Performance of Multilevel Flash Memories With Different Binary Labelings: A Multi-User Perspective

Pengfei Huang, *Student Member, IEEE*, Paul H. Siegel, *Fellow, IEEE*, and Eitan Yaakobi, *Member, IEEE*

***Abstract*— In this paper, we study the performance of different decoding schemes for multilevel flash memories where each page in every block is encoded independently. We focus on multi-level cell flash memory, which is modeled as a two-user multiple-access channel suffering from asymmetric noise. The uniform rate regions and sum rates of treating interference as noise decoding and successive cancelation decoding are investigated for a program/erase cycling model and a data retention model. We examine the effect of different binary labelings of the cell levels, as well as the impact of further quantization of the memory output (i.e., additional read thresholds). Finally, we extend our analysis to the three-level cell flash memory.**

***Index Terms*— Flash memory, multi-user system, rate region, labelings, asymmetric noise.**

I. Introduction

NAND flash memory is a versatile non-volatile data storage medium, and has been widely used in consumer electronics as well as enterprise data centers. It has many advantages over traditional magnetic recording, e.g., higher read throughput and less power consumption [3]. The basic storage unit in a NAND flash memory is a floating-gate transistor referred to as a cell. The voltage levels of a cell can be adjusted by a program operation and are used to represent the stored data. The cells typically have 2, 4, and 8 voltage levels (1, 2, and 3 bits/cell, respectively) and are referred to as single-level cell (SLC), multi-level cell (MLC), and three-level cell (TLC), respectively. Cells are organized into a rectangular array, interconnected by horizontal *wordlines* and vertical *bitlines*, that constitute a *block*. A flash memory chip comprises a collection of such blocks. During program (i.e., write) operations, the voltage level of a cell can not be decreased. In order to do so, the entire containing block must be erased and reprogrammed. Repeated program/erase (P/E) operations induce wear on the cells in the block, with a detrimental effect on the lifetime of the memory.

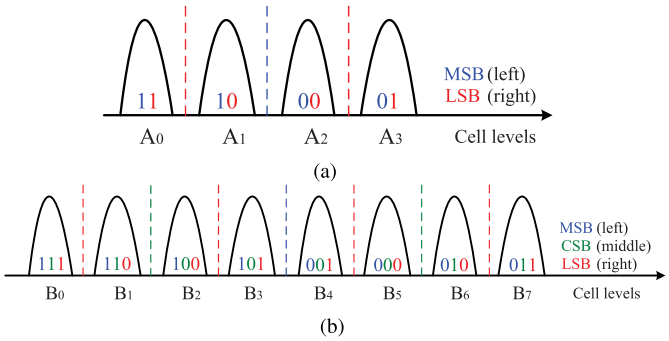


Fig. 1. (a) The four voltage levels and Gray labeling for a cell in MLC flash memories. A total of three reads are employed for decoding two pages. (b) The eight voltage levels and Gray labeling for a cell in TLC flash memories. A total of seven reads are employed for decoding three pages.

In an MLC flash memory, the two bits in the cells connected along a wordline are assigned to two separate *pages*, which represent the basic unit for program and read operations. The most significant bit (MSB) is assigned to the *lower page* while the least significant bit (LSB) is assigned to the *upper page*. We denote the four nominal voltage levels in the MLC cell as A_0, A_1, A_2, and A_3, in order of increasing voltage. The program operation is not perfect, so the actual cell levels are distributed around the nominal levels, as depicted in Fig. 1(a). Also shown in the figure is a particular mapping of 2-bit patterns, namely '11', '10', '00', and '01', to the voltage levels. We refer to this mapping as the *Gray* labeling.

Similarly, the three bits belonging to a TLC are separately mapped to three pages. We refer to the first bit as the most significant bit (MSB), the second bit as the center significant bit (CSB), and the third bit as the least significant bit (LSB). The corresponding pages are referred to as the *lower page*, *center page*, and *upper page*, respectively. Fig. 1(b) depicts the programmed cell level distributions around the eight nominal TLC voltage levels, denoted B_0, B_1, B_2, B_3, B_4, B_5, B_6, and B_7, along with the corresponding Gray labeling, '111', '110', '100', '101', '001', '000', '010', and '011'.

The channel characterization of flash memories is important for understanding fundamental limits on storage density, as well as for designing effective signal processing algorithms and error-correcting codes (ECC) [2], [11], [12], [25]. Many experiments have shown that the distribution of the readback signal (i.e., the voltage level of a cell) in flash memories is asymmetric [5], [6], [22], [28]. In [23], a mixed

Manuscript received May 2, 2016; accepted August 17, 2016. Date of publication August 26, 2016; date of current version October 11, 2016. This work was supported in part by the Center for Memory and Recording Research (CMRR) at UCSD, in part by NSF under Grant CCF-1116739 and Grant CCF-1405119, and in part by the Israel Science Foundation under Grant 1624/14. The work of E. Yaakobi was supported by CMRR at UCSD. This paper was presented at the IEEE International Symposium on Information Theory, Barcelona, Spain, July 10-15, 2016 [15].

P. Huang and P. H. Siegel are with the Center for Memory and Recording Research, Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: pehuang@ucsd.edu; psiegel@ucsd.edu).

E. Yaakobi is with the Department of Computer Science, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: yaakobi@cs.technion.ac.il).

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/JSAC.2016.2603658

normal-Laplace distribution model was proposed and shown to accurately capture this asymmetry. Several papers, e.g., [16], [21], [27], [28], [34], and [35], have recently studied the capacity of multilevel flash memory using a variety of channel models. For example, in [35], the capacity of MLC flash memory was analyzed by modeling MLC flash memory as a 4-ary input point-to-point channel with additive white Gaussian noise. The performance improvement provided by soft information obtained with multiple read thresholds – a topic to be addressed later in this paper – was also evaluated. In a similar vein, the capacity of TLC flash memory was recently studied in [27] by considering TLC flash memory as an 8-ary input point-to-point channel with asymmetric mixed normal-Laplace noise. In [28], empirical error measurements at the cell and bit levels were used to estimate the capacity of an MLC flash memory as a function of P/E cycle count. Results obtained using 4-ary discrete memoryless channel (DMC) model were compared to those based upon a union of two independent binary symmetric channel (BSC) models corresponding to the lower page and upper page.

A common feature of these prior studies of flash memory capacity is the use of point-to-point channel models. In this paper, we take a different approach, and model the flash memory as a multi-user system, where the pages correspond to independently encoded users of a shared multiple-access channel. To the best of our knowledge, this is the first time that flash memories have been examined from this perspective.

Our goal is to study the fundamental performance limits of page-oriented multilevel flash memories using various decoding schemes. Specifically, we consider both low-complexity Treating Interference as Noise (TIN) decoding and relatively high-complexity Successive Cancelation (SC) decoding for the MLC case. We first examine a general discrete memoryless multiple-access channel model with two binary inputs and a four-level output. We derive elementary conditions such that the sum rate of TIN decoding equals that of SC decoding. Then, we determine achievable rate regions and sum rates of both decoding schemes for several specific flash memory channel models, represented by channel transition matrices from cell voltage levels to quantized readback outputs. The effect of different binary labelings of the cell levels is also studied, and the optimal labelings for each decoding scheme and channel model are identified. It is shown that TIN and SC decodings both outperform Default Setting (DS) decoding, a model of current flash memory technology, which uses Gray labeling of cell levels, along with separate quantization and decoding of each page. We also study the impact of further quantization of the memory output (i.e., additional read thresholds), and the resulting effect on performance is evaluated by means of computer simulation. Finally, we generalize some of these results to the TLC case, highlighting some of the complexities that arise when the number of bits per cell increases. We remark that, although the focus of the paper is on information-theoretic analysis, some of the results provide qualitative insight into effective coding solutions.

There are conceptual connections, as well as important differences, between this work and the vast literature on multi-user communications and multilevel coded modulation for data transmission. The basic multiple-access channel model underlying many studies in the setting of wireless communications is an additive channel; that is, the received signal is the sum of all users' signals plus noise; see, for example, [4], [13], [14], [19], [31], [36]. This model has very different characteristics than the flash memory channel models that we consider, and few results are directly applicable. Multilevel coded modulation [9], [18], [20], [32] and bit-interleaved coded modulation (BICM) [8], [29], [30] have found extensive use in wireline and wireless communications. Although designed for point-to-point channels, they have been analyzed within the framework of multi-user information theory; see, for example, [7], [8], [17], [20], [29], [30], [32], [33]. It was shown in [17], [30], [32], and [29] that, on the additive white Gaussian noise (AWGN) channel, all mappings give the same sum rate under multistage decoding (i.e., SC decoding). In [32], the BICM scheme was interpreted as multilevel encoding together with parallel, independent decoding of the individual levels (i.e., TIN decoding). For PSK and QAM schemes, it was observed that the gap between the sum rates of TIN and SC decodings under Gray labeling was small over AWGN and fading channels [7], [8], [32]. As indicated by these examples, the focus of most of these studies is on channels with AWGN and fading, and the conclusions are generally not relevant to the DMC models for flash memory that we consider.

The remainder of the paper is organized as follows. In Section II, we introduce the discrete memoryless multiple-access channel model for multilevel flash memories. We then review information-theoretic characterizations of the uniform rate regions and sum rates for the three decoding schemes – TIN, SC, and DS – that we consider. We also derive some elementary but useful characterizations of those channels for which TIN and SC decoders yield the same rate regions, as well as an elementary general bound on the difference between the respective sum rates of TIN and SC decoders. In Section III, we analyze properties and relationships among the rate regions for various MLC flash memory channel models. We consider the effect of using different binary labelings of cell voltage levels, as well as the impact of employing additional read thresholds to obtain refined soft information. In Section IV, we extend the analysis to TLC flash memories, highlighting some of the complexities and qualitative differences encountered when considering more bits per cell. We conclude the paper in Section V.

Remark on notation: Throughout the paper, we follow the notation in [13]. Random variables are denoted with upper case letters (e.g., X) and their realizations with lower case (e.g., x). Calligraphic letters (e.g., $\mathcal{X}$) are used for finite sets. $\mathbb{R}$ is the set of real numbers. The discrete interval $[i : j]$ is defined as the set $\{i, i+1, \ldots, j\}$. For a length-n vector $\boldsymbol{v}$, $\boldsymbol{v}(i)$ represents the value of its ith coordinate, $i = 1, 2, \ldots, n$. We use the notation $p(x)$ to abbreviate the probability $P(X = x)$, and likewise for conditional and joint probabilities of both scalar and vector random variables, e.g., $p(y|x) = P(Y = y|X = x)$. For a probability vector $(\frac{p_1}{p_s}, \frac{p_2}{p_s}, \ldots, \frac{p_n}{p_s})$ where $p_s = \sum_{i=1}^{n} p_i$ and $p_i \geq 0$, $i = 1, 2, \ldots, n$, the entropy function is defined by $H(\frac{p_1}{p_s}, \frac{p_2}{p_s}, \ldots, \frac{p_n}{p_s}) = -\sum_{i=1}^{n} \frac{p_i}{p_s} \log_2 \frac{p_i}{p_s}$. We will also use

the function $f(x) = x\log_2 x$. Therefore, we can express the entropy function by $H(\frac{p_1}{p_s}, \frac{p_2}{p_s}, \ldots, \frac{p_n}{p_s}) = -\frac{1}{p_s}\sum_{i=1}^{n} f(p_i) + \log_2 p_s$.

II. Multiple-Access Channel Model for Flash Memories

In this section, we introduce the multiple-access channel model for multilevel flash memories, define the decoding schemes to be considered, and present some of their basic information-theoretic properties.

A. System Model

We model a multilevel flash memory as a k-user multiple-access channel with k independent inputs $X_1, \cdots, X_k$, and one output Y ($k = 2$ for MLC flash, and $k = 3$ for TLC flash).

Specifically, the readback signal $\tilde{Y} \in \mathbb{R}$ in a flash memory is expressed as

$$\tilde{Y} = \sigma(X_1, \cdots, X_k) + Z, \tag{1}$$

where $X_1, \cdots, X_k \in \{0, 1\}$ represent data from k independent pages, $Z \in \mathbb{R}$ stands for the asymmetric noise (see [23] for more details on the normal-Laplace distribution model), and σ maps an input $(x_1, \cdots, x_k)$ to a voltage level v. More specifically, σ is a bijective mapping from the set $\mathcal{T}$ which consists of all length-k binary strings to the set $\mathcal{V}$ which consists of 2^k voltage level values. For $k = 2$ (MLC flash), $\mathcal{T}_{MLC} = \{11, 10, 01, 00\}$ and $\mathcal{V}_{MLC} = \{A_0, A_1, A_2, A_3\}$. By a slight abuse of notation, we write the mapping σ as a vector $\sigma = (w_0, w_1, w_2, w_3)$ (where w_i, $i = 0, 1, 2, 3$, represent the full set of possible 2-tuples) to represent the mapping $\sigma(w_i) = A_i$ for $i = 0, 1, 2, 3$. For example, the vector $\sigma = (11, 10, 00, 01)$ corresponds to $\sigma(11) = A_0$, $\sigma(10) = A_1$, $\sigma(00) = A_2$, and $\sigma(01) = A_3$. Similarly, for $k = 3$ (TLC flash), $\mathcal{T}_{TLC} = \{111, 110, 101, 100, 011, 010, 001, 000\}$ and $\mathcal{V}_{TLC} = \{B_0, B_1, \ldots, B_7\}$. We write $\sigma = (w_0, w_1, \ldots, w_7)$ (where w_i, $i = 0, 1, \ldots, 7$, represent the full set of possible 3-tuples) to represent the mapping $\sigma(w_i) = B_i$ for $i = 0, 1, \ldots, 7$. We will refer to a *mapping* σ as a *labeling*.

During the readback process, a quantizer Q is used to quantize $\tilde{Y}$ to obtain an output Y, i.e., $Y = Q(\tilde{Y})$, where the function $Q(\cdot)$ is a mapping from $\mathbb{R}$ to a finite alphabet set $\mathcal{Y} = \{s_0, s_1, \ldots, s_{q-1}\}$ of cardinality q. Usually $q = 2^k$, but this is not necessary. The cardinality q can correspond to a large number by applying multiple reads. From an information-theoretic point of view, this means that more soft information is obtained for decoding.

B. Decoding Schemes for MLC Flash Memories

In this subsection, we investigate three decoding schemes for MLC flash memories.

Given a labeling σ and a quantizer Q, the MLC flash memory channel can be modeled as a 2-user discrete memoryless multiple-access channel $\mathcal{W}_{MLC}$: $(\mathcal{X} \times \mathcal{X}, p(y|x_1, x_2), \mathcal{Y})$, where $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{s_0, s_1, \ldots, s_{q-1}\}$, and $p(y|x_1, x_2)$ is the transition probability for any $x_1, x_2 \in \mathcal{X}$ and $y \in \mathcal{Y}$. For simplicity, denote the conditional probabilities by $p_{BD(x_1,x_2),y} \stackrel{\text{def}}{=} P(Y = y|X_1 = x_1, X_2 = x_2)$, where $BD(\cdot)$ is a function that converts a binary string into its decimal value; e.g., $p_{2,s_0} = P(Y = s_0|X_1 = 1, X_2 = 0)$.

Users $j = 1, 2$ independently encode their messages M_j into the corresponding length-n codewords x_j^n and send (write) them over the shared channel (i.e., a set of cells) to the receiver (reader). Following the notation in [13], we define a $(2^{nR_1}, 2^{nR_2}, n)$ code by two encoders $x_1^n(m_1)$ and $x_2^n(m_2)$ for messages m_1 and m_2 from message sets $[1 : 2^{nR_1}]$ and $[1 : 2^{nR_2}]$ respectively, and a decoder that assigns an estimate $(\hat{m}_1, \hat{m}_2)$ based on the received sequence y^n. We assume that the message pair (M_1, M_2) is uniform over $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$. The average probability of error is defined as $\mathrm{P}_e^{(n)} = P\{(M_1, M_2) \neq (\hat{M}_1, \hat{M}_2)\}$. A rate pair (R_1, R_2) is said to be achievable if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes such that $\lim_{n\to\infty} \mathrm{P}_e^{(n)} = 0$. The capacity region is the closure of the set of achievable rate pairs (R_1, R_2).

The capacity region of this multiple-access channel is fully characterized [10], [13]. However, in this paper, to make the analysis simple and yet representative, we focus on the *uniform* rate region for different decoding schemes. This represents the achievable region corresponding to the case that the input distributions are uniform. For other input distributions, the analysis is similar.

