_2'', \ldots, R_t'') \in \mathcal{C}_{t-1,\ell-1} \},
\]
where for all \( \ell \geq 1 \) we set \( \mathcal{C}_{t,0} = \{0, 1\}^t \) and \( \mathcal{C}_{0,0} = \{0\} \).

**Theorem 3:** For all \( t \) and \( \ell \), \( \mathcal{C}_{t,\ell} = \mathcal{C}_{t,\ell} \).

Proof: For the first direction, we prove by induction on \( t \) that for all \( \ell \geq 1 \), if \( \mathbf{R} = (R_1, \ldots, R_t) \in \mathcal{C}_{t,\ell} \), then \( \mathbf{R} \in C_{t,\ell}^{EIA:DI,\epsilon} \). Since \( C_{t,\ell}^{EIA:DI,\epsilon} = \mathcal{C}_{t,\ell} \), we conclude that \( \mathcal{C}_{t,\ell} \subseteq \mathcal{C}_{t,\ell} \).

The base of the induction is \( t = \ell \) for all \( \ell \geq 1 \). These cases are readily verified. For the induction step, let \( \mathbf{R} = (R_1, \ldots, R_t) \in \mathcal{C}_{t,\ell}, 1 \leq \ell < t \), such that \( R_1 \leq h(p) \) for \( p \in [0, 0.5] \) and for \( 2 \leq j \leq t \) \( R_j \leq pR_j' + (1 - p)R_j'' \) where \( (R_2', \ldots, R_t') \in \mathcal{C}_{t-1,\ell-1} \) and \( (R_2'', \ldots, R_t'') \in \mathcal{C}_{t-1,\ell-1} \). By the induction hypothesis, \( (R_2', \ldots, R_t') \in C_{t-1,\ell-1}^{EIA:DI,\epsilon} \) and \( (R_2'', \ldots, R_t'') \in C_{t-1,\ell-1}^{EIA:DI,\epsilon} \). Thus, we have two codes: \( C_1 \) - an \((\ell - 1)\)-change \((t - 1)\)-write ELM code which achieves the rate tuple \( (R_2', \ldots, R_t') \) and \( C_2 \) - an \( \ell \)-change \((t - 1)\)-write ELM code which achieves the rate tuple \( (R_2'', \ldots, R_t'') \). Then, we can design an \( \ell \)-change \( t \)-write ELM code, such that on the first write the encoder programs a cell with probability \( p \) for \( p \in [0, 0.5] \), and then on the next writes it applies \( C_1 \) for the cells that were programmed on the first write, and \( C_2 \) for the other cells. Thus, the rate tuple \( \mathbf{R} \) is achieved.

The second direction, \( \mathcal{C}_{t,\ell} \subseteq \mathcal{C}_{t,\ell} \), is proved by induction for all \( t \geq 1 \) we prove that \( \mathcal{C}_{t,\ell} \subseteq \mathcal{C}_{t,\ell} \) for all \( 1 \leq \ell \leq t \). The base of the induction, \( t = 1 \) and \( \ell = 1 \), is trivial. The induction assumption is that for each \( 1 \leq \ell' \leq t - 1 \), \( \mathcal{C}_{t-1,\ell'} \subseteq \mathcal{C}_{t-1,\ell'} \). For the induction step, let \( \mathbf{R} = (R_1, \ldots, R_t) \in \mathcal{C}_{t,\ell} \). By the rate tuple which is attained by the probabilities \( p_{i,j} \). Denote by \( \mathbf{R}' = (R_2', \ldots, R_t') \in \mathcal{C}_{t-1,\ell-1} \) the rate tuple which is attained by the probabilities \( p_{i,j} \). Recall that we define \( \mathcal{C}_{t-1,\ell-1} \subseteq \mathcal{C}_{t-1,\ell-1} \), and \( C_{t-1} = \{0\} \). It can be easily verified that for all \( j, 2 \leq j \leq t, R_j = p_{i,j}R_j' + (1 - p_{i,j})R_j'' \). By the induction hypothesis, \( \mathbf{R}' \in \mathcal{C}_{t-1,\ell-1} \) and \( \mathbf{R}'' \in \mathcal{C}_{t-1,\ell-1} \). Thus, by defining \( p = p_{i,0} \) we get a recursive form for \( \mathbf{R} \), and we can conclude that \( \mathcal{C}_{t,\ell} \subseteq \mathcal{C}_{t,\ell} \).

Next, using the result from Theorem 3, it is possible to find the maximum sum-rate of the EIA models, \( R_{t,\ell}^{EIA} \).

**Theorem 4:** For all \( t \) and \( \ell \),
\[
R_{t,\ell}^{EIA} = \log \sum_{i=0}^{t} \binom{t}{i},
\]
and this value is achieved for
\[
p_{1,0} = \frac{\sum_{i=0}^{t-1} \binom{t-1}{i}}{\sum_{i=0}^{t} \binom{t}{i}},
\]
where \( p_{1,0} \) are defined in Equations (2), (3) in \( \mathcal{C}_{t,\ell} \).
all \( 2 \leq \ell \leq t - 1 \),
\[
\log X_{t,\ell} = h(p) + p \log X_{t-1,\ell-1} + (1-p) \log X_{t-1,\ell}.
\]
This relation holds since
\[
h(p) + p \log X_{t-1,\ell-1} + (1-p) \log X_{t-1,\ell} = p \left( \log \left( \frac{X_{t,\ell}}{X_{t-1,\ell-1}} \right) + \log(X_{t-1,\ell}) \right) + (1-p) \left( \log \left( \frac{X_{t,\ell}}{X_{t-1,\ell}} \right) + \log(X_{t-1,\ell}) \right) = p \log X_{t,\ell} + (1-p) \log X_{t,\ell} = \log(X_{t,\ell}).
\]

### A. The EIA:DU Model - Constructions and Direct Part of Theorem 2

In this subsection, we study the EIA:DU model, that is, encoder informed all and decoder uninformed. Our main contribution is a construction of a capacity-achieving \( t \)-write EIA:DU-ELM code for the zero-error case, which assures that \( C_{t,\ell} \subseteq C_{t,\ell}^{EIA:DU,z} \). That is, the direct part of Theorem 2 is proved.

Let us start with the first non-trivial case of \( t = 3 \) and \( \ell = 2 \). Thus, we want to prove that \( C_{3,2} \subseteq C_{3,2}^{EIA:DU,z} \). Recall that,
\[
C_{3,2} = \left\{ (R_1, R_2, R_3) | R_1 \leq h(p_{1,0}), R_2 \leq 1-p_{1,0} + p_{1,0}h(p_{2,1}), R_3 \leq 1-p_{1,0}p_{2,1} \right\}.
\]
which is achieved by setting \( p_{3,0} = p_{3,1} = p_{2,0} = 0.5 \) in Equation (2). The next theorem states the existence of a construction of ELM codes for this case.

**Theorem 5:** For any \( \epsilon > 0 \) and \( p_{1,0}, p_{2,0}, p_{2,1} \in [0,0.5] \), there exists an explicit construction of a zero-error two-change three-write EIA:DU-ELM code satisfying \( R_1 \geq h(p_{1,0}) - \epsilon \), \( R_2 \geq (1-p_{1,0})h(p_{2,1}) + p_{1,0}h(p_{2,1}) - \epsilon \), and \( R_3 \geq (1-p_{1,0}p_{2,1}) - \epsilon \).

Before presenting our construction for two-change three-write EIA:DU-ELM codes, we introduce the following family of WOM codes. We then use these WOM codes as component codes in our construction of EIA:DU-ELM codes. Note that the WOM codes we use for our construction are given for \( n \to \infty \), and thus our constructions for ELM codes use such \( n \).

**Definition 6:** An \( \{n,2;M_1,M_2\}^{EIA:DU,z} \)-two-write \( q \)-ary EI:DU WOM code for the zero-error case is a coding scheme comprising of \( n \) \( q \)-ary cells. It consists of two pairs of encoding and decoding maps \( (E_{q,1},D_{q,1}) \) and \( (E_{q,2},D_{q,2}) \) which are defined as follows:

1. \( E_{q,1} : \{M_1\} \to [q]^n \) and \( D_{q,1} : Im(E_{q,1}) \to [M_1] \) such that for all \( m_1 \in [M_1] \), \( D_{q,1}(E_{q,1}(m_1)) = m_1 \).
2. \( E_{q,2} : \{M_2\} \times Im(E_{q,1}) \to [q]^n \) and \( D_{q,2} : Im(E_{q,2}) \to [M_2] \) such that for all \( (m_2,e) \in [M_2] \times Im(E_{q,1}) \), \( E_{q,2}(m_2,e) \geq c \) and \( D_{q,2}(E_{q,2}(m_2,e)) = m_2 \).

We say that \( p = (p_0,p_1,\ldots,p_{m-1}) \) is a probability vector if \( \sum_{i=0}^{m-1} p_i = 1 \) and \( p_i \geq 0 \) for all \( i \in [m] \). We distinguish between an error-probability vector that is used in Definition 1, and a probability vector. An error-probability vector is a vector of error-probabilities, and not a probability vector, i.e., the sum of the elements of an error-probability vector does not need be 1. For two positive integers \( n, q \) and a probability vector \( p = (p_0,p_1,\ldots,p_{q-1}) \), we denote by \( B(n,p) \) the set of all length-\( n \) \( q \)-ary vectors of constant composition \( w = (w_0,\ldots,w_{q-1}) \), where \( w_i = p_i n \) for \( i \in [q] \). Let \( p_1,\ldots,k \) be the probability that on the \( j \)-th write, a cell in state \( i \) is programmed to state \( k, k \geq i \).