For a channel $\mathcal{W}_{MLC}$, the **Treating Interference as Noise (TIN)** decoding scheme decodes X_1 and X_2 independently based on Y [10], [13]. Its uniform rate region $\mathscr{R}^{TIN}$ for lower page X_1 and upper page X_2 is the set of all pairs (R_1, R_2) such that $R_1 \leq I(X_1; Y)$ and $R_2 \leq I(X_2; Y)$. In $\mathscr{R}^{TIN}$, the sum rate[1] is $r_s^{TIN} = \max\{R_1 + R_2 : (R_1, R_2) \in \mathscr{R}^{TIN}\} = I(X_1; Y) + I(X_2; Y)$.

For a channel $\mathcal{W}_{MLC}$, the **Successive Cancelation (SC)** decoding scheme decodes X_1 and X_2 in some order based on Y [10], [13]. Its uniform rate region $\mathscr{R}^{SC}$ for lower page X_1 and upper page X_2 is the set of all pairs (R_1, R_2) such that $R_1 \leq I(X_1; Y|X_2)$, $R_2 \leq I(X_2; Y|X_1)$, and $R_1 + R_2 \leq I(X_1, X_2; Y)$. In $\mathscr{R}^{SC}$, the sum rate is $r_s^{SC} = \max\{R_1 + R_2 : (R_1, R_2) \in \mathscr{R}^{SC}\} = I(X_1, X_2; Y)$.

Remark 1: For TIN decoding, X_1 and X_2 are decoded independently and can be implemented in parallel. However, for SC decoding, X_1 and X_2 are decoded in a certain order. In general, TIN decoding is preferred for its low decoding complexity, but there may be a cost in performance relative to SC decoding, as reflected in the uniform rate region containment $\mathscr{R}^{TIN} \subseteq \mathscr{R}^{SC}$ and the sum rate relationship $r_s^{TIN} \leq r_s^{SC}$. □

The following theorem identifies the channels for which the sum rates of TIN decoding and SC decoding are the same.

Theorem 1: For a channel $\mathcal{W}_{MLC}$, the sum rates satisfy $r_s^{TIN} \leq r_s^{SC}$ with equality if and only if $p_{3,s_j} p_{0,s_j} = p_{2,s_j} p_{1,s_j}$ for all $j = 0, 1, \ldots, q-1$. If $r_s^{TIN} = r_s^{SC}$, then $\mathscr{R}^{TIN} = \mathscr{R}^{SC}$ and the rate region is a rectangle.

[1] In this paper, for the sake of brevity, we use the term "sum rate" to represent the maximum sum rate in the corresponding rate region.

Proof: See Appendix A. ∎

An upper bound on the difference between r_s^{SC} and r_s^{TIN} is given by the following theorem.

Theorem 2: For a channel $\mathcal{W}_{MLC}$, the rate difference $r_s^{SC} - r_s^{TIN} \leq 1$ with equality if and only if $p_{3,s_j} + p_{2,s_j} = p_{1,s_j} + p_{0,s_j}$ and $p_{3,s_j} p_{1,s_j} = p_{2,s_j} p_{0,s_j} = 0$ for all $j = 0, 1, \ldots, q-1$.

Proof: See Appendix B. ∎

From the proof of Theorem 2, it is easy to see that $r_s^{SC} - r_s^{TIN} = 1$ is satisfied if and only if $I(X_1; Y) = I(X_2; Y) = 0$ and $I(X_1, X_2; Y) = 1$. Thus, in this case, if we use TIN decoding, no information can be reliably stored in the memory (i.e., $I(X_1; Y) + I(X_2; Y) = 0$), whereas if we use SC decoding, one bit of information can be reliably stored in each cell (i.e., $I(X_1, X_2; Y) = 1$).

The third decoding scheme we consider is modeled upon current MLC flash memory technology. For this scheme, the Gray labeling $\sigma = (11, 10, 00, 01)$ is used to map binary inputs (X_1, X_2) to cell levels V. The lower page X_1 and upper page X_2 are decoded independently according to different quantization rules and a total of three reads are employed. To decode X_1, $\tilde{Y}$ is quantized by one read between voltage levels A_1 and A_2 (see Fig. 1(a)), and the corresponding output is Y_1. To decode X_2, $\tilde{Y}$ is quantized by two reads between voltage levels A_0 and A_1, and between A_2 and A_3, respectively (see Fig. 1(a)), and the corresponding output is Y_2. We call this **Default Setting (DS)** decoding, and it is used as our baseline decoding scheme. Its uniform rate region $\mathscr{R}^{DS}$ for the lower page X_1 and the upper page X_2 is the set of all pairs (R_1, R_2) such that $R_1 \leq I(X_1; Y_1)$ and $R_2 \leq I(X_2; Y_2)$. In $\mathscr{R}^{DS}$, the sum rate is $r_s^{DS} = \max\{R_1 + R_2 : (R_1, R_2) \in \mathscr{R}^{DS}\} = I(X_1; Y_1) + I(X_2; Y_2)$.

III. Performance of MLC Flash Memory With Different Decoding Schemes and Labelings

In this section, we study the uniform rate region and sum rate of several MLC flash memory channel models with different decoding schemes and labelings. The channel models, inspired by empirical observation of flash memory behavior, are defined by channel transition matrices relating voltage levels to quantized outputs. Specifically, we consider *program/erase (P/E) cycling* models (early-stage and late-stage) and a *data retention* model. The P/E cycling model is used to characterize errors caused by inter-cell interference and the wear induced by repeated program/erase operations. The data retention model is used to characterize retention errors resulting from charge leakage over time from programmed cells. Note that although we concentrate on these particular models here, similar analysis can be applied to other relevant models.

Among the $4! = 24$ possible binary labelings of the 4 nominal cell voltage levels, we initially consider 3 representative examples.

Definition 3: For MLC flash memory, the mapping σ_G=(11, 10, 00, 01) is called Gray labeling, σ_{NO}=(11, 10, 01, 00) is called Natural Order (NO) labeling, and σ_{EO}=(11, 00, 01, 10) is called Even Odd (EO) labeling.

TABLE I

Channel Transition Matrix $p_{MLC}^{E}(y|v)$ at Early-Stage of P/E Cycling for MLC Flash Memories

V	Inputs: (X_1, X_2)			Output: Y			
Levels	Gray	NO	EO	s_0	s_1	s_2	s_3
A_0	(11)	(11)	(11)	a_1	$1-a_1$	0	0
A_1	(10)	(10)	(00)	0	b_1	$1-b_1$	0
A_2	(00)	(01)	(01)	0	0	c_1	$1-c_1$
A_3	(01)	(00)	(10)	0	0	0	1

For each of these three labelings, the mapping between inputs $(X_1, X_2) \in \mathcal{I}_{MLC}$ and voltage levels $V \in \mathcal{V}_{MLC}$ is shown in Table I.

Remark 2: It is a standard practice in current MLC flash technology to program the lower page and upper page sequentially in a 2-step procedure [23]. The cell voltage level is initially set to reflect the value of the lower bit, with a '1' corresponding to the lowest level and a '0' corresponding to an intermediate level. When programming the upper bit, this voltage level is increased to reach the desired level corresponding to the binary labeling. Recall that a programming operation can not decrease the cell level. Clearly this procedure is only compatible with labelings in which two lower voltage levels share the same lower bit value, and likewise for the two upper voltage levels. The Gray labeling, which is used in practice, and the NO labeling satisfy this property, but the EO labeling does not.

The 2-step programming process also influences the behavior of the flash memory channel. The early-stage P/E cycling model and data retention model are largely independent of the programming method, but the late-stage P/E cycling model includes the effect of an incorrectly programmed lower bit on the programming of the upper bit. Hence, when analyzing the performance of any MLC flash memory labeling, we will assume that the 2-bit labels of the labeling under consideration are mapped to the corresponding labels in the Gray labeling, to which the 2-step programming process is then applied. This approach can be extended to higher density flash memory, such as TLC, as well. □

A. Quantization With Three Reads

In this subsection, we assume the quantizer Q uses three reads, placed between every pair of adjacent voltage levels, as shown in Fig. 1(a). Hence, the output alphabet $\mathcal{Y}_{MLC} = \{s_0, s_1, s_2, s_3\}$. For DS decoding, we assume that the output alphabet for the lower page X_1 is $\mathcal{Y}_{MLC}^1 = \{s_{0\cup1}, s_{2\cup3}\}$ of cardinality two, and the output alphabet for the upper page X_2 is $\mathcal{Y}_{MLC}^2 = \{s_0, s_{1\cup2}, s_3\}$ of cardinality three.[2] Thus, DS decoding also requires a total of three reads.

A.1 Performance of Gray, NO, and EO Labelings

We study the performance of MLC flash memories using the P/E cycling model, which has different channel characteristics for early and late stages of the memory lifetime.

[2]We use the notation $s_{u\cup v}$ to represent an output obtained by merging two outputs s_u and s_v in $\mathcal{Y}_{MLC}$, i.e., $P(Y = s_{u\cup v}|X_1 = x_1, X_2 = x_2) = \sum_{i\in\{u,v\}} P(Y = s_i|X_1 = x_1, X_2 = x_2)$ for any $x_1, x_2 \in \{0, 1\}$. Strictly speaking, for the upper page decoding, current MLC flash memories use an output alphabet $\mathcal{Y}_{MLC}^2 = \{s_{1\cup2}, s_{0\cup3}\}$. The resulting performance cannot exceed that obtained with the output alphabet $\{s_0, s_{1\cup2}, s_3\}$ used in this paper.

Early-Stage P/E Cycling Model: The early-stage P/E cycling channel transition matrix $p^E_{MLC}(y|v)$, for output $y \in \mathcal{Y}_{MLC}$ and voltage level $v \in \mathcal{V}_{MLC}$, reflects empirical results in [28] and is shown in Table I, where a_1, $1-a_1$, b_1, $1-b_1$, c_1, and $1-c_1$ represent non-zero probabilities.

As shown in Table I, note that the transition probability from inputs (X_1, X_2) to output Y depends upon the labeling which maps inputs to voltage levels, as well as the channel transition matrix from voltage levels to output.

Lemma 4: For channel transition matrix $p^E_{MLC}(y|v)$, using Gray labeling, we have $r_s^{TIN} = r_s^{SC}$ and $\mathscr{R}^{TIN}$=$\mathscr{R}^{SC}$. Using either NO labeling or EO labeling, we have $r_s^{TIN} < r_s^{SC}$.

Proof: With Gray labeling, in Table I, for the column $Y = s_0$, we have $p_{0,s_0} = 0$, $p_{1,s_0} = 0$, $p_{2,s_0} = 0$, and $p_{3,s_0} = a_1$. Thus, $p_{3,s_0}p_{0,s_0} = p_{2,s_0}p_{1,s_0}$. We can also verify $p_{3,s_i}p_{0,s_i} = p_{2,s_i}p_{1,s_i}$ for $i = 1, 2, 3$. Thus, from Theorem 1, we conclude $r_s^{TIN} = r_s^{SC}$ and $\mathscr{R}^{TIN} = \mathscr{R}^{SC}$. On the other hand, under NO labeling $p_{3,s_2}p_{0,s_2} \neq p_{2,s_2}p_{1,s_2}$, and under EO labeling $p_{3,s_3}p_{0,s_3} \neq p_{2,s_3}p_{1,s_3}$. Thus, from Theorem 1, for these two labelings, $r_s^{TIN} < r_s^{SC}$. ∎

Next, we calculate and compare the uniform rate regions and sum rates for the three decoding schemes under the Gray, NO, and EO labelings. The results are shown in Table II, where λ_1, λ_2, λ_3, λ_4, and λ_5 are given by

$$\lambda_1 = \frac{f(1-b_1) - f(3-b_1)}{4} + \frac{3}{2},$$

$$\lambda_2 = 1 + \frac{1}{4}\Big(f(1-a_1) + f(1+c_1) + f(1-c_1)\Big) - \frac{1}{4}\Big(f(2-c_1) + f(2-a_1+c_1)\Big),$$

$$\lambda_3 = 1 + \frac{1}{4}\Big(f(1-b_1) + f(c_1) - f(1-b_1+c_1)\Big),$$

$$\lambda_4 = 1 + \frac{1}{4}\Big(f(1-a_1) + f(b_1) + f(1-c_1)\Big) - \frac{1}{4}\Big(f(1-a_1+b_1) + f(2-c_1)\Big),$$

$$\lambda_5 = 1 - \frac{1}{4}\Big(f(1-a_1+b_1) + f(2-c_1) + f(1-b_1+c_1)\Big) + \frac{1}{4}\Big(f(1-a_1) + f(c_1) + f(1-c_1) + f(b_1) + f(1-b_1)\Big).$$

As an example, we show how to calculate λ_1; the quantities λ_2, λ_3, λ_4, and λ_5 can be obtained in a similar manner. We see that λ_1 corresponds to $I(X_1; Y_1)$ under DS decoding. Referring to Table I, we see that

$$\begin{aligned} &P(Y_1 = s_{0\cup1}|X_1 = 1) \\ &= \frac{P(Y_1 = s_{0\cup1}, X_1 = 1)}{P(X_1 = 1)} \\ &= \frac{\sum_{i=0}^{1} P(Y_1 = s_{0\cup1}, X_1 = 1, X_2 = i)}{P(X_1 = 1)} \\ &= \frac{\sum_{i=0}^{1} P(Y_1 = s_{0\cup1}|X_1 = 1, X_2 = i)P(X_1 = 1, X_2 = i)}{P(X_1 = 1)} \\ &= \frac{1+b_1}{2}. \end{aligned}$$

TABLE II

UNIFORM RATE REGIONS AND SUM RATES OF DS, TIN, AND SC DECODINGS AT EARLY-STAGE OF P/E CYCLING FOR MLC FLASH MEMORIES

Gray	DS	$\mathscr{R}_G^{DS}$	$0 \leqslant R_1 \leqslant \lambda_1,\ 0 \leqslant R_2 \leqslant \lambda_2$	$r_{s(G)}^{DS} = \lambda_1 + \lambda_2$
	TIN	$\mathscr{R}_G^{TIN}$	$0 \leqslant R_1 \leqslant \lambda_3,\ 0 \leqslant R_2 \leqslant \lambda_4$	$r_{s(G)}^{TIN} = \lambda_3 + \lambda_4$
	SC	$\mathscr{R}_G^{SC}$	$0 \leqslant R_1 \leqslant \lambda_3,\ 0 \leqslant R_2 \leqslant \lambda_4$	$r_{s(G)}^{SC} = \lambda_3 + \lambda_4$
NO	TIN	$\mathscr{R}_{NO}^{TIN}$	$0 \leqslant R_1 \leqslant \lambda_3,\ 0 \leqslant R_2 \leqslant \lambda_5$	$r_{s(NO)}^{TIN} = \lambda_3 + \lambda_5$
	SC	$\mathscr{R}_{NO}^{SC}$	$0 \leqslant R_1 \leqslant 1,\ 0 \leqslant R_2 \leqslant \lambda_4,\ R_1 + R_2 \leqslant 1 + \lambda_5$	$r_{s(NO)}^{SC} = 1 + \lambda_5$
EO	TIN	$\mathscr{R}_{EO}^{TIN}$	$0 \leqslant R_1 \leqslant \lambda_4,\ 0 \leqslant R_2 \leqslant \lambda_5$	$r_{s(EO)}^{TIN} = \lambda_4 + \lambda_5$
	SC	$\mathscr{R}_{EO}^{SC}$	$0 \leqslant R_1 \leqslant 1,\ 0 \leqslant R_2 \leqslant \lambda_3,\ R_1 + R_2 \leqslant 1 + \lambda_5$	$r_{s(EO)}^{SC} = 1 + \lambda_5$

Similarly, we calculate $P(Y_1 = s_{2\cup3}|X_1 = 1) = \frac{1-b_1}{2}$, $P(Y_1 = s_{0\cup1}|X_1 = 0) = 0$, and $P(Y_1 = s_{2\cup3}|X_1 = 0) = 1$. Noting that $P(Y_1 = s_{0\cup1}) = \frac{1+b_1}{4}$ and $P(Y_1 = s_{2\cup3}) = \frac{3-b_1}{4}$, we get

$$\begin{aligned} \lambda_1 &= I(X_1; Y_1) = H(Y_1) - H(Y_1|X_1) \\ &= H\Big(\frac{1+b_1}{4}, \frac{3-b_1}{4}\Big) - \frac{1}{2}H\Big(\frac{1+b_1}{2}, \frac{1-b_1}{2}\Big) \\ &= -\frac{1}{4}\Big(f(1+b_1) + f(3-b_1)\Big) + 2 \\ &\quad -\frac{1}{2}\Big(-\frac{1}{2}\big(f(1+b_1) + f(1-b_1)\big) + 1\Big) \\ &= \frac{f(1-b_1) - f(3-b_1)}{4} + \frac{3}{2}. \end{aligned}$$

From Table II, we have the following rate region and sum rate comparisons for the Gray, NO, and EO labelings. We denote the uniform rate region under DS decoding and Gray labeling by the term $\mathscr{R}_G^{DS}$, where the superscript and subscript represent decoding scheme and labeling, respectively. The same holds for other terms in Table II.

Theorem 5: With channel transition matrix $p^E_{MLC}(y|v)$, the rate regions satisfy $\mathscr{R}_G^{DS} \subset \mathscr{R}_G^{TIN}$, $\mathscr{R}_{NO}^{TIN} \subset \mathscr{R}_G^{TIN}$, and $\mathscr{R}_G^{SC} \subset \mathscr{R}_{NO}^{SC}$. For the sum rates, we have $r_{s(G)}^{TIN} > r_{s(G)}^{DS}$, $r_{s(G)}^{TIN} > r_{s(NO)}^{TIN}$, $r_{s(G)}^{TIN} > r_{s(EO)}^{TIN}$, and $r_{s(G)}^{SC} = r_{s(NO)}^{SC} = r_{s(EO)}^{SC}$.

Proof: Referring to Table II, we only need to show that $\lambda_3 > \lambda_1$, $\lambda_4 > \lambda_2$, $\lambda_4 > \lambda_5$, $\lambda_3 > \lambda_5$, $1 > \lambda_3$, and $\lambda_3 + \lambda_4 = 1 + \lambda_5$. Here we only give proofs of $\lambda_3 > \lambda_1$ and $\lambda_4 > \lambda_5$. Other relationships can be proved in a similar way.

We have $\lambda_3 - \lambda_1 = \frac{1}{4}\Big(f(c_1) + f(3-b_1) - f(1-b_1+c_1)\Big) - \frac{1}{2}$. Let $h_1(b_1, c_1) = f(c_1) + f(3-b_1) - f(1-b_1+c_1)$. For $0 < b_1 < 1$ and $0 < c_1 < 1$, we have $\frac{\partial h_1(b_1,c_1)}{\partial b_1} = \log_2(1-b_1+c_1) - \log_2(3-b_1) < 0$. Thus, for $0 < b_1 < 1$ and $0 < c_1 < 1$, we have $h_1(b_1, c_1) > h_1(b_1 = 1, c_1) = 2$, so $\lambda_3 > \lambda_1$.

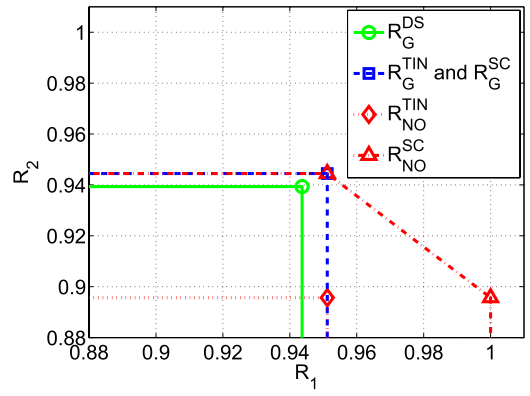
For $0 < b_1 < 1$ and $0 < c_1 < 1$, we have $f(1-b_1) + f(c_1) - f(1-b_1+c_1) = (1-b_1)\log_2\frac{1-b_1}{1-b_1+c_1} + c_1\log_2\frac{c_1}{1-b_1+c_1} < 0$, so

$$\lambda_4 - \lambda_5 = -\frac{1}{4}\Big(f(1-b_1) + f(c_1) - f(1-b_1+c_1)\Big) > 0.$$

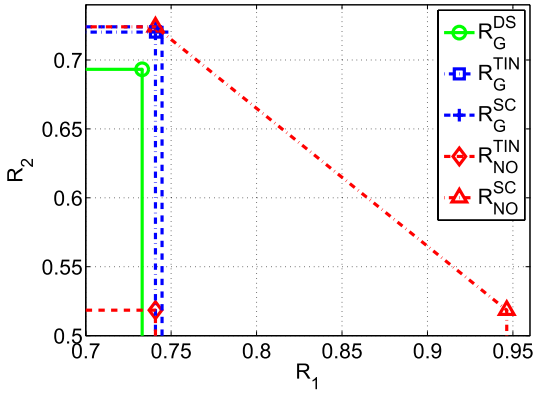
∎

Example 1: For the early-stage P/E cycling model in Table I, let $a_1 = 0.98$, $b_1 = 0.97$, and $c_1 = 0.99$.[3]

[3]Since the channel transition matrix varies from different flash chip vendors, the channel parameters are chosen to help visualize the relationship among the rate regions. For other choices of channel parameters, the relative positions of the rate regions and qualitative conclusions stay the same.



(a)



(b)

Fig. 2. (a) Uniform rate regions under Gray and NO labelings with $a_1 = 0.98$, $b_1 = 0.97$, and $c_1 = 0.99$ for the early-stage P/E cycling model. (b) Uniform rate regions under Gray and NO labelings with $\hat{a}_1 = 0.82$, $\hat{a}_2 = 0.1$, $\hat{b}_1 = 0.85$, and $\hat{c}_1 = 0.85$ for the late-stage P/E cycling model.

TABLE III

CHANNEL TRANSITION MATRIX $p^L_{MLC}(y|v)$ AT LATE-STAGE OF P/E CYCLING FOR MLC FLASH MEMORIES

V	Inputs: (X_1, X_2)			Output: Y			
Levels	Gray	NO	EO	s_0	s_1	s_2	s_3
A_0	(11)	(11)	(11)	$\hat{a}_1$	$\hat{a}_2$	0	$1-\hat{a}_1-\hat{a}_2$
A_1	(10)	(10)	(00)	0	$\hat{b}_1$	$1-\hat{b}_1$	0
A_2	(00)	(01)	(01)	0	0	$\hat{c}_1$	$1-\hat{c}_1$
A_3	(01)	(00)	(10)	0	0	0	1

The uniform rate regions under Gray and NO labelings are plotted in Fig. 2(a). It can be seen that $\mathscr{R}^{TIN}_{NO} \subset \mathscr{R}^{TIN}_{G} = \mathscr{R}^{SC}_{G} \subset \mathscr{R}^{SC}_{NO}$. The SC decoding with NO labeling gives the largest rate region.

Late-Stage P/E Cycling Model: The late-stage P/E cycling channel transition matrix $p^L_{MLC}(y|v)$, for output $y \in \mathcal{Y}_{MLC}$ and voltage level $v \in \mathcal{V}_{MLC}$, reflects measurements in [28] and its structure is shown in Table III where $\hat{a}_1$, $\hat{a}_2$, $1-\hat{a}_1-\hat{a}_2$, $\hat{b}_1$, $1-\hat{b}_1$, $\hat{c}_1$, and $1-\hat{c}_1$ represent non-zero probabilities.

Lemma 6: For channel transition matrix $p^L_{MLC}(y|v)$, we have $r_s^{TIN} < r_s^{SC}$ with Gray labeling, NO labeling or EO labeling.

Proof: For any of the three labelings, consider the column $Y = s_3$ in Table III. Since three of p_{0,s_3}, p_{1,s_3}, p_{2,s_3}, and p_{3,s_3} are positive, it is impossible to satisfy $p_{3,s_3}p_{0,s_3} = p_{2,s_3}p_{1,s_3} = 0$. Thus, from Theorem 1, we conclude $r_s^{TIN} < r_s^{SC}$. ■

TABLE IV

UNIFORM RATE REGIONS AND SUM RATES OF DS, TIN, AND SC DECODINGS AT LATE-STAGE OF P/E CYCLING FOR MLC FLASH MEMORIES

Gray	DS	$\mathscr{R}_G^{DS}$	$0 \leqslant R_1 \leqslant \tau_1,\ 0 \leqslant R_2 \leqslant \tau_2$	$r^{DS}_{s(G)} = \tau_1 + \tau_2$
	TIN	$\mathscr{R}_G^{TIN}$	$0 \leqslant R_1 \leqslant \tau_3,\ 0 \leqslant R_2 \leqslant \tau_4$	$r^{TIN}_{s(G)} = \tau_3 + \tau_4$
	SC	$\mathscr{R}_G^{SC}$	$0 \leqslant R_1 \leqslant \tau_5,\ 0 \leqslant R_2 \leqslant \tau_6,\ R_1 + R_2 \leqslant \tau_4 + \tau_5$	$r^{SC}_{s(G)} = \tau_4 + \tau_5$
NO	TIN	$\mathscr{R}_{NO}^{TIN}$	$0 \leqslant R_1 \leqslant \tau_3,\ 0 \leqslant R_2 \leqslant \tau_7$	$r^{TIN}_{s(NO)} = \tau_3 + \tau_7$
	SC	$\mathscr{R}_{NO}^{SC}$	$0 \leqslant R_1 \leqslant \tau_8,\ 0 \leqslant R_2 \leqslant \tau_6,\ R_1 + R_2 \leqslant \tau_7 + \tau_8$	$r^{SC}_{s(NO)} = \tau_7 + \tau_8$
EO	TIN	$\mathscr{R}_{EO}^{TIN}$	$0 \leqslant R_1 \leqslant \tau_4,\ 0 \leqslant R_2 \leqslant \tau_7$	$r^{TIN}_{s(EO)} = \tau_4 + \tau_7$
	SC	$\mathscr{R}_{EO}^{SC}$	$0 \leqslant R_1 \leqslant \tau_8,\ 0 \leqslant R_2 \leqslant \tau_5,\ R_1 + R_2 \leqslant \tau_7 + \tau_8$	$r^{SC}_{s(EO)} = \tau_7 + \tau_8$

Next, we calculate the uniform rate regions and the sum rates of the three decoding schemes under different labelings. The results are shown in Table IV, where τ_i, $i = 1, \ldots, 8$, are given by

$$\tau_1 = \frac{f(2-\hat{a}_1-\hat{a}_2-\hat{b}_1) - f(4-\hat{a}_1-\hat{a}_2-\hat{b}_1)}{4} + \frac{3}{2},$$

$$\tau_2 = 1 + \frac{1}{4}\Big(f(\hat{a}_2) + f(2-\hat{a}_1-\hat{a}_2) + f(1+\hat{c}_1) + f(1-\hat{c}_1)\Big) - \frac{1}{4}\Big(f(3-\hat{a}_1-\hat{a}_2-\hat{c}_1) + f(\hat{a}_2+\hat{c}_1+1)\Big),$$

$$\tau_3 = 1 + \frac{1}{4}\Big(f(1-\hat{b}_1) + f(\hat{c}_1) + f(1-\hat{a}_1-\hat{a}_2) + f(2-\hat{c}_1)\Big) - \frac{1}{4}\Big(f(1-\hat{b}_1+\hat{c}_1) + f(3-\hat{a}_1-\hat{a}_2-\hat{c}_1)\Big),$$

$$\tau_4 = 1 + \frac{1}{4}\Big(f(\hat{a}_2) + f(2-\hat{a}_1-\hat{a}_2) + f(\hat{b}_1) + f(1-\hat{c}_1)\Big) - \frac{1}{4}\Big(f(\hat{a}_2+\hat{b}_1) + f(3-\hat{a}_1-\hat{a}_2-\hat{c}_1)\Big),$$

$$\tau_5 = 1 + \frac{1}{4}\Big(f(1-\hat{a}_1-\hat{a}_2) + f(1-\hat{b}_1) + f(\hat{c}_1)\Big) - \frac{1}{4}\Big(f(2-\hat{a}_1-\hat{a}_2) + f(1-\hat{b}_1+\hat{c}_1)\Big),$$

$$\tau_6 = 1 + \frac{1}{4}\Big(f(\hat{a}_2) + f(\hat{b}_1) + f(1-\hat{c}_1)\Big) - \frac{1}{4}\Big(f(\hat{a}_2+\hat{b}_1) + f(2-\hat{c}_1)\Big),$$

$$\tau_7 = 1 - \frac{1}{4}\Big(f(\hat{a}_2+\hat{b}_1) + f(1-\hat{b}_1+\hat{c}_1) + f(3-\hat{a}_1-\hat{a}_2-\hat{c}_1)\Big) + \frac{1}{4}\Big(f(\hat{a}_2) + f(\hat{c}_1) + f(2-\hat{a}_1-\hat{a}_2-\hat{c}_1) + f(\hat{b}_1) + f(1-\hat{b}_1)\Big),$$

$$\tau_8 = 1 + \frac{1}{4}\Big(f(1-\hat{c}_1) + f(1-\hat{a}_1-\hat{a}_2) - f(2-\hat{a}_1-\hat{a}_2-\hat{c}_1)\Big).$$

From Table IV, we can infer the following rate region and sum rate relationships.

Theorem 7: With channel transition matrix $p^L_{MLC}(y|v)$, the rate regions satisfy $\mathscr{R}^{DS}_G \subset \mathscr{R}^{TIN}_G$, $\mathscr{R}^{TIN}_{NO} \subset \mathscr{R}^{TIN}_G$, and $\mathscr{R}^{SC}_G \subset \mathscr{R}^{SC}_{NO}$. For the sum rates, we have $r^{TIN}_{s(G)} > r^{DS}_{s(G)}$, $r^{TIN}_{s(G)} > r^{TIN}_{s(NO)}$, $r^{TIN}_{s(G)} > r^{TIN}_{s(EO)}$, and $r^{SC}_{s(G)} = r^{SC}_{s(NO)} = r^{SC}_{s(EO)}$.

Proof: Using Table IV, we only need to show that $\tau_3 \geq \tau_1$, $\tau_4 > \tau_2$, $\tau_4 > \tau_7$, $\tau_8 > \tau_5$, $\tau_3 > \tau_7$, and $\tau_4+\tau_5 = \tau_7+\tau_8$. Here, we give proofs of $\tau_3 \geq \tau_1$ and $\tau_4 > \tau_2$. The other relationships can be proved in a similar way. First, we compute

$$\begin{aligned}4(\tau_3-\tau_1) &= f(1-\hat{b}_1)+f(\hat{c}_1)+f(1-\hat{a}_1-\hat{a}_2)+f(2-\hat{c}_1)\\ &\quad +f(4-\hat{a}_1-\hat{a}_2-\hat{b}_1)-f(1-\hat{b}_1+\hat{c}_1)\\ &\quad -f(3-\hat{a}_1-\hat{a}_2-\hat{c}_1)-f(2-\hat{a}_1-\hat{a}_2-\hat{b}_1)\\ &\quad -f(2)\\ &= (1-\hat{b}_1+\hat{c}_1)\log_2(1+\frac{3-\hat{a}_1-\hat{a}_2-\hat{c}_1}{1-\hat{b}_1+\hat{c}_1})\\ &\quad -(1-\hat{b}_1)\log_2(1+\frac{1-\hat{a}_1-\hat{a}_2}{1-\hat{b}_1})\\ &\quad -\hat{c}_1\log_2(1+\frac{2-\hat{c}_1}{\hat{c}_1})\\ &\quad +(3-\hat{a}_1-\hat{a}_2-\hat{c}_1)\log_2(1+\frac{1-\hat{b}_1+\hat{c}_1}{3-\hat{a}_1-\hat{a}_2-\hat{c}_1})\\ &\quad -(1-\hat{a}_1-\hat{a}_2)\log_2(1+\frac{1-\hat{b}_1}{1-\hat{a}_1-\hat{a}_2})\\ &\quad -(2-\hat{c}_1)\log_2(1+\frac{\hat{c}_1}{2-\hat{c}_1}).\end{aligned}$$

Note that the function $t\log_2(1+1/t)$ is concave. Define $t_1 = \frac{1-\hat{b}_1}{1-\hat{a}_1-\hat{a}_2}$, $t_2 = \frac{\hat{c}_1}{2-\hat{c}_1}$, $r_1 = \frac{1-\hat{a}_1-\hat{a}_2}{3-\hat{a}_1-\hat{a}_2-\hat{c}_1}$, and $r_2 = \frac{2-\hat{c}_1}{3-\hat{a}_1-\hat{a}_2-\hat{c}_1}$. Then we have

$$\begin{aligned}(r_1t_1+r_2t_2)\log_2\big(1+1/(r_1t_1+r_2t_2)\big) &\geq r_1t_1\log_2(1+1/t_1)\\ &\quad +r_2t_2\log_2(1+1/t_2).\end{aligned}$$

That is,

$$\begin{aligned}&(1-\hat{b}_1+\hat{c}_1)\log_2(1+\frac{3-\hat{a}_1-\hat{a}_2-\hat{c}_1}{1-\hat{b}_1+\hat{c}_1})\\ &\geq (1-\hat{b}_1)\log_2(1+\frac{1-\hat{a}_1-\hat{a}_2}{1-\hat{b}_1})+\hat{c}_1\log_2(1+\frac{2-\hat{c}_1}{\hat{c}_1}).\end{aligned}$$

Similarly, we have

$$\begin{aligned}&(3-\hat{a}_1-\hat{a}_2-\hat{c}_1)\log_2(1+\frac{1-\hat{b}_1+\hat{c}_1}{3-\hat{a}_1-\hat{a}_2-\hat{c}_1})\\ &\geq (1-\hat{a}_1-\hat{a}_2)\log_2(1+\frac{1-\hat{b}_1}{1-\hat{a}_1-\hat{a}_2})\\ &\quad +(2-\hat{c}_1)\log_2(1+\frac{\hat{c}_1}{2-\hat{c}_1}).\end{aligned}$$

Therefore, $\tau_3-\tau_1 \geq 0$.