A family of two-write \( q \)-ary capacity-achieving EI:DU WOM codes was constructed recently by Shpilka [21]. Particularly, given \( \epsilon > 0 \) and probability vectors \( p_{1,0}, p_{2,0}, p_{2,1} \), Shpilka [21] constructed a family of two-write \( q \)-ary EI:DU WOM codes that match these probability vectors on the first and second writes. We state this result formally.

**Lemma 7:** [21] For all \( (i,j) \in \{(1,0),(2,0),(2,1),\ldots,(2,q-2)\} \), let \( p_{2,i} = (p_{2,i-1},p_{2,i+1},\ldots,p_{2,q-2}) \) be a probability vector. Then, for all \( \epsilon > 0 \) there exists an \( \{n,2;M_1,M_2\}^{EIA:DU,z} \)-two-write \( q \)-ary EI:DU WOM code satisfying:
\[
\text{Im}(E_{q,1}) \subseteq B(n,p_{1,0}) \quad \text{and} \quad R_1 = \log M_1 \geq h(p_{1,0}) - \epsilon.
\]
For all \( c_1 \in \text{Im}(E_{q,1}), m_2 \in [M_2], \) and \( c_2 = E_{q,2}(m_2,c_1) \), the following condition holds. For \( i \in [q] \), let \( c_2^i \) be a length-\( w_{1,i} \) substring of \( c_2 \) at all locations \( k \) with value \( i \) before the second write, that is, \( c_{1,k}^i = c_{2,i} \). Then, \( c_2^i \in B(w_{1,i},p_{2,i}) \).

We refer to the family of WOM codes from Lemma 7 as an \( \{n,2;M_1,M_2\}^{EIA:DU,\epsilon}(p_{1,0},p_{2,0},\ldots,p_{2,q-2}) \)-WOM code, where \( M_1 = 2^R_0n \) and \( M_2 = 2^R_2n \) are determined as the maximal possible values based on \( \epsilon \), which tends to zero, and the probability vectors \( p_{2,i} \).

For the case \( q = 2 \), for shorthand, given \( p_{1,0},p_{2,1} = p \) we denote these codes by \( \{n,2;M_1,M_2\}^{EIA:DU,\epsilon}(p_{1,0},p_{2,0},\ldots,p_{2,q-2}) \) WOM code, where \( R_1 = \log M_1 \geq h(p) - \epsilon \) and \( R_2 = \log M_2 \geq 1 - p - \epsilon \).

We are now ready to present a construction of two-change three-write EIA:DU-ELM codes which establishes the result in Theorem 5.

**Construction 9:** Given \( p_{1,0},p_{2,0},p_{2,1} \in [0,0.5] \) and \( \epsilon > 0 \), we construct an \( \{n,3,2;M_1,M_2,M_3\}^{EIA:DU,z} \)-ELM code where \( M_j = 2^{R_j}n \) for \( j = 1,2,3 \) such that \( R_1 \geq h(p_{1,0}) - \epsilon \), \( R_2 \geq (1-p_{1,0})h(p_{2,1}) + p_{1,0}h(p_{2,1}) - \epsilon \), and \( R_3 \geq (1-p_{1,0}p_{2,1}) - \epsilon \).

We use the following two WOM codes.

1. \( L_1 \) \( = (p_{1,0},p_{1,0}-1,p_{1,0}-2) \) (1 = p_{1,0}, p_{1,0}, 0).
2. \( L_2 \) \( = (p_{2,0}-1,p_{2,0}-2) \) (0, 2, p_{2,0}, 0).
3. \( L_3 \) \( = (p_{2,1}-1,p_{2,1}-2) \) (1, p_{2,1}, 2, p_{2,1}).

Let \( C_1 \) be an \( \{n,2, M_1, M_2\}^{EIA:DU,\epsilon}(p_{1,0},p_{2,0},p_{2,1}) \)-two-write

\(^{3}\text{We assume here that } p_i \text{ is a rational number and } n \text{ is large enough such that } p_i n \text{ is an integer for } i \in [q].\)
ternary EI:DU WOM code from Lemma 7 with two pairs of encoder/decoder \((E_{3,1}, D_{3,1})\) and \((E_{3,2}, D_{3,2})\).

2) Let \(\rho_1 = p_{1,0}p_{2,1}\), and \(C_2\) be an \([n, 2; M_1, M_3]_{EIA:DU}^z(\epsilon, p_1)\) two-write binary EI:DU WOM code from Lemma 8 with two pairs of encoder/decoder \((E_{2,1}, D_{2,1})\) and \((E_{2,2}, D_{2,2})\).

The three pairs of encoder/decoder mappings \((E_{i}^{EIA:DU}, D_{j}^{EIA:DU})\) for \(j = 1, 2, 3\) are defined as follows.

**First write:** \(E_{1}^{EIA:DU}(m_1) = E_{3,1}(m_1)\) for all \(m_1 \in [M_1]\). Similarly, \(D_{1}^{EIA:DU}(c_1) = D_{3,1}(c_1)\). Note that since we chose the probability to program level 2 in the first write of \(C_1\) to be zero, the output of the encoder \(E_{3,1}\) is indeed a binary vector, so \(E_{1}^{EIA:DU}\) and \(D_{1}^{EIA:DU}\) are well defined.

**Second write:** The idea is to use the second write encoder \(E_{3,2}\) of \(C_1\) with the probability vectors \(p_{2,0}\) and \(p_{2,1}\), and notice that here we write all cells to levels 1 or 2. Then, we can view this “ternary word” as a binary word. Let \(c_1 = (c_{1,1}, \ldots, c_{1,n}) \in \text{Im}(E_{1}^{EIA:DU})\) be the cell-state vector after the first write, and note that this is a binary vector. The encoder/decoder \((E_{2}^{EIA:DU}, D_{2}^{EIA:DU})\) are defined formally as follows. For all \((m_2, c_1) \in [M_2] \times \text{Im}(E_{1}^{EIA:DU})\),

\[
c_2 = E_{2}^{EIA:DU}(m_2, c_1) = c_2' \mod 2,
\]

where \(c_2' = E_{3,2}(m_2, c_1) \in [3]^n\). Furthermore, for all \(c_2 \in \text{Im}(E_{2}^{EIA:DU})\),

\[
D_{2}^{EIA:DU}(c_2) = D_{3,2}(c_2') = m_2,
\]

where \(c_2' = 2 \cdot 1 - c_2\), that is, \(c_{2,1} = 1\) if \(c_{2,1} = 1\) and \(c_{2,1} = 2\) if \(c_{2,1} = 0\).

**Third write:** Let \(c_2\) be the cell-state vector after the second write. We note that the encoder on the third write knows the program-count vector \(v_2 \in [3]^n\), but the decoder does not have this information. Among the \(n\) cells, there are \(p_{1,0}n\) cells which have been programmed twice, where \(p_{1,0} = p_{1,0}p_{2,1}\), and therefore (only) these cells cannot be programmed on this write. Hence, the encoder can interpret the vector \(v_2\) as a length-\(n\) binary vector indicating for each cell whether it can be programmed on this write. We denote this vector by \(c_2'\), so \(c_{2,1}' = 1\) if and only if \(v_{2,1} = 1\). We will use the code \(C_2\) to encode and decode on this write, and we denote by \(\bar{e}\) the bitwise complement of a binary vector \(e\). Specifically, the encoder/decoder mappings are defined as follows. For all \(m_3 \in [M_3]\) and \(v_2 \in N_2\),

\[
E_{3}^{EIA:DU}(m_3, v_2) = \overline{E_{2,2}(m_3, c_2')}.
\]

Furthermore, for all \(c_3 \in \text{Im}(E_{3}^{EIA:DU})\),

\[
D_{3}^{EIA:DU}(c_3) = D_{2,2}(c_3').
\]

To illustrate Construction 9, we present the following example.

**Example 1:** Let \(n = 7, p_{1,0} = 3/7, p_{2,0} = 1/2\), and \(p_{2,1} = 1/3\), we construct a \([7, 3; M_1, M_3]_{EIA:DU}^z\) three-change-two-write ELM code as follows.

- In the first write, we encode a message \(m_1\) to obtain a binary vector of length 7, e.g., \(c_1 = (1, 1, 1, 0, 0, 0, 0)\).
- In the second write, to encode a binary vector \(m_2\), in the first step, we use the second write encoder \(E_{3,2}\) of the ternary code \(C_1\) in Lemma 7 with probability \(p_{2,0} = (0, 1/2, 1/2)\) and \(p_{2,1} = (2/3, 1/3)\) to obtain \(c_2' = E_{3,2}(c_1, m_2)\), e.g., \(c_2' = (2, 1, 1, 1, 1, 2, 2)\). In the second step, we can replace symbol 2 by symbol 0 in the vector \(c_2'\) to obtain the binary vector \(c_2 = (0, 1, 1, 1, 1, 0, 0)\). So, \(c_2\) is the output in the second write. We observe that it is not difficult to decode the vector \(c_2\) to obtain the message \(m_2\).