Next, we compute $\tau_4-\tau_2 = \frac{1}{4}\Big(f(\hat{b}_1)+f(\hat{a}_2+\hat{c}_1+1)-f(\hat{a}_2+\hat{b}_1)-f(1+\hat{c}_1)\Big)$. Let $\hat{h}_1(\hat{a}_2,\hat{b}_1,\hat{c}_1) = f(\hat{b}_1)+f(\hat{a}_2+\hat{c}_1+1)-f(\hat{a}_2+\hat{b}_1)-f(1+\hat{c}_1)$. For $0<\hat{a}_2<1$, $0<\hat{b}_1<1$, and $0<\hat{c}_1<1$, we have $\frac{\partial\hat{h}_1(\hat{a}_2,\hat{b}_1,\hat{c}_1)}{\partial\hat{a}_2} = \log_2(\hat{a}_2+\hat{c}_1+1)-\log_2(\hat{a}_2+\hat{b}_1) > 0$. Thus, for $0<\hat{a}_2<1$, $0<\hat{b}_1<1$, and $0<\hat{c}_1<1$, $\hat{h}_1(\hat{a}_2,\hat{b}_1,\hat{c}_1) > \hat{h}_1(\hat{a}_2=0,\hat{b}_1,\hat{c}_1) = 0$, so $\tau_4 > \tau_2$. ■

Example 2: For the late-stage P/E cycling model in Table III, let $\hat{a}_1 = 0.82$, $\hat{a}_2 = 0.1$, $\hat{b}_1 = 0.85$, and $\hat{c}_1 = 0.85$. The uniform rate regions under Gray and NO labelings are plotted in Fig. 2(b). We see that $\mathscr{R}_{NO}^{TIN} \subset \mathscr{R}_{G}^{TIN} \subset \mathscr{R}_{G}^{SC} \subset \mathscr{R}_{NO}^{SC}$. Note that unlike the early-stage P/E cycling model in Example 1, here the region $\mathscr{R}_{G}^{TIN}$ is strictly included in $\mathscr{R}_{G}^{SC}$.

Remark 3: For both P/E cycling models, Theorems 5 and 7 imply that, for TIN decoding, among the 3 labelings, Gray labeling gives the largest sum rate, which is also larger than the sum rate of DS decoding. Moreover, compared to NO labeling, Gray labeling generates a larger uniform rate region for TIN decoding, but a smaller one for SC decoding.

For the early-stage P/E cycling model, we can also draw several conclusions from Lemma 4, Theorem 5, and Table II that provide insight into efficient coding schemes.

First, the sum rate of TIN decoding under Gray labeling is the same as that of SC decoding under any of Gray, NO, and EO labelings. This provides a symmetric capacity-achieving coding solution based on good codes for the point-to-point channel. With Gray labeling, we only need to use two point-to-point symmetric capacity-achieving codes, e.g., polar codes [1], for the lower and upper pages, to achieve the rates $I(X_1;Y)$ and $I(X_2;Y)$, respectively. The two pages can be decoded independently.

Second, for NO labeling or EO labeling, with SC decoding, the rate pair $(R_1 = 1, R_2 = \lambda_5)$ can be achieved, which implies that no coding is required for the lower page X_1. This also suggests us a very simple coding solution. We only need to apply a symmetric capacity-achieving, point-to-point code to the upper page X_2 to achieve rate $I(X_2;Y)$, and no coding is needed for the lower page X_1. To recover the data, we first decode the upper page. Then, we can determine the lower page using the decoded data from the upper page, the binary labeling, the channel transition matrix, and the output Y. For example, referring to the NO labeling in Table I, we see that if the correctly decoded bit of the upper page is $X_2 = 1$ and the output is $Y = s_2$, then the lower page bit must be $X_1 = 0$.

For the late-stage P/E cycling model, from Lemma 6, Theorem 7, and Table IV, we see that the sum rate of TIN decoding under Gray labeling is strictly less than that of SC decoding. The gap Δ between the two sum rates is

$$\begin{aligned}\Delta &= r_{s(G)}^{SC} - r_{s(G)}^{TIN} = \tau_5-\tau_3\\ &= \frac{1}{4}\Big(f(3-\hat{a}_1-\hat{a}_2-\hat{c}_1)-f(2-\hat{a}_1-\hat{a}_2)-f(2-\hat{c}_1)\Big).\end{aligned}$$

To bound the gap Δ, let $\hat{a} = \hat{a}_1+\hat{a}_2$ and $\hat{h}(\hat{a},\hat{c}_1) = f(3-\hat{a}-\hat{c}_1)-f(2-\hat{a})-f(2-\hat{c}_1)$. For $0<\hat{a}<1$ and $0<\hat{c}_1<1$, we have $\frac{\partial\hat{h}(\hat{a},\hat{c}_1)}{\partial\hat{a}} = \log_2(2-\hat{a})-\log_2(3-\hat{a}-\hat{c}_1) < 0$, and $\frac{\partial\hat{h}(\hat{a},\hat{c}_1)}{\partial\hat{c}_1} = \log_2(2-\hat{c}_1)-\log_2(3-\hat{a}-\hat{c}_1) < 0$. Therefore, $\frac{\hat{h}(\hat{a}=1,\hat{c}_1=1)}{4} < \Delta < \frac{\hat{h}(\hat{a}=0,\hat{c}_1=0)}{4}$; that is, $0<\Delta<\frac{3\log_2 3-4}{4} = 0.1887$.

If we impose constraints $\eta_{\hat{a}} \leq \hat{a}_1+\hat{a}_2 < 1$ and $\eta_{\hat{c}_1} \leq \hat{c}_1 < 1$, we have $0<\Delta \leq \frac{1}{4}\Big(f(3-\eta_{\hat{a}}-\eta_{\hat{c}_1})-f(2-\eta_{\hat{a}})-f(2-\eta_{\hat{c}_1})\Big)$. For example, for $\eta_{\hat{a}} = 0.95$ and $\eta_{\hat{c}_1} = 0.85$, we get $0<\Delta\leq 0.00246$. In general, the gap Δ is very small. □

Data Retention Model: For the data retention model, the structure of the channel transition matrix $p_{MLC}^{DR}(y|\upsilon)$, for output $y \in \mathcal{Y}_{MLC}$ and voltage level $\upsilon \in \mathcal{V}_{MLC}$, is shown in Table V, where $\tilde{a}_1$, $1-\tilde{a}_1$, $\tilde{b}_1$, $1-\tilde{b}_1$, $\tilde{c}_1$, and $1-\tilde{c}_1$

TABLE V
CHANNEL TRANSITION MATRIX $p_{MLC}^{DR}(y|v)$ OF DATA RETENTION MODEL FOR MLC FLASH MEMORIES

V	Inputs: (X_1, X_2)			Output: Y			
Levels	Gray	NO	EO	s_0	s_1	s_2	s_3
A_0	(11)	(11)	(11)	1	0	0	0
A_1	(10)	(10)	(00)	$1-\tilde{a}_1$	$\tilde{a}_1$	0	0
A_2	(00)	(01)	(01)	0	$1-\tilde{b}_1$	$\tilde{b}_1$	0
A_3	(01)	(00)	(10)	0	0	$1-\tilde{c}_1$	$\tilde{c}_1$

represent non-zero probabilities. In contrast to the early-stage P/E cycling model, where errors are caused by upward drift of cell voltages, in the data retention model, errors arise from downward drift of the cell voltages.

Analysis and results for the data retention model are very similar to those of the early-stage P/E cycling model. We state only one representative result here, without a detailed proof.

Lemma 8: For channel transition matrix $p_{MLC}^{DR}(y|v)$, using Gray labeling, we have $r_s^{TIN} = r_s^{SC}$ and $\mathscr{R}^{TIN} = \mathscr{R}^{SC}$. Using either NO labeling or EO labeling, we have $r_s^{TIN} < r_s^{SC}$.

A.2 Extension to All Labelings

In this subsection, we extend the analysis to the entire set of labelings. There exist a total of 4! = 24 labelings. In order to categorize and analyze these 24 labelings, we take advantage of the algebraic structure of permutation groups, and consider a labeling σ as a permutation π in the symmetric group $\mathcal{S}_4$. This is the group whose elements are all the permutation operations that can be performed on the 4 distinct elements in $\mathcal{T}_{MLC}$, and whose group operation, denoted as $*$, is the composition of such permutation operations. A labeling $\sigma = (w_0, w_1, w_2, w_3)$ corresponds to the permutation $\pi = (w_0, w_1, w_2, w_3)$ in $\mathcal{S}_4$, where the permutation vector $\pi = (w_0, w_1, w_2, w_3)$ is defined to represent $\pi(11) = w_0$, $\pi(10) = w_1$, $\pi(01) = w_2$, and $\pi(00) = w_3$, e.g., $\pi = (11, 10, 01, 00)$ is the identity permutation in $\mathcal{S}_4$. The group operation $*$ of two permutations π_1 and π_2 is defined as their composition and results in another permutation $\pi_3 = \pi_1 * \pi_2$. In other words, $\pi_1 * \pi_2$ is the function that maps any element $w \in \mathcal{T}_{MLC}$ to $\pi_1\big(\pi_2(w)\big)$. Note that the rightmost permutation is applied first. For example, $(10, 11, 00, 01) * (11, 10, 00, 01) = (10, 11, 01, 00)$.

Lemma 9: In the symmetric group $\mathcal{S}_4$, G_0=$\{(11, 10, 01, 00)$, $(10, 11, 00, 01)$, $(01, 00, 11, 10)$, $(00, 01, 10, 11)\}$ forms a normal subgroup (the Klein four-group).

Proof: The element $(11, 10, 01, 00)$ in G_0 is the identity element in $\mathcal{S}_4$. We can verify that the 4 elements in G_0 have the following properties: 1) the composition of the identity element and any element is that element itself; 2) the composition of any non-identity element with itself is the identity element; 3) the composition of two distinct non-identity elements is the third non-identity element. Thus, G_0 is the Klein four-group. ■

With the subgroup G_0 in Lemma 9, we partition $\mathcal{S}_4$ into G_0 and its 5 cosets, each of size 4: $\mathcal{S}_4 = G_0 \cup G_1 \cup G_2 \cup \bar{G}_0 \cup \bar{G}_1 \cup \bar{G}_2$, where G_1=G_0*(11, 10, 00, 01); G_2=G_0*(11, 00, 01, 10); $\bar{G}_0 = G_0 * (11, 01, 10, 00)$; $\bar{G}_1 = G_0 * (11, 01, 00, 10)$; $\bar{G}_2 = G_0 * (11, 00, 10, 01)$.

In the following, we will treat each vector in every coset as a labeling. For example, G_0 includes $\sigma_{NO} = (11, 10, 01, 00)$, G_1 includes $\sigma_G = (11, 10, 00, 01)$, and G_2 includes $\sigma_{EO} = (11, 00, 01, 10)$. The following two lemmas give properties of the uniform rate regions for different labelings. We assume an *arbitrary* channel transition matrix $p_{MLC}(y|v)$, for output $y \in \mathcal{Y}_{MLC}$ and voltage level $v \in \mathcal{V}_{MLC}$, is given. The first lemma leverages the symmetries within the Klein four-group and its cosets to deduce the relationship of the rate regions of different labelings.

Lemma 10: With an arbitrary channel transition matrix $p_{MLC}(y|v)$, for TIN decoding, the 4 labelings in each of G_0, G_1, G_2, $\bar{G}_0$, $\bar{G}_1$, and $\bar{G}_2$ give the same uniform rate region $\mathscr{R}^{TIN}$ and sum rate r_s^{TIN}. For SC decoding, the 4 labelings in each of G_0, G_1, G_2, $\bar{G}_0$, $\bar{G}_1$, and $\bar{G}_2$ give the same uniform rate region $\mathscr{R}^{SC}$, and all 24 labelings in $\mathcal{S}_4$ give the same sum rate r_s^{SC}.

Proof: This is based on the fact that the 4 labelings in each of G_0, G_1, G_2, $\bar{G}_0$, $\bar{G}_1$, and $\bar{G}_2$ are interchangeable by one of the following three operations: 1) in position X_1, change 0 to 1 and 1 to 0; 2) in position X_2, change 0 to 1 and 1 to 0; 3) in both positions X_1 and X_2, change 0 to 1 and 1 to 0. For example, in G_0, $(11, 10, 01, 00)$ is transformed to $(01, 00, 11, 10)$ by changing 0 to 1 and 1 to 0 in position X_1, is transformed to $(10, 11, 00, 01)$ by changing 0 to 1 and 1 to 0 in position X_2, and is transformed to $(00, 01, 10, 11)$ by changing 0 to 1 and 1 to 0 in both positions X_1 and X_2. Since the distributions for X_1 and X_2 are uniform, the values of $I(X_1; Y)$, $I(X_2; Y)$, $I(X_1; Y|X_2)$, and $I(X_2; Y|X_1)$ under a labeling σ_1 are the same as those under a labeling σ_2 which is obtained by one of the above three operations on the labeling σ_1. Thus, for a fixed decoding scheme (TIN or SC), the uniform rate region and sum rate under the labeling σ_1 are the same as those under the labeling σ_2. Therefore, for a fixed decoding scheme (TIN or SC), the 4 labelings in each coset give the same uniform rate region and sum rate. For the sum rate of SC decoding, for all 24 labelings, $I(X_1, X_2; Y)$ is the same due to the uniform distributions for X_1 and X_2. ■

Lemma 11: With an arbitrary channel transition matrix $p_{MLC}(y|v)$, for TIN decoding, if the labelings in G_i, $i = 0, 1, 2$, give a uniform rate region: $R_1 \le \varphi_1$ and $R_2 \le \varphi_2$, then the labelings in $\bar{G}_i$ give a uniform rate region: $R_1 \le \varphi_2$ and $R_2 \le \varphi_1$. For SC decoding, if the labelings in G_i give a uniform rate region: $R_1 \le \psi_1$, $R_2 \le \psi_2$, and $R_1 + R_2 \le \psi_3$, then the labelings in $\bar{G}_i$ give a uniform rate region: $R_1 \le \psi_2$, $R_2 \le \psi_1$, and $R_1 + R_2 \le \psi_3$.

Proof: This is based on the fact that the 4 labelings in G_i, i=0, 1, 2, are transformed (one-to-one) to the 4 labelings in $\bar{G}_i$ by swapping the values in positions X_1 and X_2. For example, $(11, 10, 01, 00)$ in G_0 is transformed to $(11, 01, 10, 00)$ in $\bar{G}_0$. With the uniform distributions for X_1 and X_2, X_1 (or X_2) with labeling $(11, 10, 01, 00)$ is equivalent to X_2 (or X_1) with labeling $(11, 01, 10, 00)$. Thus, for a fixed decoding scheme (TIN or SC), the uniform rate region under labeling $(11, 10, 01, 00)$ will become the one under labeling $(11, 01, 10, 00)$ by swapping the constraints on R_1 and R_2. From Lemma 10, it follows that the uniform rate region under labelings in G_0 will become the one under labelings in $\bar{G}_0$ by

swapping the constraints on R_1 and R_2. The same conclusion holds for the labelings in G_i and $\bar{G}_i$, $i = 1, 2$. ■

Remark 4: Theorems 5 and 7, and Lemmas 10 and 11, imply that for both P/E cycling models, with TIN decoding, the 8 labelings in G_1 (including Gray labeling) and $\bar{G}_1$ produce the largest sum rate among all 24 labelings. With SC decoding, all of the 24 labelings give the same sum rate. □

Finally, we examine the uniform rate region that can be achieved by using multiple labelings in $\mathcal{S}_4$ within a codeword in a time-sharing fashion. Define $\mathscr{R}_{\mathcal{S}_4}^{TIN} = Conv\Big(\bigcup_{\sigma\in\mathcal{S}_4}\mathscr{R}_{\sigma}^{TIN}\Big)$, the convex hull of uniform rate regions of all 24 labelings for TIN decoding. Define $\mathscr{R}_{\mathcal{S}_4}^{SC} = Conv\Big(\bigcup_{\sigma\in\mathcal{S}_4}\mathscr{R}_{\sigma}^{SC}\Big)$, the convex hull of uniform rate regions of all 24 labelings for SC decoding. Through time-sharing of different labelings, we obtain the following lemma.

Lemma 12: For TIN decoding, any point $(R_1, R_2) \in \mathscr{R}_{\mathcal{S}_4}^{TIN}$ can be achieved. For SC decoding, any point $(R_1, R_2) \in \mathscr{R}_{\mathcal{S}_4}^{SC}$ can be achieved.

Proof: We first show that any point $(R_1, R_2) \in \mathscr{R}_{\mathcal{S}_4}^{TIN}$ can be achieved. From Carathéodory's Theorem [10], any point (R_1, R_2) in $\mathscr{R}_{\mathcal{S}_4}^{TIN}$ can be represented as a convex combination of 3 points in $\bigcup_{\sigma\in\mathcal{S}_4}\mathscr{R}_{\sigma}^{TIN}$. Without loss of generality, we assume $(R_1, R_2) = \alpha_1(R_1^1, R_2^1) + \alpha_2(R_1^2, R_2^2) + \alpha_3(R_1^3, R_2^3)$, where $\alpha_1, \alpha_2, \alpha_3 \geq 0$ and $\sum_{i=1}^{3}\alpha_i = 1$. Points (R_1^1, R_2^1), (R_1^2, R_2^2), and (R_1^3, R_2^3) in $\bigcup_{\sigma\in\mathcal{S}_4}\mathscr{R}_{\sigma}^{TIN}$ are achievable under some labelings. Consider three sequences of codes, achieving (R_1^1, R_2^1), (R_1^2, R_2^2), and (R_1^3, R_2^3), respectively. For each block length n, consider the $(2^{\alpha_1 nR_1^1}, 2^{\alpha_1 nR_2^1}, \alpha_1 n)$, $(2^{\alpha_2 nR_1^2}, 2^{\alpha_2 nR_2^2}, \alpha_2 n)$, and $(2^{\alpha_3 nR_1^3}, 2^{\alpha_3 nR_2^3}, \alpha_3 n)$ codes from the given three sequences of codes, respectively. By a standard time-sharing argument [10], a fourth $(2^{nR_1}, 2^{nR_2}, n)$ code can be constructed from the above three codes (details are omitted here due to space limitation). Thus, any point $(R_1, R_2) \in \mathscr{R}_{\mathcal{S}_4}^{TIN}$ can be achieved. In a similar manner, one can show that any point $(R_1, R_2) \in \mathscr{R}_{\mathcal{S}_4}^{SC}$ is achievable. ■

Moreover, for the early-stage P/E cycling model in Table I, the rate region $\mathscr{R}_{\mathcal{S}_4}^{SC}$ can be determined explicitly.

Theorem 13: For the early-stage P/E cycling model, $\mathscr{R}_{\mathcal{S}_4}^{SC}$ is the set of all pairs (R_1, R_2) such that $R_1 \leq 1$, $R_2 \leq 1$, and $R_1 + R_2 \leq I(X_1, X_2; Y) = 1 + \lambda_5$.

Proof: Using Table II, Lemma 10, and Lemma 11, we can calculate the convex hull of the uniform rate regions of all 24 labelings. We can also see that the convex hull of the uniform rate regions of two labelings $(11, 10, 01, 00)$ and $(11, 01, 10, 00)$ are enough to achieve $\mathscr{R}_{\mathcal{S}_4}^{SC}$. ■

Example 3: For the early-stage of P/E cycling, let $a_1 = 0.98$, $b_1 = 0.97$, and $c_1 = 0.99$. The uniform rate regions $\mathscr{R}_{\mathcal{S}_4}^{TIN}$, $\mathscr{R}_{\mathcal{S}_4}^{SC}$, and $\mathscr{R}_{G}^{DS}$ are plotted in Fig. 3. It can be seen that $\mathscr{R}_{G}^{DS} \subset \mathscr{R}_{\mathcal{S}_4}^{TIN} \subset \mathscr{R}_{\mathcal{S}_4}^{SC}$, and the line connecting the two corner points in $\mathscr{R}_{\mathcal{S}_4}^{TIN}$ is on the line connecting the two corner points in $\mathscr{R}_{\mathcal{S}_4}^{SC}$. Moreover, either R_1 or R_2 can achieve rate 1 with SC decoding. For the late-stage P/E cycling model, let $\hat{a}_1 = 0.82$, $\hat{a}_2 = 0.1$, $\hat{b}_1 = 0.85$, and $\hat{c}_1 = 0.85$. The uniform rate regions $\mathscr{R}_{\mathcal{S}_4}^{TIN}$, $\mathscr{R}_{\mathcal{S}_4}^{SC}$, and $\mathscr{R}_{G}^{DS}$ are plotted in Fig. 4. For this case, there is a gap between the line connecting the two corner points in $\mathscr{R}_{\mathcal{S}_4}^{TIN}$ and the one connecting the two corner points in $\mathscr{R}_{\mathcal{S}_4}^{SC}$. This property implies that SC decoding will give a larger sum rate than TIN decoding.

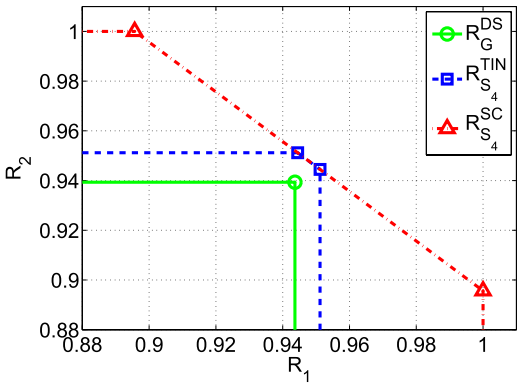


Fig. 3. Uniform rate regions $\mathscr{R}_{\mathcal{S}_4}^{TIN}$, $\mathscr{R}_{\mathcal{S}_4}^{SC}$, and $\mathscr{R}_{G}^{DS}$ with $a_1 = 0.98$, $b_1 = 0.97$, and $c_1 = 0.99$ for the early-stage P/E cycling model.

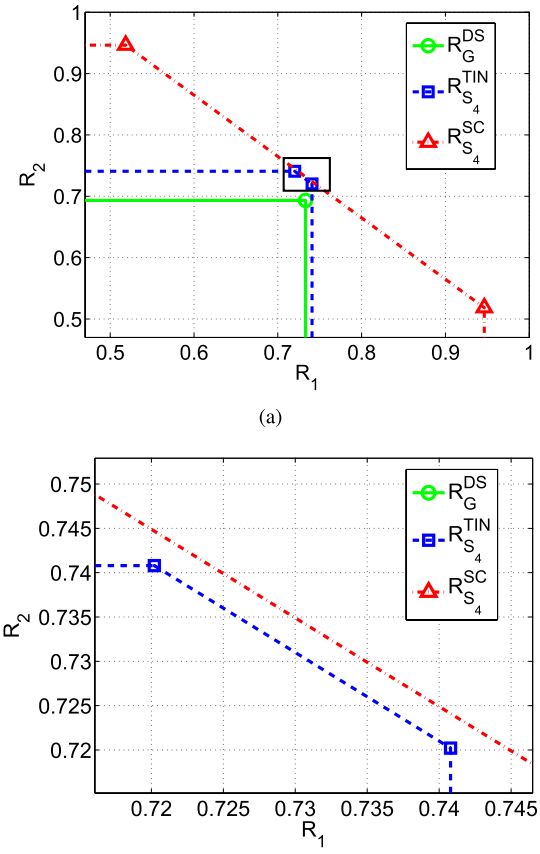


Fig. 4. (a) Uniform rate regions $\mathscr{R}_{\mathcal{S}_4}^{TIN}$, $\mathscr{R}_{\mathcal{S}_4}^{SC}$, and $\mathscr{R}_{G}^{DS}$ with $\hat{a}_1 = 0.82$, $\hat{a}_2 = 0.1$, $\hat{b}_1 = 0.85$, and $\hat{c}_1 = 0.85$ for the late-stage P/E cycling model, where the two curves (blue and red) in the black rectangle are enlarged and shown in (b).

Remark 5: Although we focus on the P/E cycling and data retention models, our analysis can be extended to other channel models. As a simple example, we consider a channel model whose channel transition matrix $p_{MLC}^{C}(y|v)$ reflects both upward and downward drift of voltage levels due to the combined effects of P/E cycling and data retention.

TABLE VI

CHANNEL TRANSITION MATRIX $p^{C}_{MLC}(y|v)$ FOR COMBINED EFFECTS OF P/E CYCLING AND DATA RETENTION FOR MLC FLASH MEMORIES

V	Inputs: (X_1, X_2)			Output: Y			
Levels	Gray	NO	EO	s_0	s_1	s_2	s_3
A_0	(11)	(11)	(11)	$\bar{a}_1$	$1-\bar{a}_1$	0	0
A_1	(10)	(10)	(00)	$\bar{b}_2$	$\bar{b}_1$	$1-\bar{b}_1-\bar{b}_2$	0
A_2	(00)	(01)	(01)	0	$\bar{c}_2$	$\bar{c}_1$	$1-\bar{c}_1-\bar{c}_2$
A_3	(01)	(00)	(10)	0	0	$1-\bar{d}_1$	$\bar{d}_1$

The structure of the transition matrix is shown in Table VI, where $\bar{a}_1$, $1-\bar{a}_1$, $\bar{b}_1$, $\bar{b}_2$, $1-\bar{b}_1-\bar{b}_2$, $\bar{c}_1$, $\bar{c}_2$, $1-\bar{c}_1-\bar{c}_2$, $\bar{d}_1$, and $1-\bar{d}_1$ represent non-zero probabilities. From Theorem 1, we readily conclude that $r_s^{TIN} < r_s^{SC}$ with any of the 24 possible labelings since, for any labeling, three of the probabilities p_{0,s_1}, p_{1,s_1}, p_{2,s_1}, and p_{3,s_1} are positive. □

B. Quantization With Increased Number of Reads

In the previous analysis, we used a quantizer with three read thresholds. We now investigate the improvement in performance that can be obtained by applying additional reads to obtain more refined soft information.

For a channel $\mathcal{W}_{MLC}$, assume there is an output set $\mathcal{Y}^{q}_{MLC} = \{s_0, s_1, \ldots, s_{q-1}\}$ obtained by a set of $q-1$ reads. Denote the corresponding uniform rate regions for TIN and SC decodings by $\mathscr{R}^{TIN}$ and $\mathscr{R}^{SC}$, and the sum rates for TIN and SC decodings by r_s^{TIN} and r_s^{SC}.

Now, introduce one more read threshold to split one of the outputs $s_0, s_1, \ldots, s_{q-1}$. Without loss of generality, we split s_0 into s_0^1 and s_0^2 to obtain a new output set $\hat{\mathcal{Y}}^{q+1}_{MLC} = \{s_0^1, s_0^2, s_1, s_2, \ldots, s_{q-1}\}$. The resulting uniform rate regions for TIN and SC decodings are denoted by $\hat{\mathscr{R}}^{TIN}$ and $\hat{\mathscr{R}}^{SC}$, respectively, and the corresponding sum rates by $\hat{r}_s^{TIN}$ and $\hat{r}_s^{SC}$, respectively. The following lemma shows that for both TIN and SC decoding, one-step progressive quantization produces rate regions that contain the original rate regions. Thus the sum rates are not decreased, and in fact they become strictly larger except when the transition probabilities satisfy very specific conditions after quantization.

Lemma 14: For uniform rate regions, under TIN and SC decodings, $\mathscr{R}^{TIN} \subseteq \hat{\mathscr{R}}^{TIN}$ and $\mathscr{R}^{SC} \subseteq \hat{\mathscr{R}}^{SC}$. For sum rates, under TIN and SC decodings, $r_s^{TIN} \le \hat{r}_s^{TIN}$ and $r_s^{SC} \le \hat{r}_s^{SC}$.

Proof: See Appendix C. ■

In the following example, we show by means of computer simulation that the performance of MLC flash can be improved through the use of additional reads.

Example 4: Following [23], we assume that the readback cell voltage has the normal-Laplace distribution $\mathcal{NL}(\mu, \nu, \alpha, \beta)$. The corresponding cumulative distribution function (cdf) for all real y is

$$F(y) = \Phi\left(\frac{y-\mu}{\nu}\right) - \left(\phi\left(\frac{y-\mu}{\nu}\right) \cdot \frac{\beta\Re(\alpha\nu - (y-\mu)/\nu) - \alpha\Re(\beta\nu + (y-\mu)/\nu)}{\alpha+\beta}\right),$$

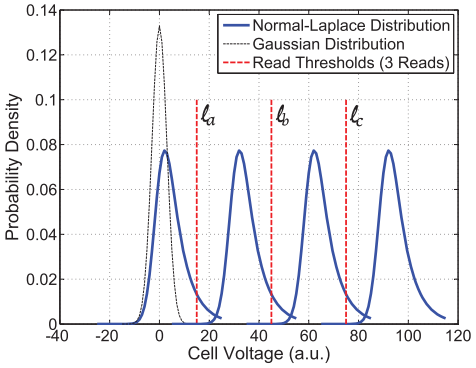


Fig. 5. Channel model for MLC flash memories with cell voltage modeled as the normal-Laplace distribution.

where Φ and ϕ are the cdf and probability density function (pdf) of a standard normal random variable and $\Re$ is Mills' ratio $\Re(z) = \frac{1-\Phi(z)}{\phi(z)}$ [26]. The readback cell voltage distributions for inputs A_0, A_1, A_2, and A_3 are defined to be $\mathcal{NL}(0, 3, 1/6, 1)$, $\mathcal{NL}(30, 3, 1/6, 1)$, $\mathcal{NL}(60, 3, 1/6, 1)$, and $\mathcal{NL}(90, 3, 1/6, 1)$, respectively. Here, the parameters of the normal-Laplace distributions are chosen to qualitatively reflect the distributions reported in the literature and to illustrate the effect of output quantization using multiple reads. The cell voltage is in arbitrary units (a.u.), chosen merely for convenience.

In the standard 3-read setting, the read thresholds are placed at positions $\ell_a = 15$, $\ell_b = 45$, and $\ell_c = 75$. We use a vector to represent these positions, $\vec{L}_3 = (15, 45, 75)$. On either side of each of the positions ℓ_a, ℓ_b, and ℓ_c, we define an additional t read positions, regularly spaced at intervals of d units. For example, setting $t = 1$ and $d = 3$, we specify a total of 9 reads centered at ℓ_a, ℓ_b, and ℓ_c, with resulting read position vector $\vec{L}_9$=(12, 15, 18, 42, 45, 48, 72, 75, 78). Similarly, for a total of 15 reads, we set $t = 2$ and $d = 3$, and for a total of 21 reads, we set $t = 3$ and $d = 3$.

For the Gray labeling, the uniform rate regions under TIN and SC decodings are plotted in Fig. 6. As expected, the rate regions of TIN decoding and SC decoding are similar. Additional reads can significantly improve the rates of both lower and upper pages. For the NO labeling, the uniform rate regions under TIN and SC decodings are plotted in Fig. 7. With either TIN decoding or SC decoding, additional reads effectively enhance the rate of the upper page. However, with SC decoding, the rate improvement of the lower page is very limited.

IV. EXTENSION TO TLC FLASH MEMORY

In this section, we extend our analysis from MLC to TLC flash memory. Although similar tools can be applied, we will see that the characterization of the rate regions becomes more intricate as the number of bits per cell increases.