- In the last write, the encoder has all information and know that the first cell is programmed twice and the program-count vector is \(v_2 = (2, 1, 1, 1, 0, 0)\). The encoder also view the vector \(v_2\) as a binary vector \(c_2' = (1, 0, 0, 0, 0, 0)\) where the first cell (=1) is not programmable. Using the second-write encoder \(E_{3,2}\) of the EI:DU WOM code \(C_2\), we can encode the message \(m_3\) to obtain \(c_3' = E_{3,2}(m_3, c_2')\), e.g., \(c_3' = (1, 0, 0, 0, 1, 1, 1, 1)\). We now take the bitwise complement of \(c_3'\) to obtain \(E_{3,2}(m_3, v_2) = c_3 = (0, 1, 1, 1, 0, 0, 0)\). So, all the three messages are \(c_1 = (1, 1, 1, 0, 0, 0, 0)\), \(c_2 = (1, 1, 1, 1, 1, 0, 0)\), and \(c_3 = (0, 1, 1, 1, 0, 0, 0)\).

We now present the proof of Theorem 5.

**Proof of Theorem 5:** Let \(R_j(C_i)\) be the rate of the WOM code \(C_i\) on the \(j\)-th write. For any \(\epsilon > 0\) and \(p_{1,0}, p_{2,0}, p_{2,1} \in [0, 0.5]\), we choose the codes \(C_1\) and \(C_2\) in Construction 9 to satisfy

\[
R_1(C_1) \geq h(p_{1,0}) - \epsilon = h(p_{1,0}) - \epsilon,
\]

where \(p_{1,0} = (1 - p_{1,0}, p_{1,0})\),

\[
R_2(C_1) \geq h(p_{1,0})h(p_{2,0}) + h(p_{1,0}) - h(p_{2,1}) - \epsilon
\]

and

\[
R_2(C_2) \geq 1 - p_1 - \epsilon = 1 - p_{1,0}p_{2,1} - \epsilon.
\]

The result follows from the fact that the rate tuple of the two-change three-write ELM code is \((R_1(C_1), R_2(C_1), R_2(C_2))\).

The solution for the case \(t = 3, \ell = 2\) is generalized for any \(t\) and \(\ell\) in the following theorem.

**Theorem 10:** For all \(t\) and \(\ell, \ell \leq E_{1}^{EIA:DU} \cdot z\), that is, for any \(\epsilon > 0\) and a rate \(t\)-tuple \((R_1, \ldots, R_t) \in \ell_{t, \ell}\), there exists a zero-error \(\ell\)-change \(t\)-write EI:DU WOM code \(C\) such that its rate on the \(j\)-th write is at least \(R_j - \epsilon\) for all \(1 \leq j \leq t\), that is, \(R_j(C) \geq R_j - \epsilon\).

To prove Theorem 10, we construct a zero-error \(\ell\)-change \(t\)-write ELM code. The idea is to generalize Construction 9. Hence, for any given \(j\), we use \(q\)-ary EI:DU WOM code from Lemma 7 to program all cells up to the two highest levels \(q - 1\) and \(q - 2\). So, the decoder can look at \(q - 1\) as \(0\) and \(q - 2\) as \(1\) to decode the original message. We now present the construction formally as follows.

**Construction 11:** Given \(p_{j, i} \in [0, 0.5]\), for all \(i \in \ell_{t, \ell}, 1 \leq j \leq t\), and \(\epsilon > 0\), we construct an \([n, t, \ell; M_1, \ldots, M_t]_{EIA:DU}^z\) ELM code where \(M_1 = \left(p_{1,0}, \ldots, M_1\right)\) and \(M_2 = 2^{nR_j}\) for \(2 \leq j \leq t\) such that

\[
R_j \geq \sum_{i=0}^{n-1} Q_{j-1,h(p_{j, i})} - \epsilon,
\]

where \(Q_{j-1,h(p_{j, i})}\) is defined
in Equation (1). The \( t \) pairs of encoder/decoder mappings
\((E_j^{EIA:DU}, D_j^{EIA:DU})\) are defined as follows.

**First write:** Given \( p_{1,0} \), we program all words of length \( n \), weight \( |p_{1,0}| \) as on the first write of Construction 9. Hence, \( M_1 = (p_{1,0}) \) and the rate on the first write satisfies \( R_1 \geq h(p_{1,0}) - \epsilon \).

**j-th write, \( 2 \leq j \leq t \):** Let \( m = \min\{j, \ell\} \).

We denote the cell-state vector and the cell-program-count vector after the \( j-1 \) write by \( c_{j-1} = (c_{j-1,1}, \ldots, c_{j-1,n}) \in \text{Im}(E_j^{EIA:DU}) \) and \( v_{j-1} = (v_{j-1,1}, \ldots, v_{j-1,n}) \in B(n, q_{j-1}) \subset [m]^n \), respectively, where \( q_{j-1} = (Q_{j-1,0}, Q_{j-1,1}, \ldots, Q_{j-1,n-1}) \) and \( Q_{j,i} \) are defined in Equation (1). To program on the \( j \)-th write, we use the two-write \((2m + 1)\)-ary WOM code from Lemma 7, 
\[ [n, 2; M_1, M_2] \in \text{EIA:DU}, 2^{M_2}] \]\
where \( p_{1,0} \) is the encoded word and \( h(p_{1,0}) - \epsilon \). We now define the encoder/decoder (\( E_j^{EIA:DU}, D_j^{EIA:DU} \)) formally as follows.

For all \( m_j \in [M_j] \) and \( v_{j-1} \in \text{Im}(E_{m_j}) \),
\[ c_j = E_j^{EIA:DU}(m_j, v_{j-1}) = c_j'(\text{mod}2), \]
where \( c_j' = E_{m_j}^{EIA:DU}(m_j, v_{j-1}) \in [2m + 1]^n \). Furthermore, for all \( c_j \in \text{Im}(E_j^{EIA:DU}) \),
\[ D_j^{EIA:DU}(c_j) = D_{m_j}(c_j') = m_j, \]
where \( c_j' = 2m - 1 \) if \( c_{j,i} = 1 \) and \( c_{j,i} = 2m \) if \( c_{j,i} = 0 \).

**Proof of Theorem 10:** Given all parameters as in Construction 11, the rate of this ELM code on the first write is \( R_1 \geq h(p_{1,0}) - \epsilon \). Now, we consider the \( j \)-th write. Since we used the WOM code in Lemma 7 to program the \( j \)-th write of the ELM code, the rate on this write is exactly the rate on the second write of the used WOM code. Hence, the rate in the \( j \)-th write of the ELM code is \( R_j \geq \sum_{i=0}^{m-1} Q_{j,i} h(p_{j,i}) - \epsilon \).

**Remark 1:** In this section, we provide an explicit construction of zero-error two-change three-write EIA:DU ELM code and generalize the result to construct a zero-error \( \ell \)-change \( \ell \)-write EIA:DU ELM code. Since Shpilka [21] provided a pair of polynomial time encoding/decoding algorithms of a family of two-write WOM codes, the encoder and decoder in Theorem 10 also run in polynomial time. As shown in Theorem 10, using these constructions, we can achieve any rate in the capacity region and thus achieve the maximum sum-rate when the length \( n \) tends to infinity. However, for a fixed value of \( n \), we can only achieve a high sum-rate but cannot achieve the maximum sum-rate. Furthermore, Shpilka's technique only works for large block length [21]. Hence, for small value of block length \( n \), we need other constructions to obtain a high sum-rate, for example, Construction 19 that will be presented later.

**B. The EIA:DIA Model - Converse Part of Theorem 2**

In this section, we prove the converse part of Theorem 2 for the EIA:DIA model \( \epsilon \)-error case. That is, we prove that \( C_{EIA:DIA, \epsilon} \subset C_{t, \ell} \).

For this direction we need to prove that if there exists an \([n, t, \ell; M_1, \ldots, M_t]\) EIA:DIA, ELM code where \( p_\epsilon = (p_\epsilon_1, \ldots, p_\epsilon_\ell) \), then
\[ \left( \frac{\log M_1}{n} - \epsilon_1, \frac{\log M_2}{n} - \epsilon_2, \ldots, \frac{\log M_t}{n} - \epsilon_\ell \right) \in C_{t, \ell}, \]
where \( (\epsilon_1, \epsilon_2, \ldots, \epsilon_\ell) \) tends to 0 if \( p_\epsilon \) tends to 0 and \( n \) tends to infinity. In our proof \( \epsilon_j = \frac{H(p_{\epsilon_j}) + p_{\epsilon_j} \log(M_j)}{n} \), and therefore \( \epsilon_j \to 0 \) when \( p_{\epsilon_j} \to 0 \) and \( n \to \infty \).