Similar to the MLC case, given a labeling σ and a quantizer Q, the TLC flash memory channel can be modeled as a 3-user discrete memoryless multiple-access channel $\mathcal{W}_{TLC}$: $(\mathcal{X} \times \mathcal{X} \times \mathcal{X}, p(y|x_1, x_2, x_3), \mathcal{Y})$, where $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{s_0, s_1, \ldots, s_{q-1}\}$, and $p(y|x_1, x_2, x_3)$ is the

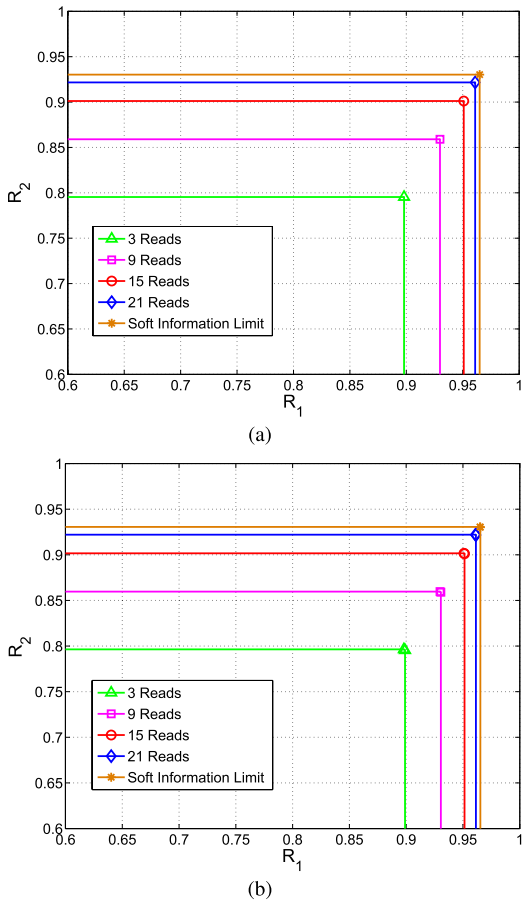


Fig. 6. Uniform rate regions under Gray labeling with different number of reads: (a) using TIN decoding, and (b) using SC decoding.

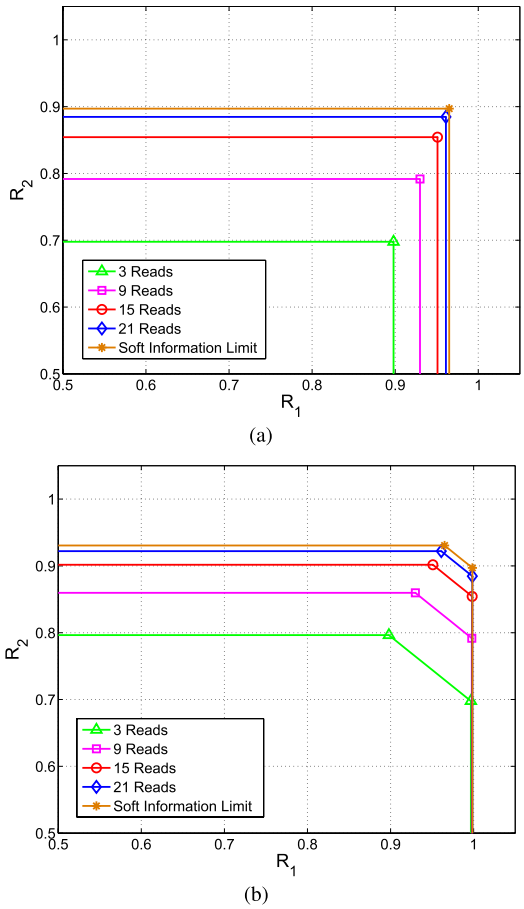


Fig. 7. Uniform rate regions under NO labeling with different number of reads: (a) using TIN decoding, and (b) using SC decoding.

transition probability for any $x_1, x_2, x_3 \in \mathcal{X}$ and $y \in \mathcal{Y}$. We again use a simplified notation for the transition probabilities, $p_{BD(x_1,x_2,x_3),y} \stackrel{\text{def}}{=} P(Y = y|X_1 = x_1, X_2 = x_2, X_3 = x_3)$; e.g., $p_{5,s_0} = P(Y = s_0|X_1 = 1, X_2 = 0, X_3 = 1)$.

A. Rate Regions for TIN and SC Decoding Schemes

For a channel $\mathcal{W}_{TLC}$, with TIN decoding, the uniform rate region $\mathscr{R}^{TIN}$ is the set of all rate tuples (R_1, R_2, R_3) such that $R_i \leq I(X_i; Y)$ for all $i = 1, 2, 3$. The sum rate is $r_s^{TIN} = \max\{\sum_{i=1}^3 R_i : (R_1, R_2, R_3) \in \mathscr{R}^{TIN}\} = \sum_{i=1}^3 I(X_i, Y)$. With SC decoding, the uniform rate region $\mathscr{R}^{SC}$ is the set of all rate tuples (R_1, R_2, R_3) such that

1) $R_1 \leq I(X_1; Y|X_2, X_3)$, $R_2 \leq I(X_2; Y|X_1, X_3)$, $R_3 \leq I(X_3; Y|X_1, X_2)$;
2) $R_1 + R_2 \leq I(X_1, X_2; Y|X_3)$, $R_1 + R_3 \leq I(X_1, X_3; Y|X_2)$, $R_2 + R_3 \leq I(X_2, X_3; Y|X_1)$;
3) $R_1 + R_2 + R_3 \leq I(X_1, X_2, X_3; Y)$.

The sum rate is $r_s^{SC} = \max\{\sum_{i=1}^3 R_i : (R_1, R_2, R_3) \in \mathscr{R}^{SC}\} = I(X_1, X_2, X_3; Y)$.

The following theorem, analogous to Theorem 1, characterizes the channels for which the sum rates of TIN decoding and SC decoding are the same.

Theorem 15: For a channel $\mathcal{W}_{TLC}$, the sum rates satisfy $r_s^{TIN} \leq r_s^{SC}$, with equality if and only if

$$(p_{7,s_j} + p_{6,s_j})(p_{1,s_j} + p_{0,s_j}) = (p_{5,s_j} + p_{4,s_j})(p_{3,s_j} + p_{2,s_j}), \tag{2a}$$

$$p_{7,s_j}(p_{4,s_j} + p_{2,s_j} + p_{0,s_j}) = p_{6,s_j}(p_{5,s_j} + p_{3,s_j} + p_{1,s_j}), \tag{2b}$$

$$p_{5,s_j}(p_{6,s_j} + p_{2,s_j} + p_{0,s_j}) = p_{4,s_j}(p_{7,s_j} + p_{3,s_j} + p_{1,s_j}), \tag{2c}$$

$$p_{3,s_j}(p_{6,s_j} + p_{4,s_j} + p_{0,s_j}) = p_{2,s_j}(p_{7,s_j} + p_{5,s_j} + p_{1,s_j}), \tag{2d}$$

$$p_{1,s_j}(p_{6,s_j} + p_{4,s_j} + p_{2,s_j}) = p_{0,s_j}(p_{7,s_j} + p_{5,s_j} + p_{3,s_j}), \tag{2e}$$

for all $j = 0, 1, \ldots, q-1$. If $r_s^{TIN} = r_s^{SC}$, then $\mathscr{R}^{TIN} = \mathscr{R}^{SC}$ and the rate region is a cube.

Proof: See Appendix D. ■

Along the lines of Theorem 2, the following theorem provides an upper bound on the difference between the sum rates r_s^{SC} and r_s^{TIN}.

Theorem 16: For a channel $\mathcal{W}_{TLC}$, the rate difference $r_s^{SC} - r_s^{TIN} \leq 2$ with equality if and only if

$$p_{7,s_j} + p_{6,s_j} + p_{5,s_j} + p_{4,s_j} = p_{3,s_j} + p_{2,s_j} + p_{1,s_j} + p_{0,s_j}, \tag{3a}$$

$$p_{7,s_j} + p_{5,s_j} + p_{3,s_j} + p_{1,s_j} = p_{6,s_j} + p_{4,s_j} + p_{2,s_j} + p_{0,s_j}, \tag{3b}$$

$$(p_{7,s_j} + p_{6,s_j})(p_{3,s_j} + p_{2,s_j}) = (p_{5,s_j} + p_{4,s_j})(p_{1,s_j} + p_{0,s_j}) = 0, \tag{3c}$$

$$p_{7,s_j} p_{6,s_j} = p_{5,s_j} p_{4,s_j} = p_{3,s_j} p_{2,s_j} = p_{1,s_j} p_{0,s_j} = 0, \tag{3d}$$

for all $j = 0, 1, \ldots, q-1$.

Proof: We bound the difference $r_s^{SC} - r_s^{TIN}$ as

$$\begin{aligned} r_s^{SC} - r_s^{TIN} &= I(X_1; X_2|Y) + I(X_1, X_2; X_3|Y) \\ &= H(X_1|Y) - H(X_1|X_2, Y) + H(X_3|Y) \\ &\quad - H(X_3|X_1, X_2, Y) \\ &\overset{(a)}{\leq} H(X_1) + H(X_3) - H(X_1|X_2, Y) \\ &\quad - H(X_3|X_1, X_2, Y) \\ &\overset{(b)}{\leq} H(X_1) + H(X_3) = 2, \end{aligned}$$

where step (a) follows from $H(X_1|Y) \leq H(X_1)$ and $H(X_3|Y) \leq H(X_3)$, and step (b) follows from $H(X_1|X_2, Y) \geq 0$ and $H(X_3|X_1, X_2, Y) \geq 0$. Thus, $r_s^{SC} - r_s^{TIN} = 2$ if and only if 1) $H(X_1|Y) = H(X_1) = H(X_3|Y) = H(X_3) = 1$, and 2) $H(X_1|X_2, Y) = H(X_3|X_1, X_2, Y) = 0$.

The condition $H(X_1|Y) = \sum_{j=0}^{q-1}(\sum_{i=0}^{7} \frac{p_{i,s_j}}{8})$ $H(X_1|Y = s_j) = 1$ holds if and only if condition (3a) in the statement of the theorem holds. Similarly, $H(X_3|Y) = 1$ corresponds to condition (3b), $H(X_1|X_2, Y) = 0$ corresponds to condition (3c), and $H(X_3|X_1, X_2, Y) = 0$ corresponds to condition (3d). ■

B. Performance of TLC Flash Memory With Different Decoding Schemes and Labelings

We now investigate the uniform rate region and sum rate of two TLC flash memory channel models, using TIN and SC decoding and different labelings of cell voltage levels. We first formally define the generalizations of the Gray labeling and NO labeling.

Definition 17: For TLC flash memory, labeling σ_G=(111, 110, 100, 101, 001, 000, 010, 011) is called Gray labeling, and σ_{NO}=(111, 110, 101, 100, 011, 010, 001, 000) is called Natural Order (NO) labeling.

For each labeling, the mapping between inputs $(X_1, X_2, X_3) \in \mathcal{T}_{TLC}$ and voltage levels $V \in \mathcal{V}_{TLC}$ is shown in Table VII.

B.1 Performance of Gray and NO Labelings

We fix our quantizer Q with seven reads, which are placed between every pair of neighboring voltage levels, as shown in Fig. 1(b). Hence, the output alphabet is $\mathcal{Y}_{TLC} = \{s_0, s_1, \ldots, s_7\}$.

We consider a simple P/E cycling model, called the *early-stage P/E cycling* model, characterized by the channel transition matrix $p_{TLC}^{E}(y|v)$, for $y \in \mathcal{Y}_{TLC}$ and $v \in \mathcal{V}_{TLC}$, shown in Table VII, where ε_i and $1 - \varepsilon_i$, $i = 0, 1, \ldots, 6$, represent non-zero probabilities. This model reflects the phenomenon of upward shift of the cell voltage. A related model is the *data retention* model, where the channel transition matrix $p_{TLC}^{DR}(y|v)$ reflects the downward shift of cell voltage. The structure of the transition matrix is shown in Table VIII, where ϵ_i and $1 - \epsilon_i$, $i = 1, 2, \ldots, 7$, represent non-zero probabilities. We give an analysis of the early-stage P/E cycling model below. The same techniques can be applied to the data retention model, yielding analogous conclusions. More sophisticated TLC channel models can be developed by means of further experimental studies; once they are obtained, a similar analysis can be carried out.

TABLE VII

CHANNEL TRANSITION MATRIX $p_{TLC}^{E}(y|v)$ AT EARLY-STAGE OF P/E CYCLING FOR TLC FLASH MEMORIES

V	Inputs: (X_1, X_2, X_3)		Output: Y							
Levels	Gray	NO	s_0	s_1	s_2	s_3	s_4	s_5	s_6	s_7
B_0	(111)	(111)	ε_0	1- ε_0	0	0	0	0	0	0
B_1	(110)	(110)	0	ε_1	$1-\varepsilon_1$	0	0	0	0	0
B_2	(100)	(101)	0	0	ε_2	$1-\varepsilon_2$	0	0	0	0
B_3	(101)	(100)	0	0	0	ε_3	$1-\varepsilon_3$	0	0	0
B_4	(001)	(011)	0	0	0	0	ε_4	1-ε_4	0	0
B_5	(000)	(010)	0	0	0	0	0	ε_5	1-ε_5	0
B_6	(010)	(001)	0	0	0	0	0	0	ε_6	1-ε_6
B_7	(011)	(000)	0	0	0	0	0	0	0	1

TABLE VIII

CHANNEL TRANSITION MATRIX $p_{TLC}^{DR}(y|v)$ OF DATA RETENTION MODEL FOR TLC FLASH MEMORIES

V	Inputs: (X_1, X_2, X_3)		Output: Y							
Levels	Gray	NO	s_0	s_1	s_2	s_3	s_4	s_5	s_6	s_7
B_0	(111)	(111)	1	0	0	0	0	0	0	0
B_1	(110)	(110)	$1-\epsilon_1$	ϵ_1	0	0	0	0	0	0
B_2	(100)	(101)	0	$1-\epsilon_2$	ϵ_2	0	0	0	0	0
B_3	(101)	(100)	0	0	$1-\epsilon_3$	ϵ_3	0	0	0	0
B_4	(001)	(011)	0	0	0	$1-\epsilon_4$	ϵ_4	0	0	0
B_5	(000)	(010)	0	0	0	0	$1-\epsilon_5$	ϵ_5	0	0
B_6	(010)	(001)	0	0	0	0	0	$1-\epsilon_6$	ϵ_6	0
B_7	(011)	(000)	0	0	0	0	0	0	$1-\epsilon_7$	ϵ_7

For the early-stage P/E cycling model, we have the following lemma.

Lemma 18: For channel transition matrix $p_{TLC}^{E}(y|v)$, using Gray labeling, we have $r_s^{TIN} = r_s^{SC}$ and $\mathscr{R}^{TIN} = \mathscr{R}^{SC}$. Using NO labeling, we have $r_s^{TIN} < r_s^{SC}$.

Proof: With Gray labeling, we can verify that the conditions, i.e., the set of equations (2a)-(2e), in Theorem 15, are satisfied for $j = 0, 1, \ldots, 7$. Thus, from Theorem 15, we conclude that $r_s^{TIN} = r_s^{SC}$ and $\mathscr{R}^{TIN}$=$\mathscr{R}^{SC}$. While under NO labeling $p_{7,s_2}(p_{4,s_2} + p_{2,s_2} + p_{0,s_2}) \neq p_{6,s_2}(p_{5,s_2} + p_{3,s_2} + p_{1,s_2})$. Thus, in this case, $r_s^{TIN} < r_s^{SC}$. ■

Next, for the early-stage P/E cycling model, we make the following observations about the corner points of uniform rate regions for TIN and SC decodings under Gray and NO labelings. First, with Gray labeling, according to Theorem 15 and Lemma 18, the rate regions for TIN and SC decoding have a common unique corner point: $\mathbb{P}_G^0 = \big(I(X_1; Y), I(X_2; Y), I(X_3; Y)\big)$. Second, with NO labeling, for TIN decoding, the corner point is: $\mathbb{P}_{NO}^0 = \big(I(X_1; Y), I(X_2; Y), I(X_3; Y)\big)$. [4] For SC decoding, we have

[4] Although there is a term $I(X_1; Y)$ in both $\mathbb{P}_G^0$ and $\mathbb{P}_{NO}^0$, they represent different values due to the use of different labelings for the two points $\mathbb{P}_G^0$ and $\mathbb{P}_{NO}^0$. The same applies to other common terms.

six corner points:

$$\mathbb{P}^1_{NO} = \big(I(X_1;Y), I(X_2;Y|X_1), I(X_3;Y|X_1,X_2)\big),$$
$$\mathbb{P}^2_{NO} = \big(I(X_1;Y), I(X_2;Y|X_1,X_3), I(X_3;Y|X_1)\big),$$
$$\mathbb{P}^3_{NO} = \big(I(X_1;Y|X_2), I(X_2;Y), I(X_3;Y|X_1,X_2)\big),$$
$$\mathbb{P}^4_{NO} = \big(I(X_1;Y|X_2,X_3), I(X_2;Y), I(X_3;Y|X_2)\big),$$
$$\mathbb{P}^5_{NO} = \big(I(X_1;Y|X_3), I(X_2;Y|X_1,X_3), I(X_3;Y)\big),$$
$$\mathbb{P}^6_{NO} = \big(I(X_1;Y|X_2,X_3), I(X_2;Y|X_3), I(X_3;Y)\big).$$

We have the following structure of these corner points.