Let \( X_j \) be a length-\( n \) binary vector where \( X_{j,k} = 1 \) if and only if the \( k \)-th cell is intended to be programmed on the \( j \)-th write. Similarly, \( Y_j \) is a length-\( n \) binary vector, where \( Y_{j,k} = 1 \) if and only if the value of the \( k \)-th cell was successfully changed on the \( j \)-th write, that is, \( Y_j = c_j \oplus c_{j-1} \). Note that the encoder knows the number of times each cell was programmed. Therefore, we can assume that a cell is not intended to be programmed more than \( \ell \) times. Furthermore, the decoder also knows the number of times each cell was programmed. Thus we assume that \( X_j = Y_j \) where \( X_j \) is the encoded word and \( Y_j \) is the input of the decoder.

Let \( S_1, \ldots, S_t \) be independent random variables, where \( S_j \) is uniformly distributed over the messages set \([M_j]\), and \( \hat{S}_j \) is the decoding result on the \( j \)-th write. Let \( V_j \) be an independent random variable on \( N_j \), the set of all cell-programs-count vectors after the first \( j \) writes. The data processing yields the following Markov chain:
\[ S_j|V_j = \hat{S}_j|V_j = Y_j|V_j = \hat{S}_j|V_j \]
and therefore, \( I(X_j; Y_j|V_j) \geq I(S_j; \hat{S}_j|V_j) \).

Additionally,
\[ I(S_j; \hat{S}_j|V_j) = H(S_j|V_j) - H(S_j|\hat{S}_j, V_j) \]
\[ \geq H(S_j) - H(S_j|\hat{S}_j) \]
\[ \geq \log(M_j) - H(p_{\epsilon_j}) - p_{\epsilon_j} \log(M_j). \]

The first inequality follows from the independence of \( V_j \) and \( S_j \) which implies that \( H(S_j|V_j) = H(S_j) \), and from the fact that conditioning does not increase the entropy. The second inequality follows from Fano's inequality [7, p. 38] \( H(S_j|\hat{S}_j) \leq H(p_{\epsilon_j}) + p_{\epsilon_j} \log(M_j) \).

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
\[
\frac{1}{n} I(X_j; Y_j|V_{j-1}) \leq \frac{1}{n} H(Y_j|V_{j-1})
\]
\[
\leq \frac{1}{n} \sum_{k=0}^{n-1} H(Y_{j,k}|V_{j-1,k})
\]
\[
= H(Y_{j,L}|V_{j-1,L})
\]
\[
= \sum_{i=0}^{\ell} Pr(V_{j-1,L} = i) H(Y_{j,L}|V_{j-1,L} = i)
\]
\[
= \sum_{i=0}^{\ell-1} Pr(V_{j-1,L} = i) H(Y_{j,L}|V_{j-1,L} = i)
\]
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
\[
H(Y_{j,L}|V_{j-1,L}, L) = \sum_{k=0}^{n-1} Pr(L=k) H(Y_{j,k}|V_{j-1,L}, L=k)
\]
and step (d) follows from \(H(Y_{j,L}|V_{j-1,L} = \ell) = 0\).

Now, we set \(p_{j,i} = Pr(X_j = 1|N_j-1_L = i) = Pr(Y_{j,L} = 1|N_j-1_L = i)\), and thus we can conclude that \(Q_{j,i} = Pr(N_j,L = i)\) where \(Q_{j,i}\) is calculated in Equation (1), and then
\[
\log(M_j) - \epsilon_j \leq \frac{1}{n} I(X_j; Y_j|N_{j-1})
\]
\[
\leq \sum_{i=0}^{\ell-1} Pr(N_{j-1,L} = i) H(Y_{j,L}|N_{j-1,L} = i)
\]
\[
= \sum_{i=0}^{\ell-1} Q_{j,i} - \epsilon_j h(p_{j,i})
\]
where \(\epsilon_j = \frac{H(p_{j,i}) + p_{j,i} \log(M_j)}{n}\), and the converse part is implied.

By Theorem 10 in Subsection III-A and by the proof of the converse part in Subsection III-B we completed the proof of Theorem 2. Furthermore by Theorem 4 we conclude the following corollary.

Corollary 12: For all \(t, \ell, C_{\ell,t} = \tilde{C}_{\ell,t}\) is the capacity region for all the EIA models for both the zero-error and the \(\epsilon\)-error cases and is denoted by \(C_{EIA}\). The maximum sum-rate of all the EIA models is \(R_{EIA} = \log(\ell)!\).

IV. THE CAPACITY OF THE EIP:DIA MODEL

In this section we discuss the capacity region and the maximum sum-rate of the EIP:DIA model. Recall that if \(\ell = 1\) then by definition, EIP is equivalent to EIA and this model is equivalent to the known WOM model. Thus, in this section we assume that \(\ell > 1\). We focus on the \(\epsilon\)-error case and present the capacity region of this model. The zero-error case is harder to solve, and is left for future research. However the \(\epsilon\)-error case provides an upper bound for the zero-error case. Note that the EU:DI WOM model is simpler than the EIP:DIA ELM model, and even though its exact capacity for the zero-error case is still not known for general \(t\). As done in the EIA models, let us denote by \(c_{j,i} \in [t + 1]\), the length-\(n\) binary vector which represents the memory state after the \(j\)-th write, where \(c_0 = 0\).

For \(1 \leq j \leq t\) and \(i \in [\ell + 1]\), we define the probabilities \(p_{j,i}, p_{j,i}\), and \(Q_{j,i}\) as follows. \(p_{j,k}\) is the probability of programming a cell on the \(j\)-th write given that the value of this cell was \(k\), \(k \in \{0,1\}\), and \(Q_{j,i}\) is the probability of a cell to be programmed exactly \(i\) times after the first \(j\) writes. Additionally, let \(Q_{j,e}, Q_{j,0}\) be the probability of a cell to be programmed an even, odd number of times after the first \(j\) writes, respectively. Formally, \(Q_{j,i}\) is defined recursively by using the probabilities \(p_{j',0}\) and \(p_{j',1}\) for \(j' \leq j\). We now assume that \(\ell\) is even. The case of an odd \(\ell\) is defined similarly.

We define \(Q_{j,i}\) for \(j > 0\) as follows. For even \(i \geq 0\),
\[
Q_{j,i} = \begin{cases} 
Q_{j-1,i-1}p_{j,1} + Q_{j-1,i}(1 - p_{j,0}), & \text{if } 0 < i < \ell, \\
Q_{j-1,i-1}p_{j,1} + Q_{j-1,i}, & \text{if } i = \ell, \\
Q_{j-1,i}(1 - p_{j,0}), & \text{if } i = 0, 
\end{cases}
\]
and for odd \(i > 0\), \(Q_{j,i} = Q_{j-1,i-1}p_{j,0} + Q_{j-1,i}(1 - p_{j,1})\). The base \(j = 0\), is \(Q_{0,0} = 1\) and \(Q_{0,i} = 0\) for \(i > 0\). Furthermore, let \(Q_{j,e} = \sum_{i=0}^{\ell/2} Q_{j,2i}\) and \(Q_{j,0} = \sum_{i=1}^{\ell/2} Q_{j,2i-1}\).

Next, we define the rates region \(C_{\ell,t}\) which will be proved to be the capacity region of the EIP:DIA model for the \(\epsilon\)-error case. We present here the definition for even \(\ell\), while the odd case is defined similarly.

\[
\tilde{C}_{\ell,t} = \{(R_{1}, R_{2}, \ldots, R_{t})|y|1 \leq j \leq t:\n\]
\[
R_{j} \leq Q_{j-1,0}h(p_{j,1}) + (Q_{j-1,e} - Q_{j-1,0})h(p_{j,0}),
\]
\[
\sum_{i=0}^{\ell/2} Q_{j,2i} \leq 0, 0.5\} \text{and } Q_{j,e}, Q_{j,0}, Q_{j,\ell} \text{ are defined above}.\n\]

For example, for \(t = 3, \ell = 2\), we have that
\[
\tilde{C}_{3,2} = \tilde{C}_{3,2} = \{(R_{1}, R_{2}, R_{3})|R_{1} \leq h(p_{1,0}),
\]
\[
R_{2} \leq 1 - p_{1,0}h(p_{2,1}),
\]
\[
R_{3} \leq 1 - p_{1,0}p_{2,1}, \text{ and } p_{2,1} \in [0, 0.5]\},
\]
which is achieved by substituting \(p_{3,0} = p_{3,1} = p_{2,0} = 0.5\) in Equations (2) and (5). Using the region \(\tilde{C}_{\ell,t}\), the next theorem characterizes the capacity region of the EIP models for the \(\epsilon\)-error case.

Theorem 13: The rates region \(\tilde{C}_{\ell,t}\) is the capacity region of \(t\)-write \(\ell\)-change ELM EIP:DIA model for the \(\epsilon\)-error case. That is, \(\tilde{C}_{\ell,t} = C_{\ell,t}^{EIP:DIA}\).

Proof: To show the achievable region, we should prove that for each \(\epsilon > 0\) and \((R_{1}, R_{2}, \ldots, R_{t}) \in \tilde{C}_{\ell,t}\), there exists an \([n, t; M_{1}, \ldots, M_{t}]_{\ell}\) ELM code, where for all \(1 \leq j \leq t, \log M_{j} \geq R_{j} - \epsilon\) and \(p_{e} = (p_{e_{1}}, \ldots, p_{e_{t}}) \leq (\epsilon, \ldots, \epsilon)\). We use the well-known random channel-coding theorem [7, p. 200] on each write. We describe the encoding and decoding on each write.

The \(j\)-th write presents a DMC with the input length-\(n\) binary vector \(X_{j}\) and the output is \((Z_{j-1}, Y_{j})\), where
For the EIA models, i.e.,

By substituting $p$ and $Z$ for $g \in \{z, \epsilon\}$ and $\tilde{Y} \in \{\{A, I, P, U\}\}$, was stated in Section III to be equal to $C_{t, \ell}$, while in Section IV we presented the capacity region of the EIP:DIA model for the $\epsilon$-error case, $C_{t, \ell}^EIP:DIA,\epsilon$.

The next theorem proves that for $t > \ell \geq 3$ the maximum sum-rate of the EIP:DIA model for the epsilon-error case is smaller than the maximum sum-rate of the EIA models. Hence, the capacity region $C_{t, \ell}^EIP:DIA,\epsilon$ is a proper subset of the capacity region $C_{t, \ell}^EIA$ for these parameters. Recall that for $\ell = 2$ these regions were shown to be the same in Theorem 14, and therefore the maximum sum-rates of these models for $\ell = 2$ are the same too.

**Theorem 15:** For $t > \ell \geq 3$, $R_{t, \ell}^EIP:DIA,\epsilon < R_{t, \ell}^EIA$, and hence $C_{t, \ell}^EIP:DIA,\epsilon \subset C_{t, \ell}^EIA$.

**Proof:** Let $\tilde{R} = (\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_t)$ be a rate tuple which achieves the maximum sum-rate $R_{t, \ell}^EIP:DIA,\epsilon$, and denote by $p_{j,0}, p_{j,1},$ and $Q_{j,i}$, $1 \leq j \leq t$ and $i \in [\ell + 1]$, the probabilities which attain $\tilde{R}$ in $C_{t, \ell}$.

Now we present a rate tuple $R = (R_1, R_2, \ldots, R_t) \in C_{t, \ell}$. Then, we conclude that $R \in C_{t, \ell}^EIA \setminus C_{t, \ell}^EIP:DIA,\epsilon$, which implies that $R_{t, \ell}^EIP:DIA,\epsilon < R_{t, \ell}^EIA$ and $C_{t, \ell}^EIP:DIA,\epsilon \subset C_{t, \ell}^EIA$.

We assume now that $\ell$ is even, while the proof for the odd case is similar. Since $R$ is maximal rate tuple we have $p_{t-1,0} = \tilde{p}_{t-1,0} = \tilde{p}_{t-1,1} = 0.5$. For all $j$ and $i$ such that $1 \leq j \leq t - 2$ and $i \in [\ell]$, we define $p_{j,i} = \tilde{p}_{j,i}$ by $i \equiv i \mod 2$. In 2. Add. 2, for all $i \in [\ell - 1]$, $p_{t-1,i} = 0.5$, $p_{t-1,i-1} = \tilde{p}_{t-1,i-1}$, and for all $i$, $p_{i,0} = 0.5$.

Thus, for all $j$ and $i$, $1 \leq j \leq t - 2$ and $i \in [\ell]$, $R_j = \tilde{R}_j$ and $Q_{j,i} = \tilde{Q}_{j,i}$. For the $(t - 1)$-th write we have, $R_{t-1} = 1 - Q_{t-2,\ell-1} - Q_{t-2,\ell-1} + Q_{t-2,\ell-1}h(\tilde{p}_{t-1,1})$, while $R_{t-1} = \tilde{R}_{t-1} = Q_{t-2,\ell-1}h(\tilde{p}_{t-1,1}) + (Q_{t-2,\ell-1} - Q_{t-2,\ell-1})$, and for the last write $R_t = \tilde{R}_t = 1 - Q_{t-1,1}$.

Now we prove that $\tilde{p}_{t-1,1} < 0.5$ which immediately implies that $R_{t-1} > \tilde{R}_{t-1}$ and thus completes the proof. Recall that $\tilde{R}_{t-1} = 1 - Q_{t-1,1} = 1 - Q_{t-2,\ell-1} - Q_{t-2,\ell-1}p_{t-1,1}$. Thus, given the probabilities for the first $t - 2$ writes, in order to achieve the maximal rate tuple $\tilde{R}$, we have to maximize $R_{t-1} + \tilde{R}_t$. That is, we choose $\tilde{p}_{t-1,1}$ which maximizes $Q_{t-2,\ell-1}h(\tilde{p}_{t-1,1}) - Q_{t-2,\ell-1}p_{t-1,1}$. The derivative is $Q_{t-2,\ell-1}log\frac{1-p_{t-1,1}}{\tilde{p}_{t-1,1}} - Q_{t-2,\ell-1} - 1$, and the maximum is obtained for $\tilde{p}_{t-1,1} = 1/(1 + 2Q_{t-2,\ell-1}/Q_{t-2,\ell-1})$. Since $R$ is maximal and $t > \ell \geq 3$, we have $Q_{t-2,\ell-1} = 0$, and therefore $\tilde{p}_{t-1,1} \neq 0.5$.

We can summarize the results regarding the capacity region of the EIP:DIA model in the following corollary.
Corollary 16: For all $t > \ell$ the following holds

$$C_{t,\ell}^{EIP:DIA,\epsilon} = C_{t,\ell} \subseteq C_{t,\ell}^{EIA}.$$  

Furthermore,

- For $t > \ell = 2$ all these regions are equal, in particular,
  $$C_{t,2}^{EIP:DIA,\epsilon} = C_{t,2}^{EIA},$$
- For $t > \ell \geq 3$, $C_{t,\ell}^{EIP:DIA,\epsilon} \subseteq C_{t,\ell}^{EIA}$ and $R_{t,\ell}^{EIP:DIA,\epsilon} < R_{t,\ell}^{EIA}$.

VI. THE CAPACITY OF EU:DIA MODEL

In this section, we study the EU:DIA model for the $\epsilon$-error case, and provide the capacity region of this model. As in the EIP:DIA model, the capacity region for the zero-error case and the exact maximum sum-rate are left for future research.

For $1 \leq j < t$ and $i \in [\ell + 1]$, let $p_j$ be the probability of programming a cell on the $j$-th write, and $Q_{j,i}$ denotes the probability of a cell to be programmed exactly $i$ times on the first $j$ writes. Additionally, let $Q_{j,i}:Q_{j,o}$ be the probabilities of a cell to be programmed even, odd number of times on the first $j$ writes, respectively. Formally, $Q_{j,i}$ is defined recursively by using $p_j$ probabilities for $j' < j$. For $j \geq 1$,

$$Q_{j,i} = \begin{cases} Q_{j-1,i-1}p_j + Q_{j-1,i}(1-p_j), & \text{if } 0 < i \leq \ell, \\ Q_{j-1,i-1}p_j + Q_{j-1,i}, & \text{if } i = \ell, \\ Q_{j-1,i-1}(1-p_j), & \text{if } i = 0, \\ \end{cases}$$

where $Q_{0,0} = 1$ and $Q_{0,i} = 0$ for $i > 0$.

Then, we define the region $\overline{C}_{t,\ell}^{EU:DIA,\epsilon}$ which is proved later in this section to be the capacity region $C_{t,\ell}^{EU:DIA,\epsilon}$.

$$\overline{C}_{t,\ell}^{EU:DIA,\epsilon} = \{(R_1, R_2, \ldots, R_t) | \forall 1 \leq j \leq t : R_j \leq h(p_j) - Q_{j-1,i}p_j, p_j \in [0, 0.5], Q_{j,i} \text{ is defined above} \}.$$  

(7)

The next theorems establish the capacity region of the EU:DIA model for the $\epsilon$-error case and compare between this model and the EIP:DIA model. The techniques applied for the EU:DIA model are very similar to the proofs in Section IV. The proofs of Theorems 17 and 18 are similar to the proofs of Theorems 13 and 15, respectively. Therefore, these proofs are moved to Appendix.

Theorem 17: The rates region $\overline{C}_{t,\ell}^{EU:DIA,\epsilon}$ is the capacity region of t-write $\ell$-change ELM EU:DIA model for the $\epsilon$-error case. That is, $\overline{C}_{t,\ell} = C_{t,\ell}^{EU:DIA,\epsilon}$.

Theorem 18: For $t > \ell \geq 2$, $R_{t,\ell}^{EU:DIA,\epsilon} < R_{t,\ell}^{EIP:DIA,\epsilon}$, and hence $C_{t,\ell}^{EU:DIA,\epsilon} \subseteq C_{t,\ell}^{EIP:DIA,\epsilon}$.

VII. THE EIP:DU MODEL

In this section, we study the EIP:DU model and its sum-rate. First, we note that $C_{t,\ell}^{EIP:DU,\epsilon} \subseteq C_{t,\ell}^{EIP:DIA,\epsilon}$ for all $t, \ell$, and thus,

$$R_{t,\ell}^{EIP:DU,\epsilon} \leq R_{t,\ell}^{EIP:DIA,\epsilon} \leq \log \left( \sum_{i=0}^{\ell} \binom{t}{i} \right).$$  

(8)

That is, we obtain an upper bound of the maximum sum-rate $R_{t,\ell}^{EIP:DU,\epsilon}$. Note that for $t > \ell \geq 3$ this upper bound is not tight (Theorem 15). We are now interested in some good lower bounds for the maximum sum-rate. Our goal is to provide several constructions with high sum-rate. We first present a general construction for the zero-error case and then show how to obtain higher sum-rate for the $\epsilon$-error case with $t = 3, \ell = 2$.