Theorem 19: With channel transition matrix $p^E_{TLC}(y|v)$, the corner points satisfy $\mathbb{P}^0_G = \mathbb{P}^1_{NO}$. For SC decoding, we have point $\mathbb{P}^2_{NO} = \big(I(X_1;Y), 1, I(X_3;Y|X_1)\big)$, points $\mathbb{P}^3_{NO} = \mathbb{P}^4_{NO} = \big(1, I(X_2;Y), I(X_3;Y|X_1,X_2)\big)$, and points $\mathbb{P}^5_{NO} = \mathbb{P}^6_{NO} = \big(1,1,I(X_3;Y)\big)$.

Proof: See Appendix E. ∎

Let the uniform rate regions of TIN decoding and SC decoding under Gray labeling be $\mathscr{R}^{TIN}_G$ and $\mathscr{R}^{SC}_G$, respectively, and under NO labeling be $\mathscr{R}^{TIN}_{NO}$ and $\mathscr{R}^{SC}_{NO}$, respectively. We have the following theorem on the uniform rate regions.

Theorem 20: With channel transition matrix $p^E_{TLC}(y|v)$, the uniform rate regions satisfy $\mathscr{R}^{TIN}_{NO} \subset \mathscr{R}^{TIN}_G = \mathscr{R}^{SC}_G \subset \mathscr{R}^{SC}_{NO}$.

Proof: From Lemma 18, we have $\mathscr{R}^{TIN}_G = \mathscr{R}^{SC}_G$. From Theorem 19, it is clear that $\mathscr{R}^{SC}_G \subset \mathscr{R}^{SC}_{NO}$. Here, we only need to show that $\mathscr{R}^{TIN}_{NO} \subset \mathscr{R}^{TIN}_G$. We prove it by showing that $\mathbb{P}^0_G(1) = \mathbb{P}^0_{NO}(1)$, $\mathbb{P}^0_G(2) > \mathbb{P}^0_{NO}(2)$, and $\mathbb{P}^0_G(3) > \mathbb{P}^0_{NO}(3)$. First, from Theorem 19, $\mathbb{P}^0_G(1) = \mathbb{P}^1_{NO}(1)$. Since $\mathbb{P}^0_{NO}(1) = \mathbb{P}^1_{NO}(1)$, we have $\mathbb{P}^0_G(1) = \mathbb{P}^0_{NO}(1)$. Next, we show $\mathbb{P}^0_G(2) > \mathbb{P}^0_{NO}(2)$ as follows:

$$\begin{aligned}
&\mathbb{P}^0_G(2) - \mathbb{P}^0_{NO}(2)\\
&= \frac{1}{2}H\Big(\frac{\varepsilon_0}{4}, \frac{1-\varepsilon_0+\varepsilon_1}{4}, \frac{1-\varepsilon_1}{4}, \frac{\varepsilon_4}{4}, \frac{1-\varepsilon_4+\varepsilon_5}{4}, \frac{1-\varepsilon_5}{4}\Big)\\
&\quad + \frac{1}{2}H\Big(\frac{\varepsilon_2}{4}, \frac{1-\varepsilon_2+\varepsilon_3}{4}, \frac{1-\varepsilon_3}{4}, \frac{\varepsilon_6}{4}, \frac{1-\varepsilon_6+1}{4}\Big)\\
&\quad - \frac{1}{2}H\Big(\frac{\varepsilon_0}{4}, \frac{1-\varepsilon_0+\varepsilon_1}{4}, \frac{1-\varepsilon_1}{4}, \frac{\varepsilon_6}{4}, \frac{1-\varepsilon_6+1}{4}\Big)\\
&\quad - \frac{1}{2}H\Big(\frac{\varepsilon_2}{4}, \frac{1-\varepsilon_2+\varepsilon_3}{4}, \frac{1-\varepsilon_3+\varepsilon_4}{4}, \frac{1-\varepsilon_4+\varepsilon_5}{4}, \frac{1-\varepsilon_5}{4}\Big)\\
&= \frac{1}{8}\Big(f(1-\varepsilon_3+\varepsilon_4) - f(1-\varepsilon_3) - f(\varepsilon_4)\Big)\\
&= \frac{1}{8}\Big((1-\varepsilon_3)\log_2\Big(\frac{1-\varepsilon_3+\varepsilon_4}{1-\varepsilon_3}\Big) + \varepsilon_4\log_2\Big(\frac{1-\varepsilon_3+\varepsilon_4}{\varepsilon_4}\Big)\Big) > 0.
\end{aligned}$$

Similarly, we show $\mathbb{P}^0_G(3) > \mathbb{P}^0_{NO}(3)$:

$$\begin{aligned}
&\mathbb{P}^0_G(3) - \mathbb{P}^0_{NO}(3)\\
&= \frac{1}{2}H\Big(\frac{\varepsilon_0}{4}, \frac{1-\varepsilon_0}{4}, \frac{\varepsilon_2}{4}, \frac{1-\varepsilon_2}{4}, \frac{\varepsilon_4}{4}, \frac{1-\varepsilon_4}{4}, \frac{\varepsilon_6}{4}, \frac{1-\varepsilon_6}{4}\Big)\\
&\quad + \frac{1}{2}H\Big(\frac{\varepsilon_1}{4}, \frac{1-\varepsilon_1}{4}, \frac{\varepsilon_3}{4}, \frac{1-\varepsilon_3}{4}, \frac{\varepsilon_5}{4}, \frac{1-\varepsilon_5}{4}, \frac{1}{4}\Big)\\
&\quad - \frac{1}{2}H\Big(\frac{\varepsilon_0}{4}, \frac{1-\varepsilon_0}{4}, \frac{\varepsilon_3}{4}, \frac{1-\varepsilon_3+\varepsilon_4}{4}, \frac{1-\varepsilon_4}{4}, \frac{1}{4}\Big)\\
&\quad - \frac{1}{2}H\Big(\frac{\varepsilon_1}{4}, \frac{1-\varepsilon_1+\varepsilon_2}{4}, \frac{1-\varepsilon_2}{4}, \frac{\varepsilon_5}{4}, \frac{1-\varepsilon_5+\varepsilon_6}{4}, \frac{1-\varepsilon_6}{4}\Big)\\
&= \frac{1}{8}\Big(f(1-\varepsilon_1+\varepsilon_2) - f(1-\varepsilon_1) - f(\varepsilon_2)\Big)\\
&\quad + \frac{1}{8}\Big(f(1-\varepsilon_3+\varepsilon_4) - f(1-\varepsilon_3) - f(\varepsilon_4)\Big)\\
&\quad + \frac{1}{8}\Big(f(1-\varepsilon_5+\varepsilon_6) - f(1-\varepsilon_5) - f(\varepsilon_6)\Big) > 0.
\end{aligned}$$

∎

Remark 6: Similar to the MLC flash case (see Remark 3), for the TLC flash early-stage P/E cycling model, we can derive from Lemma 18, Theorem 19, and Theorem 20 several insights into efficient coding schemes. First, the sum rate of TIN decoding under Gray labeling is the same as that of SC decoding. This implies that an ECC solution for Gray labeling can use three point-to-point symmetric capacity-achieving codes for three pages separately. Second, with NO labeling, for SC decoding, 6 corner points are reduced to 4 points. Moreover, the rate $\big(R_1 = 1, R_2 = 1, R_3 = I(X_3;Y)\big)$ can be achieved, which means that we can apply a point-to-point symmetric capacity-achieving code to the page X_3, leaving the pages X_1 and X_2 uncoded.

Thanks to the similarity between the data retention model in Table VIII and the early-stage P/E cycling model in Table VII, the conclusions of Lemma 18, Theorem 19, and Theorem 20 also apply to the data retention model. □

Example 5: For the early-stage P/E cycling model with channel transition matrix $p^E_{TLC}(y|v)$ in Table VII, let $\varepsilon_i = 0.9$, $i = 0, 1, \ldots, 6$. The uniform rate regions of TIN and SC decodings under Gray and NO labelings are plotted in Fig. 8. It is shown that $\mathscr{R}^{TIN}_{NO} \subset \mathscr{R}^{TIN}_G = \mathscr{R}^{SC}_G \subset \mathscr{R}^{SC}_{NO}$. For SC decoding under NO labeling, the corner point $\mathbb{P}^5_{NO} = \mathbb{P}^6_{NO} = (R_1 = 1, R_2 = 1, R_3 = 0.5878)$ is achieved. For this particular point, coding is only needed for the page X_3.

B.2 Extension to All Labelings

Now, we extend the analysis to the entire set of $8! = 40320$ labelings. As in the case of MLC flash memory, we treat a labeling σ as a permutation π in the symmetric group $\mathcal{S}_8$ with group operation $*$. We write a permutation π in $\mathcal{S}_8$ as a vector $\pi = (w_0, w_1, w_2, w_3, w_4, w_5, w_6, w_7)$ (where w_i, $i = 0, 1, \ldots, 7$, represent the full set of possible 3-tuples) to represent that $\pi(111) = w_0$, $\pi(110) = w_1$, $\pi(101) = w_2$, $\pi(100) = w_3$, $\pi(011) = w_4$, $\pi(010) = w_5$, $\pi(001) = w_6$, and $\pi(000) = w_7$.

In the following, we will use a decimal labeling to represent a binary labeling for simplicity. For example, the binary labeling (111, 110, 101, 100, 011, 010, 001, 000) will be written as (7, 6, 5, 4, 3, 2, 1, 0). By utilizing the algebraic structure of the symmetric group $\mathcal{S}_8$, we have following two lemmas on the uniform rate region and sum rate, valid for any channel model characterized by an *arbitrary* channel transition matrix $p_{TLC}(y|v)$ for $y \in \mathcal{Y}_{TLC}$ and $v \in \mathcal{V}_{TLC}$. We omit their proofs, since they are similar to the MLC case.

Lemma 21: In the symmetric group $\mathcal{S}_8$, D_0={(7, 6, 5, 4, 3, 2, 1, 0), (6, 7, 4, 5, 2, 3, 0, 1), (5, 4, 7, 6, 1, 0, 3, 2), (4, 5, 6, 7, 0, 1, 2, 3), (3, 2, 1, 0, 7, 6, 5, 4), (2, 3, 0, 1, 6, 7, 4, 5), (1, 0, 3, 2, 5, 4, 7, 6), (0, 1, 2, 3, 4, 5, 6, 7)} forms an abelian subgroup which includes σ_{NO} =(7, 6, 5, 4, 3, 2, 1, 0). With an arbitrary channel transition matrix $p_{TLC}(y|v)$,

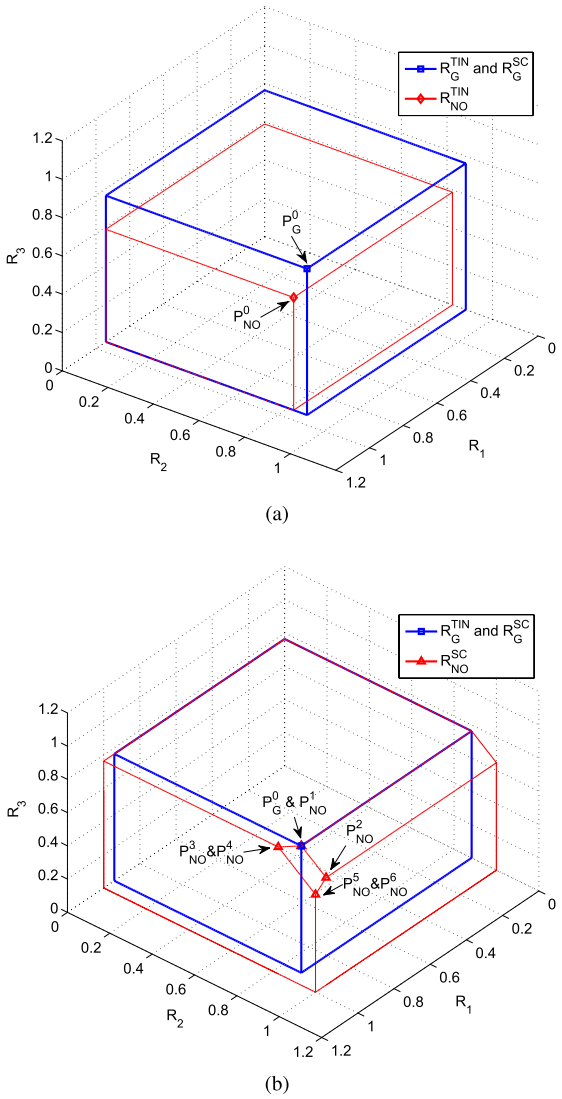


Fig. 8. (a) Illustration of uniform rate regions for TIN decoding under Gray labeling, $\mathscr{R}_G^{TIN}$; SC decoding under Gray labeling, $\mathscr{R}_G^{SC}$; and TIN decoding under NO labeling, $\mathscr{R}_{NO}^{TIN}$. (b) Illustration of uniform rate regions for TIN decoding under Gray labeling, $\mathscr{R}_G^{TIN}$; SC decoding under Gray labeling, $\mathscr{R}_G^{SC}$; and SC decoding under NO labeling, $\mathscr{R}_{NO}^{SC}$.

all labelings in D_0 give the same uniform rate region for a given decoding scheme (TIN or SC decoding).

Lemma 22: In the symmetric group $\mathcal{S}_8$, $\hat{D}_0$=$\{D_0 \cup D_0*(7, 6, 3, 2, 5, 4, 1, 0) \cup D_0*(7, 3, 5, 1, 6, 2, 4, 0) \cup D_0*(7, 5, 6, 4, 3, 1, 2, 0) \cup D_0*(7, 3, 6, 2, 5, 1, 4, 0) \cup D_0*(7, 5, 3, 1, 6, 4, 2, 0)\}$ forms an abelian subgroup which includes σ_{NO} =(7, 6, 5, 4, 3, 2, 1, 0). With an arbitrary channel transition matrix $p_{TLC}(y|v)$, all labelings in $\hat{D}_0$ give the same sum rate for a given decoding scheme (TIN or SC decoding).

Remark 7: With subgroup D_0 in Lemma 21, we can partition $\mathcal{S}_8$ into D_0 and its 5039 cosets, each of size 8. All labelings in each coset give the same uniform rate region for a given decoding scheme (TIN or SC decoding). With subgroup $\hat{D}_0$ in Lemma 22, the group $\mathcal{S}_8$ can be partitioned into $\hat{D}_0$ and its 839 cosets, each of size 48. All labelings in each coset give the same sum rate for TIN decoding. Note that for SC decoding, all labelings in $\mathcal{S}_8$ give the same sum rate, due to the uniform distributions for X_1, X_2, and X_3.

Thus, for an arbitrary channel transition matrix $p_{TLC}(y|v)$, for TIN decoding, among all the labelings in $\mathcal{S}_8$, the 48 labelings in the coset $\hat{D}_0 * (7, 6, 4, 5, 1, 0, 2, 3)$, which includes Gray labeling, give the largest sum rate, which equals that of SC decoding. However, for some particular channel transition matrix, for TIN decoding, more than 48 labelings may give the largest sum rate. For example, for the channel transition matrix $p_{TLC}^E(y|v)$ in Table VII, for TIN decoding, among all the labelings in $\mathcal{S}_8$, we find that the 144 labelings in the set $\tilde{D} = \{\hat{D}_0 * (7, 6, 4, 5, 1, 0, 2, 3) \cup \hat{D}_0 * (7, 6, 4, 0, 2, 3, 1, 5) \cup \hat{D}_0 * (7, 6, 2, 3, 1, 5, 4, 0)\}$ give the largest sum rate. □

Finally, we discuss the uniform rate region if we are allowed to use multiple labelings together for each codeword, instead of one labeling in the entire TLC flash memory. Define $\mathscr{R}_{\mathcal{S}_8}^{TIN} = Conv\left(\bigcup_{\sigma \in \mathcal{S}_8} \mathscr{R}_\sigma^{TIN}\right)$, the convex hull of uniform rate regions of all 40320 labelings for TIN decoding. Define $\mathscr{R}_{\mathcal{S}_8}^{SC} = Conv\left(\bigcup_{\sigma \in \mathcal{S}_8} \mathscr{R}_\sigma^{SC}\right)$, the convex hull of uniform rate regions of all 40320 labelings for SC decoding. Through time-sharing of different labelings, for TIN decoding, any point $(R_1, R_2, R_3) \in \mathscr{R}_{\mathcal{S}_8}^{TIN}$ can be achieved. For SC decoding, any point $(R_1, R_2, R_3) \in \mathscr{R}_{\mathcal{S}_8}^{SC}$ can be achieved. Moreover, for the early-stage P/E cycling model with channel transition matrix $p_{TLC}^E(y|v)$ in Table VII, the region $\mathscr{R}_{\mathcal{S}_8}^{SC}$ can be determined explicitly. The following theorem, analogous to Theorem 13, is stated without proof.