The following construction provides a family of $\ell$-change t-write EIP:DU ELM codes for the zero-error case.

Construction 19: Let $(k_1, \ldots, k_l)$ be such that $1 \leq k_i \leq t$ for $1 \leq i \leq \ell$ and $\sum_{i=1}^{\ell} k_i = t$. Let $[n, k_i; M_{j_1+1}, \ldots, M_{j_2+k_i}]^{EIP:DU,\epsilon}$ be a binary $k_i$-write EIP:DU WRM code for $1 \leq i \leq \ell$ with sum-rate $R_i$ where $j_i = 0$ and $j_{i+1} = \sum_{j=1}^{i} k_j$. Each of which consists of $n$ bits and $k_i$ pairs of encoding and decoding maps $\mathbf{E}_{i}^{EIP:DU,\epsilon} = \mathbf{D}_{i}^{EIP:DU,\epsilon}$ for $1 \leq h \leq k_i$. We define an $[n, t, 1, M_1, \ldots, M_\ell]^{EIP:DU,\epsilon}$ t-change w-write ELM code consists of $n$ bits and $t$ pairs of encoders and decoders $\mathbf{E}_{j}^{EIP:DU,\ell} = \mathbf{D}_{j}^{EIP:DU,\ell}$ where $\mathbf{E}_{j}^{EIP:DU,\ell} = \mathbf{D}_{j}^{EIP:DU,\ell}$ for $1 \leq j \leq t$. The maximum sum-rate of the ELM codes from Construction 19 is $R_{\text{sum}} \geq \sum_{i=1}^{\ell} \log(k_i + 1) - \epsilon/\ell$ since for $1 \leq i \leq \ell$, $R_i \geq \log(k_i + 1) - \epsilon/\ell$ and $R_{\text{sum}} = \sum_{i=1}^{\ell} R_i$. Hence, in order to maximize the sum-rate, our goal is to maximize the value of $\sum_{i=1}^{\ell} \log(k_i + 1)$ given that $\sum_{i=1}^{\ell} k_i = t$. Assume that $t = k\ell + r$, $r \in [\ell]$, then this maximum value will be achieved when choosing $k_1 = \cdots = k_{r} = k + 1$ and $k_{r+1} = \cdots = k_{t} = k$. The next corollary summarizes this result.

Corollary 20: For all $t$ and $\ell$, where $t = k\ell + r$, $r \in [\ell]$,

$$R_{t,\ell}^{EIP:DU,\epsilon} \geq r \log(k + 2) + (\ell - r) \log(k + 1)$$

$$= \ell \log \left( \frac{t}{\ell} + 1 \right) + (t \mod \ell) \log \left( \frac{1}{1 - \frac{1}{k+1}} \right).$$

Proof: We choose $(k_1, \ldots, k_l)$ such that $k_1 = \cdots = k_r = k + 1$ and $k_{r+1} = \cdots = k_t = k$ and thus $\sum_{i=1}^{\ell} k_i = t$. We note that $k = \left\lfloor \frac{t}{\ell} \right\rfloor$ and $r = t \mod \ell$. Since we presented in Construction 19 an $[n, t, 1, M_1, \ldots, M_\ell]^{EIP:DU,\epsilon}$ t-change t-write ELM code with sum-rate $R_{\text{sum}} = \sum_{i=1}^{t} R_i \geq r \log(k + 2) + (\ell - r) \log(k + 1) - \epsilon$ for any $\epsilon > 0$, we obtain the result in Corollary 20.

From the above corollary, we have a lower bound of the maximum sum-rate of the EIP:DU model. Recall that $R_{t,\ell}^{EIP:DU,\epsilon} \leq R_{t,\ell}^{DIA,\epsilon} = \log \left( \sum_{i=0}^{t} \binom{t}{i} \right)$, that is the exact maximum sum-rate of the EIA model is an upper bound of the maximum sum-rate of the EIP:DU model. Hence, we obtain a lower bound and an upper bound of the maximum sum-rate of the EIP:DU ELM model. We note that when $t \leq \ell$, we always achieve the full capacity, that is, the maximum sum-rate is $t$. When $t > \ell$, the maximum sum-rate is difficult to compute exactly and there is a gap between the above lower and upper bounds. We illustrate the results for $\ell = 2, t \in [3, 23]$ in the following figure.

The following result shows that for $\ell = 2$ the sum-rate of the ELM code from Construction 19 is already close to the upper bound when $t$ is large and $n \rightarrow \infty$. 

Authorized licensed use limited to: Technion Israel Institute of Technology. Downloaded on February 23,2022 at 17:36:38 UTC from IEEE Xplore. Restrictions apply.
Proposition 21: For \( \ell = 2 \) and \( t \geq 3 \), \( R_{t,2}^{EIP;DU,z} \geq R_{t,2}^{EIA} - 1 \).

Proof: Recall that \( R_{t,2}^{EIA} = \log \sum_{i=0}^{t} {t \choose i} = \log \left( \frac{t^2 + t + 2 \ell}{2} \right) \).

When \( t \) is even, there exists a positive integer \( t_1 \) such that \( t = 2t_1 \). In this case,

\[
R_{t,2}^{EIP;DU,z} \geq 2 \log(t_1 + 1) = \log(t_1^2 + 2t_1 + 1)
\]

and,

\[
R_{t,2}^{EIA} = \log \left( \frac{4t_1^2 + 2t_1 + 1}{2} \right).
\]

Hence,

\[
R_{t,2}^{EIA} - R_{t,2}^{EIP;DU,z} \leq 4t_1^2 + 2t_1 + 1 \leq 2(2t_1 + 1) = 2 \leq 2 = 1.
\]

When \( t \) is odd, there exists a positive integer \( t_2 \) such that \( t = 2t_2 + 1 \). In this case,

\[
R_{t,2}^{EIP;DU,z} \geq \log(t_2 + 1) + \log(t_2 + 2) = \log(t_2^2 + 3t_2 + 2)
\]

and,

\[
R_{t,2}^{EIA} = \log \left( \frac{4t_2^2 + 6t_2 + 4}{2} \right).
\]

Hence,

\[
R_{t,2}^{EIA} - R_{t,2}^{EIP;DU,z} \leq 4t_2^2 + 6t_2 + 4 \leq 2(2t_2 + 3t_2 + 2) = 2 \leq 2 = 1.
\]

In conclusion, the proposition is proven.

We note that when \( t = 3 \) and \( \ell = 2 \), the maximum achievable sum-rate of the codes in Construction 19 is \( \log 6 \approx 2.585 \), while the upper bound is \( \log 7 \approx 2.807 \). Lastly, we show how to improve this result for the \( c \)-error case.

The main ideas of the following construction are as follows. On the first two writes, we follow exactly the first two writes of EIA:DU ELM code when \( t \leq 2 \). For the third write, \( t \geq 3 \), and \( \ell = 2 \). In order to overcome this difficulty, we use the following family of binary EU:DU WOM codes.

**Definition 22:** An \( [n,2; M_1, M_2]_{2}^{EU;DU,(p_1,p_2)} \) two-write binary EU:DU WOM code is a coding scheme comprising of \( n \) bits. It consists of two pairs of encoding and decoding maps \((\mathcal{E}_j^{EU;DU}, \mathcal{D}_j^{EU;DU})\) for \( j = 1,2 \).

For the map \( \mathcal{E}_j^{EU;DU} \), \( \text{Im}(\mathcal{E}_j^{EU;DU}) \) is its image and \( \text{Im}^*(\mathcal{E}_j^{EU;DU}) \) is the set of all the cell-state vectors which can be obtained after the \( j \)-th write. We note that \( \text{Im}(\mathcal{E}_0^{EU;DU}) = \text{Im}^*(\mathcal{E}_0^{EU;DU}) = \{(0, \ldots , 0)\} \) and \( \text{Im}^*(\mathcal{E}_2^{EU;DU}) = \{\max\{c_1, c_2\} \mid c_i \in \text{Im}(\mathcal{E}_i^{DU;EU}) \} \).

The encoding and decoding maps are defined as follows. For \( j = 1,2 \),

\[
\mathcal{E}_j^{EU;DU} : [M_j] \rightarrow \mathcal{B}(n,(1-p_j, p_j))
\]

and

\[
\mathcal{D}_j^{EU;DU} : \text{Im}^*(\mathcal{E}_j^{EU;DU}) \rightarrow [M_j]
\]

such that for all \( m \in [M_j] \),

\[
\sum_{(m,e) \in \text{Im}(\mathcal{E}_j^{EU;DU})} Pr(m) Pr(e) I_m(D_j^{EU;DU}(\max\{c_1, c_2\} \mid c_i \in \text{Im}(\mathcal{E}_i^{DU;EU}) )) \leq p_e.
\]

Two-write binary EU:DU WOM codes have been studied for a long time [23]. Recently, in [11] several constructions of EU:DU WOM codes were presented. Assume that there exists a capacity achieving code for the \( Z \) channel, then the following result for EU:DU WOM codes can be received based upon the constructions from [11].

**Lemma 23:** [11] For all \( 0 \leq p_1, p_2 \leq 0.5 \) and \( \epsilon > 0 \) there exists an \( [n, 2; M_1, M_2]_{EU;DU,(0,\epsilon)} \) two-write binary EU:DU WOM code satisfying:

\[
\cdot c_1 \in \mathcal{B}(n,(1-p_1, p_1)), \quad R_1 = \log \frac{M_1}{n} \geq h(p_1) - \epsilon.
\]

\[
\cdot c_2 \in \mathcal{B}(n,(1-p_2, p_2)), \quad R_2 = \frac{n \log M_2}{n} \geq h(\frac{p_1}{p_2}) - \epsilon
\]

where \( c_i \in \text{Im}(\mathcal{E}_i^{EU;DU}) \) for \( i = 1,2 \).

We refer to the family of WOM codes from Lemma 23 as an \( [n, 2; M_1, M_2]_{EU;DU,(0,\epsilon)} \) WOM code, where \( M_1 = 2R_1n \) and \( M_2 = 2R_2n \) are determined as the maximal possible values based on \( \epsilon \), which tends to zero, and the probabilities \( p_1 \) and \( p_2 \).

We are now ready to present a construction of two-change three-write EIP:DU ELM code.

**Construction 24:** Given \( p_0, p_1, p_2, p_1, p_3 \in [0,0.5] \), we use the following two codes:

- An \([n, 3, 2; M_1, M_2, M_3]_{EIA;DU,z}^{EIA;DU,z}\) code from Construction 9 with the first two pairs of encoder/decoder \((\mathcal{E}_i^{EIA;DU,z}, \mathcal{D}_i^{EIA;DU,z})\) for \( i = 1,2 \).

- An \([n, 2; M_1, M_2]_{EU;DU,(0,\epsilon)}^{EU;DU,(0,\epsilon),(c_1, p_1, p_3)}\) two-write binary EU:DU WOM code from Lemma 23, \( p_1 = p_1 \text{p_2} \), with the pair of encoder/decoder in the second write \((\mathcal{E}_2^{EU;DU}, \mathcal{D}_2^{EU;DU})\).

We construct \([n, 3; M_1, M_2, M_3]_{EIP;DU,(0,0,\epsilon)}^{EIP;DU,(0,0,\epsilon)}\) two-change three-write EIP:DU ELM code where its 3 pairs of encoding/decoding maps \((\mathcal{E}_j^{EIP;DU}, \mathcal{D}_j^{EIP;DU})\) for \( j = 1,2,3 \) are defined as follows.

1. For \( i = 1,2 \), \( \mathcal{E}_i^{EIP;DU} = \mathcal{E}_i^{EIA;DU} \) and \( \mathcal{D}_i^{EIP;DU} = \mathcal{D}_i^{EIA;DU} \).

- That is, the first two writes of this EIP:DU
ELM code are exactly the same as the first two writes of the EIA:DU ELM code from Construction 9.

(2) After the first two writes, we note that \( p_{1,0} \) cells are already programmed twice, and thus cannot be programmed this time. Hence, we use the pair of encoder/decoder \((E_{\text{EU:DU}}^3, E_{\text{DU:EU}}^3)\) to encode/decode information. The pair of encoder/decoder in the third write is defined formally as follows:

\[
E_{\text{DU:EU}}^3 : [M_3] \times \text{Im}^*(E_{\text{EU:DU}}^3) \rightarrow [2]^n
\]

such that for all \( m_3 \in [M_3] \) and \( c_2 \in \text{Im}^*(E_{\text{EU:DU}}^3) \),

\[
E_{\text{DU:EU}}^3 (m_3, c_2) = E_{\text{EU:DU}}^3 (m_3).
\]

Furthermore,

\[
D_{\text{DU:EU}}^3 : \text{Im}^*(E_{\text{EU:DU}}^3) \rightarrow [M_3]
\]

such that for all \( c_3^* \in \text{Im}^*(E_{\text{DU:EU}}^3) \),

\[
D_{\text{DU:EU}}^3 (c_3^*) = D_{\text{EU:DU}}^3 (c_3^*) = m_3.
\]

On the second write, it is clear that \( R_1 \geq h(p_{1,0}) - \epsilon \) and \( R_2 \geq (1 - p_{1,0}) h(p_{2,0}) + p_{1,0} h(p_{2,1}) - \epsilon \). In the third write, \( R_3 \geq h(p_{1,0} p_{2,1} p_3) - p_3 h(p_{1,0} p_{2,1}) - \epsilon \) for all \( \epsilon > 0 \).

Therefore, the following region is achievable for the \( \epsilon \)-error case:

\[
C_{3,2,1}^{\text{EU:DU}} = \{(R_1, R_2, R_3) : R_1 \leq h(p_{1,0}), R_2 \leq (1 - p_{1,0}) h(p_{2,0}) + p_{1,0} h(p_{2,1}), R_3 \leq h(p_{1,0} p_{2,1} p_3) - p_3 h(p_{1,0} p_{2,1}), p_{1,0} p_{2,1} p_3 \in [0, 1]\}.
\]

The sum-rate of the above code is \( R_{\text{sum}} = R_1 + R_2 + R_3 \geq h(p_{1,0}) + (1 - p_{1,0}) h(p_{2,0}) + p_{1,0} h(p_{2,1}) + h(p_{1,0} p_{2,1} p_3) - p_3 h(p_{1,0} p_{2,1}) - \epsilon \) for all \( \epsilon > 0 \). By choosing \( p_{1,0} = 3/7, p_{2,0} = 1/2, p_{2,1} = 2/3, \) and \( p_3 = 1/2 \), we obtain the sum-rate \( R_{\text{sum}} = R_1 + R_2 + R_3 \approx 2.64 \).

Remark 2: In this section, we construct a family of zero-error \( \epsilon \)-change-twrite EIP:DU ELM codes for any \( \ell \) and \( t \). Using some efficient encoding/decoding algorithms of the well-known binary t-write EI:DU WOM codes, we can encode/decode our EIP:DU ELM codes efficiently in polynomial time. When \( n \) tends to infinity, we can obtain some codes with high sum-rate and can get a lower bound on the maximal sum-rate of the EIP:DU model. We note that the lower bound is not tight even though it is close to the upper bound. We actually improve the lower bound for the \( \epsilon \)-error case when \( \ell = 2 \) and \( t = 3 \) in Construction 24. Using some known polynomial time encoding/decoding algorithms of a two-write EU:DU WOM code in Lemma 23 [11], the encoding and decoding algorithms in Construction 24 also run in polynomial time. Since the exact capacity region and the maximum sum-rate of the EIP:DU model are not known yet, we expect to have better constructions in near future.

VIII. CONCLUSION

In this paper, we have proposed and studied a new coding scheme, called ELM codes. This family of codes can be used to increase the endurance of resistive memories by rewriting codes. This new family of rewriting codes generalizes the well-known WOM codes. We investigated the coding schemes of nine different models which depend upon the knowledge of the encoder and the decoder. In all these models, we focused on the capacity region and the achievable maximum sum-rate. In several important models, we also presented constructions of ELM codes with high sum-rate and some constructions of capacity-achieving codes. For future work, we are interested in practical constructions of capacity-achieving codes with efficient encoding/decoding algorithms, especially in the EIP:DU model.

APPENDIX

Theorem 13 - The Converse Part: The rates region \( \mathcal{C}_{e,t} \) is a superset of the capacity region of \( t \)-write \( \ell \)-change ELM EIP:DU model for the \( \epsilon \)-error case. That is,\( C_{e,t}^{\text{EU:DU,\ell}} \subseteq \mathcal{C}_{e,t} \).

Proof: Let \( S_j, \hat{S}_j, V_j, 1 \leq j \leq t, \) and \( L \) be defined as in the proof of the converse part in Theorem 2. Thus, exactly as proved in Theorem 2, we have \( I(X_j; Y_j|V_{j-1}) \geq I(S_j; \hat{S}_j|V_{j-1}), I(S_j; \hat{S}_j|V_{j-1}) \geq \log(M_j) - H(p_{e_j}) - p_{e_j} \log(M_j) \), and

\[
\frac{1}{n} I(X_j; Y_j|V_{j-1}) \leq \sum_{i=0}^{\ell-1} \Pr(V_{j-1,L} = i) H(Y_j, L|V_{j-1,L} = i).
\]

Now, we set \( p_{1,0} = \Pr(X_{j,L} = 1|V_{j-1,L} \mod 2 = 0) \) and similarly \( p_{1,1} = \Pr(X_{j,L} = 0|V_{j-1,L} \mod 2 = 1) \). Thus, for even \( i < \ell \)

\[
H(Y_j, L|V_{j-1,L} = i) = H(p_{1,0}) \quad \text{and for odd } i \leq \ell \quad H(Y_j, L|V_{j-1,L} = i) = H(p_{1,1}).
\]

We also define for \( i \in [\ell + 1] \)

\[
Q_{j,i} = \Pr(V_{j,L} = i) \quad \text{and note that } Q_{j,i} \text{ can be calculated as in Equation (4), and we use the notations } Q_{j,0} \text{ and } Q_{j,e} \text{ as defined above. Then,}
\]

\[
\log(M_j) - \epsilon_j \leq \frac{1}{n} I(X_j; Y_j|V_{j-1}) \leq \sum_{i=0}^{\ell-1} \Pr(V_{j-1,L} = i) H(Y_j, L|V_{j-1,L} = i) = \sum_{i=1}^{\ell/2} Q_{j-1,2i-1} h(p_{j,1}) + Q_{j-1,2i-2} h(p_{j,0}) \geq h(p_{j,1}) \sum_{i=1}^{\ell/2} Q_{j-1,2i-1} + h(p_{j,0}) \sum_{i=1}^{\ell/2} Q_{j-1,2i-2} = Q_{j-1,\epsilon} h(p_{j,1}) + (Q_{j-1,\epsilon} - Q_{j-1,\epsilon}) h(p_{j,0}),(\epsilon_j = H(p_{e_j}) + p_{e_j} \log(M_j)) ,
\]

and the claim is implied.

Theorem 17: The rates region \( \mathcal{C}_{e,t} \) is the capacity region of \( t \)-write \( \ell \)-change ELM EU:DU model for the \( \epsilon \)-error case. That is, \( \mathcal{C}_{e,t} = C_{e,t}^{\text{EU:DU,\ell}} \).

Proof: To show the achievable region, we should prove that for each \( \epsilon > 0 \) and \( \{R_1, R_2, \ldots, R_t\} \in \mathcal{C}_{e,t} \), there exists an \( \{n, t, M_1, \ldots, M_t\} \in C_{e,t}^{\text{EU:DU,\ell}} \), an ELM code, where for all \( 1 \leq j \leq t, \frac{\log M_j}{n} \geq R_j - \epsilon \) and \( p_{e_j} = (p_{e_1}, \ldots, p_{e_t}) \leq (\epsilon, \ldots, \epsilon) \). We use the well-known random channel-coding theorem [7, p. 200] on each write. We describe the encoding and decoding on each write.
The $j$-th write presents a DMC with the input length-$n$ binary vector $X_j$ and the output is $(Z_{j-1}, Y_j)$, where $Z_{j-1} \in \{0,1\}^n$ represents the times each cell was programmed before the $j$-th write, and $Y_j \in \{0,1\}^n$ represent the state of the memory after the $j$-th write. Let $x_j = X_{j,k}, z_j = Z_{j-1,k}$, and $y_j = Y_{j,k}$ for some index $k$. By the random coding theorem, for $n$ large enough, the following region is achievable

$$\left\{ \{R_1, \ldots, R_t\} \mid \forall 1 \leq j \leq t, R_j \leq I(x_j; y_j) \right\}.$$

By the definitions and notations of the probabilities $p_j$, and $Q_{j,i}$,$r_i$,

$$I(x_j; (z_{j-1}, y_j)) = H(z_{j-1}, y_j) - H(z_{j-1}, y_j|x_j) = H(z_{j-1}) + H(y_j|z_{j-1}) - H(z_{j-1}, y_j|x_j) = \sum_{i=0}^{\ell} Pr(z_{j-1} = i)H(y_j|z_{j-1} = i).$$

Step (a) follows from $H((z_{j-1}, y_j|x_j)) = H(z_{j-1}|x_j)$ since $y_j$ is a function of $x_j$, $z_{j-1}$, and $H(z_{j-1}|x_j) = H(z_{j-1})$ because $z_{j-1}$ is independent on $x_j$. Step (b) is implied by $H(y_j|z_{j-1} = \emptyset) = 0$. Hence, we can achieve the region $C_{t,\ell}$ for the $\ell$-change $t$-write WDM EIP:DIA model for the $\epsilon$-error case.

The proof of the converse part is similar to the proof of this part in Theorem 2. Let $S_j, S_j, V_j, 1 \leq j \leq t$, and $L$ be defined as in the proof of the converse part in Theorem 2. Thus, exactly as proved in Theorem 2, we have $I(X_j; Y_j|V_j-1) \geq I(S_j; S_j|V_j-1), \ I(S_j; S_j|V_j-1) \geq \log(M_j) - H(p_e) - p_e \log(M_j)$, and

$$\frac{1}{n} I(X_j; Y_j|V_j-1) \leq \sum_{i=0}^{\ell-1} Pr(V_j-1, L = i)H(Y_j, L|V_j-1, L = i).$$

Now, we set $p_1 = Pr(X_j, L = 1)$. Thus, for $i < \ell$ $H(Y_j, L|V_j-1, L = i) = h(p_j)$. We also define for $i \in \{0,1\}$

$$Q_{j,i} = Pr(V_j, i = i),$$

and we note that $Q_{j,i}$ can be calculated as in Equation (4). Then

$$\log(M_j) - \epsilon_j \leq \frac{1}{n} I(X_j; Y_j|V_j-1) \leq \sum_{i=0}^{\ell-1} Pr(V_j-1, L = i)H(Y_j, L|V_j-1, L = i) = \sum_{i=0}^{\ell-1} Q_{j,i}h(p_j) = (1 - Q_{j,1})h(p_j),$$

where $\epsilon_j = \frac{H(p_e) + p_e \log(M_j)}{n}$, and the theorem is implied. 

Proof: Let $\mathbf{R} = (R_1, R_2, \ldots, R_t)$ be a rate tuple which achieves the maximum sum-rate $R_{\ell}^{EIP:DIA,\epsilon}$, and we denote by $\mathbf{R}_j$ and $Q_{j,i}$, $1 \leq j \leq t$ and $i \in \{0,1\}$, the probabilities which attain $\mathbf{R}$ in $C_{t,\ell}$.

Now we present a rate tuple $\tilde{\mathbf{R}} = (\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_t) \in \bar{C}_{t,\ell}$. Then, we conclude that $\mathbf{R} \in C_{t,\ell}^{EIP:DIA,\epsilon} \setminus C_{t,\ell}^{EIP:DIA,\epsilon}$, which implies that $R_{\ell}^{EIP:DIA,\epsilon} < R_{\ell}^{EIP:DIA,\epsilon}$ and $C_{t,\ell}^{EIP:DIA,\epsilon} \subseteq C_{t,\ell}^{EIP:DIA,\epsilon}$.

We assume now that $\ell$ is even, while the proof for the odd case is similar. Since $\tilde{\mathbf{R}}$ achieves maximum sum-rate we have $\tilde{p}_i = 0.5$. For all $j$ and $i$, $1 \leq j \leq t - 1$ and $i \in \{0\}$, we define $\tilde{p}_{j,0} = \tilde{p}_{j,1} = \tilde{p}_j$. In addition, $\tilde{p}_t, 0 = 0.5, \tilde{p}_t, 1 = \tilde{p}_{t-1}, 0 = 0.5$.

Thus, for all $j$ and $i$, $1 \leq j \leq t - 2$ and $i \in \{0\}$, $\tilde{R}_j = \tilde{R}_j$ and $\tilde{Q}_{j,i} = Q_{j,i}$.

For the $(t-1)$-th write we have, $R_{t-1} = \tilde{Q}_{t-2,0}h(\tilde{p}_{t-1}) + (1 - \tilde{Q}_{t-2,0})h(\tilde{p}_{t-1})$, and for the last write $\tilde{R}_t = \tilde{R}_t = \tilde{Q}_{t-1,1}$.

Now we prove that $\tilde{p}_{t-1} < 0.5$ which immediately implies that $R_{t-1} > R_{t-1}$ and thus completes the proof. Recall that $\tilde{R}_t = \tilde{Q}_{t-1,1} = 1 - \tilde{Q}_{t-2,0} - \tilde{Q}_{t-2,1} - \tilde{p}_{t-1}$. Thus, given the probabilities for the first $t - 2$ writes, in order to achieve the maximal rate tuple $\tilde{\mathbf{R}}$ we have to maximize $\tilde{R}_{t-1} + \tilde{R}_t$. That is, we choose $\tilde{p}_{t-1}$ which maximizes $(1 - \tilde{Q}_{t-2,0})h(\tilde{p}_{t-1}) - (1 - Q_{t-2,1} - \tilde{Q}_{t-2,1})h(\tilde{p}_{t-1})$, and the maximum is obtained for $\tilde{p}_{t-1} = 1/(1 + 2 \tilde{Q}_{t-2,1} - 1/(1 - Q_{t-2,1}))$. Since $\tilde{\mathbf{R}}$ is maximal and $t > \ell \geq 2$.

\begin{thebibliography}{1}
\end{thebibliography}
Yeow Meng Chee (Senior Member, IEEE) received the B.Math. degree in computer science and combinatorics and optimization and the M.Math. and Ph.D. degrees in computer science from the University of Waterloo, Waterloo, ON, Canada, in 1988, 1989, and 1996, respectively. He was a Professor with the School of Physical and Mathematical Sciences and the Interim Dean of science with Nanyang Technological University. He is currently a Professor with the Department of Industrial Systems Engineering and Management and the Associate Vice President of innovation and enterprise with the National University of Singapore. His research interests include the interplay between combinatorics and computer science/engineering, particularly combinatorial design theory, coding theory, extremal set systems, and electronic design automation.