Theorem 23: For the early-stage P/E cycling model, the rate region $\mathscr{R}_{\mathcal{S}_8}^{SC}$ is the set of all rate tuples (R_1, R_2, R_3) such that

1) $R_1 \le 1,\ R_2 \le 1,\ R_3 \le 1$;
2) $R_1 + R_2 + R_3 \le I(X_1, X_2, X_3; Y)$.

V. Conclusion

We analyzed the performance of multilevel flash memories with different decoding schemes and cell voltage labelings from a multi-user perspective. In MLC flash memory, we showed that both TIN and SC decoding outperform the current default decoding scheme in terms of both uniform rate region and sum rate. For the P/E cycling model, with TIN decoding, we found that 8 labelings, including the standard Gray labeling, offer the largest sum rate among all 24 possible labelings. The sum rate of TIN decoding under Gray labeling equals that of SC decoding at the early-stage of P/E cycling, and is smaller than but close to that of SC decoding at the late-stage of P/E cycling. It was also shown that additional read thresholds can effectively enhance the rate region and sum rate. For TLC flash memory, the structure and properties of all 40320 labelings were investigated. We studied the early-stage P/E cycling model, which is similar to the data retention model. For TIN decoding, 144 labelings including Gray labeling give the largest sum rate among all 40320 labelings. Moreover, surprisingly, for SC decoding under NO labeling, the rate $(R_1 = 1, R_2 = 1, R_3)$ can be achieved, implying a simple coding solution in which the first two of the three pages can remain uncoded.

APPENDIX A
PROOF OF THEOREM 1

Proof: We bound the value $r_s^{SC} - r_s^{TIN}$ as follows

$$\begin{aligned} r_s^{SC} - r_s^{TIN} &= I(X_1, X_2; Y) - I(X_1; Y) - I(X_2; Y) \\ &= I(X_2; Y|X_1) - I(X_2; Y) \\ &= H(X_2|X_1) - H(X_2|X_1, Y) \\ &\quad -\Big(H(X_2) - H(X_2|Y)\Big) \\ &\overset{(a)}{=} I(X_1; X_2|Y) \\ &= \sum_{j=0}^{q-1}\Big(\sum_{i=0}^{3}\frac{p_{i,s_j}}{4}\Big) I(X_1; X_2|Y = s_j) \geq 0, \end{aligned}$$

where in step (a) we use $H(X_2|X_1) = H(X_2)$ which follows from the fact that X_1 and X_2 are independent.

Now, $I(X_1; X_2|Y) \geq 0$ with equality if and only if X_1 and X_2 are conditionally independent given $Y = s_j$, i.e., $P(X_1, X_2|Y = s_j) = P(X_1|Y = s_j)P(X_2|Y = s_j)$. Thus, we need to check four cases:

1) $P(X_1 = 1, X_2 = 1|Y = s_j) = P(X_1 = 1|Y = s_j)P(X_2 = 1|Y = s_j)$, i.e., $\frac{p_{3,s_j}}{\sum_{i=0}^{3} p_{i,s_j}} = \frac{p_{3,s_j}+p_{2,s_j}}{\sum_{i=0}^{3} p_{i,s_j}} \frac{p_{3,s_j}+p_{1,s_j}}{\sum_{i=0}^{3} p_{i,s_j}}$.

2) $P(X_1 = 1, X_2 = 0|Y = s_j) = P(X_1 = 1|Y = s_j)P(X_2 = 0|Y = s_j)$, i.e., $\frac{p_{2,s_j}}{\sum_{i=0}^{3} p_{i,s_j}} = \frac{p_{3,s_j}+p_{2,s_j}}{\sum_{i=0}^{3} p_{i,s_j}} \frac{p_{2,s_j}+p_{0,s_j}}{\sum_{i=0}^{3} p_{i,s_j}}$.

3) $P(X_1 = 0, X_2 = 1|Y = s_j) = P(X_1 = 0|Y = s_j)P(X_2 = 1|Y = s_j)$, i.e., $\frac{p_{1,s_j}}{\sum_{i=0}^{3} p_{i,s_j}} = \frac{p_{1,s_j}+p_{0,s_j}}{\sum_{i=0}^{3} p_{i,s_j}} \frac{p_{3,s_j}+p_{1,s_j}}{\sum_{i=0}^{3} p_{i,s_j}}$.

4) $P(X_1 = 0, X_2 = 0|Y = s_j) = P(X_1 = 0|Y = s_j)P(X_2 = 0|Y = s_j)$, i.e., $\frac{p_{0,s_j}}{\sum_{i=0}^{3} p_{i,s_j}} = \frac{p_{1,s_j}+p_{0,s_j}}{\sum_{i=0}^{3} p_{i,s_j}} \frac{p_{2,s_j}+p_{0,s_j}}{\sum_{i=0}^{3} p_{i,s_j}}$.

To satisfy conditions 1) – 4), we have $p_{3,s_j}p_{0,s_j} = p_{2,s_j}p_{1,s_j}$ for all $j = 0, 1, \ldots, q-1$.

Finally, assuming $r_s^{TIN} = r_s^{SC}$, i.e., $I(X_1; Y) + I(X_2; Y) = I(X_1, X_2; Y)$, since $I(X_1, X_2; Y) = I(X_1; Y) + I(X_2; Y|X_1) = I(X_2; Y) + I(X_1; Y|X_2)$, we have $I(X_1; Y) = I(X_1; Y|X_2)$ and $I(X_2; Y) = I(X_2; Y|X_1)$, which means $\mathscr{R}^{TIN} = \mathscr{R}^{SC}$. ∎

APPENDIX B
PROOF OF THEOREM 2

Proof: We bound the value $r_s^{SC} - r_s^{TIN}$ as

$$\begin{aligned} r_s^{SC} - r_s^{TIN} &= I(X_1, X_2; Y) - I(X_1; Y) - I(X_2; Y) \\ &= I(X_1; X_2|Y) = H(X_1|Y) - H(X_1|X_2, Y) \\ &\overset{(a)}{\leq} H(X_1) - H(X_1|X_2, Y) \overset{(b)}{\leq} H(X_1) = 1, \end{aligned}$$

where step (a) follows from $H(X_1|Y) \leq H(X_1)$, and step (b) is due to $H(X_1|X_2, Y) \geq 0$. Thus, $r_s^{SC} - r_s^{TIN} = 1$ if and only if 1) $H(X_1|Y) = H(X_1) = 1$, and 2) $H(X_1|X_2, Y) = 0$.

The condition $H(X_1|Y) = \sum_{j=0}^{q-1}(\sum_{i=0}^{3}\frac{p_{i,s_j}}{4})H(X_1|Y = s_j) = 1$ holds if and only if $p_{3,s_j} + p_{2,s_j} = p_{1,s_j} + p_{0,s_j}$ for all $j = 0, 1, \ldots, q-1$. It means that even if Y is given, X_1 is still completely random to the observer. Similarly, $H(X_1|X_2, Y) = 0$ requires that $p_{3,s_j}p_{1,s_j} = p_{2,s_j}p_{0,s_j} = 0$ for all $j = 0, 1, \ldots, q-1$. It indicates that if X_2 and Y are obtained, then X_1 can be determined. ∎

APPENDIX C
PROOF OF LEMMA 14

Proof: We first prove that the sum rate $\hat{r}_s^{SC} \geq r_s^{SC}$, i.e., $I(X_1, X_2; \hat{Y}) \geq I(X_1, X_2; Y)$, and give the condition when the equality holds. The value $I(X_1, X_2; Y)$ can be expressed as follows:

$$\begin{aligned} I(X_1, X_2; Y) &= H(Y) - H(Y|X_1, X_2) \\ &= H\Big(\frac{\sum_{i=0}^{3} p_{i,s_0}}{4}, \frac{\sum_{i=0}^{3} p_{i,s_1}}{4}, \ldots, \frac{\sum_{i=0}^{3} p_{i,s_{q-1}}}{4}\Big) \\ &\quad -\frac{1}{4}\sum_{i=0}^{3} H(p_{i,s_0}, p_{i,s_1}, \ldots, p_{i,s_{q-1}}). \end{aligned}$$

Similarly, the value $I(X_1, X_2; \hat{Y})$ is

$$\begin{aligned} &I(X_1, X_2; \hat{Y}) = H(\hat{Y}) - H(\hat{Y}|X_1, X_2) \\ &= H\Big(\frac{\sum_{i=0}^{3} p_{i,s_0^1}}{4}, \frac{\sum_{i=0}^{3} p_{i,s_0^2}}{4}, \frac{\sum_{i=0}^{3} p_{i,s_1}}{4}, \ldots, \frac{\sum_{i=0}^{3} p_{i,s_{q-1}}}{4}\Big) \\ &\quad -\frac{1}{4}\sum_{i=0}^{3} H(p_{i,s_0^1}, p_{i,s_0^2}, p_{i,s_1}, \ldots, p_{i,s_{q-1}}) \\ &\overset{(a)}{=} H\Big(\frac{\sum_{i=0}^{3} p_{i,s_0}}{4}, \frac{\sum_{i=0}^{3} p_{i,s_1}}{4}, \ldots, \frac{\sum_{i=0}^{3} p_{i,s_{q-1}}}{4}\Big) \\ &\quad +\frac{\sum_{i=0}^{3} p_{i,s_0}}{4} H\Big(\frac{\sum_{i=0}^{3} p_{i,s_0^1}}{\sum_{i=0}^{3} p_{i,s_0}}, \frac{\sum_{i=0}^{3} p_{i,s_0^2}}{\sum_{i=0}^{3} p_{i,s_0}}\Big) \\ &\quad -\frac{1}{4}\sum_{i=0}^{3} H(p_{i,s_0}, p_{i,s_1}, \ldots, p_{i,s_{q-1}}) \\ &\quad -\frac{1}{4}\sum_{i=0}^{3} p_{i,s_0} H\Big(\frac{p_{i,s_0^1}}{p_{i,s_0}}, \frac{p_{i,s_0^2}}{p_{i,s_0}}\Big), \end{aligned}$$

where step (a) is from the grouping property of entropy [24] and $p_{i,s_0} = p_{i,s_0^1} + p_{i,s_0^2}$ for $i = 0, 1, 2, 3$.

The difference between $I(X_1, X_2; \hat{Y})$ and $I(X_1, X_2; Y)$ is

$$\begin{aligned} &I(X_1, X_2; \hat{Y}) - I(X_1, X_2; Y) \\ &= \frac{\sum_{i=0}^{3} p_{i,s_0}}{4} H\Big(\frac{\sum_{i=0}^{3} p_{i,s_0^1}}{\sum_{i=0}^{3} p_{i,s_0}}, \frac{\sum_{i=0}^{3} p_{i,s_0^2}}{\sum_{i=0}^{3} p_{i,s_0}}\Big) \\ &\quad -\frac{1}{4}\sum_{i=0}^{3} p_{i,s_0} H\Big(\frac{p_{i,s_0^1}}{p_{i,s_0}}, \frac{p_{i,s_0^2}}{p_{i,s_0}}\Big) \\ &= \frac{\sum_{i=0}^{3} p_{i,s_0}}{4}\Big(-\frac{1}{\sum_{i=0}^{3} p_{i,s_0}}\big(f(\sum_{i=0}^{3} p_{i,s_0^1}) + f(\sum_{i=0}^{3} p_{i,s_0^2})\big) \\ &\qquad +\log_2(\sum_{i=0}^{3} p_{i,s_0})\Big) \\ &\quad -\frac{1}{4}\sum_{i=0}^{3} p_{i,s_0}\Big(-\frac{1}{p_{i,s_0}}\big(f(p_{i,s_0^1}) + f(p_{i,s_0^2})\big) + \log_2(p_{i,s_0})\Big) \\ &= \frac{1}{4}\Big(f(\sum_{i=0}^{3} p_{i,s_0}) - f(\sum_{i=0}^{3} p_{i,s_0^1}) - f(\sum_{i=0}^{3} p_{i,s_0^2})\Big) \\ &\quad -\frac{1}{4}\sum_{i=0}^{3}\Big(f(p_{i,s_0}) - f(p_{i,s_0^1}) - f(p_{i,s_0^2})\Big). \end{aligned}$$

Note that the difference is only related to the probabilities p_{i,s_0}, p_{i,s_0^1}, and p_{i,s_0^2} for $i = 0, 1, 2, 3$.

To prove that $I(X_1, X_2; \hat{Y}) \geq I(X_1, X_2; Y)$, we define a new function $g(u_1, u_2) = f(u_1 + u_2) - f(u_1) - f(u_2)$ and utilize its properties. We first prove $g(u_1 + v_1, u_2 + v_2) \geq g(u_1, u_2) + g(v_1, v_2)$ as follows:

$$\begin{aligned} & g(u_1+v_1, u_2+v_2) - \Big(g(u_1,u_2) + g(v_1,v_2)\Big) \\ &= (u_1+u_2)\log_2(1+\frac{v_1+v_2}{u_1+u_2}) + (v_1+v_2)\log_2(1+\frac{u_1+u_2}{v_1+v_2}) \\ &\quad -\Big(u_1\log_2(1+\frac{v_1}{u_1}) + v_1\log_2(1+\frac{u_1}{v_1})\Big) \\ &\quad -\Big(u_2\log_2(1+\frac{v_2}{u_2}) + v_2\log_2(1+\frac{u_2}{v_2})\Big). \end{aligned}$$

The function $t\log_2(1 + 1/t)$ is concave. Let $t_1 = \frac{u_1}{v_1}$, $t_2 = \frac{u_2}{v_2}$, $r_1 = \frac{v_1}{v_1+v_2}$, and $r_2 = \frac{v_2}{v_1+v_2}$. We have $(r_1t_1 + r_2t_2)\log_2\Big(1+1/(r_1t_1+r_2t_2)\Big) \geq r_1t_1\log_2(1+1/t_1)+r_2t_2\log_2(1+1/t_2)$; that is, $(u_1 + u_2)\log_2(1 + \frac{v_1+v_2}{u_1+u_2}) \geq u_1\log_2(1 + \frac{v_1}{u_1}) + u_2\log_2(1+\frac{v_2}{u_2})$. Similarly, $(v_1+v_2)\log_2(1+\frac{u_1+u_2}{v_1+v_2}) \geq v_1\log_2(1+\frac{u_1}{v_1})+ v_2\log_2(1 + \frac{u_2}{v_2})$. Therefore, $g(u_1 + v_1, u_2 + v_2) \geq g(u_1, u_2) + g(v_1, v_2)$, where equality holds if and only if $\frac{u_1}{u_2} = \frac{v_1}{v_2}$.

Now, we apply $g(u_1 + v_1, u_2 + v_2) \geq g(u_1, u_2) + g(v_1, v_2)$ twice and have

$$\begin{aligned} & f(\sum_{i=0}^{3} p_{i,s_0}) - f(\sum_{i=0}^{3} p_{i,s_0^1}) - f(\sum_{i=0}^{3} p_{i,s_0^2}) \\ &= g(p_{0,s_0^1} + p_{1,s_0^1} + p_{2,s_0^1} + p_{3,s_0^1}, p_{0,s_0^2} + p_{1,s_0^2} \\ &\quad + p_{2,s_0^2} + p_{3,s_0^2}) \\ &\geq g(p_{0,s_0^1} + p_{1,s_0^1}, p_{0,s_0^2} + p_{1,s_0^2}) \\ &\quad + g(p_{2,s_0^1} + p_{3,s_0^1}, p_{2,s_0^2} + p_{3,s_0^2}) \\ &\geq g(p_{0,s_0^1}, p_{0,s_0^2}) + g(p_{1,s_0^1}, p_{1,s_0^2}) + g(p_{2,s_0^1}, p_{2,s_0^2}) \\ &\quad + g(p_{3,s_0^1}, p_{3,s_0^2}) \\ &= \sum_{i=0}^{3} \Big(f(p_{i,s_0}) - f(p_{i,s_0^1}) - f(p_{i,s_0^2})\Big), \end{aligned}$$

where equality holds if and only if $\frac{p_{0,s_0^1}}{p_{0,s_0^2}} = \frac{p_{1,s_0^1}}{p_{1,s_0^2}} = \frac{p_{2,s_0^1}}{p_{2,s_0^2}} = \frac{p_{3,s_0^1}}{p_{3,s_0^2}}$. Thus, we have proved $I(X_1, X_2; \hat{Y}) \geq I(X_1, X_2; Y)$.

With the same proof technique, we can prove $I(X_1; \hat{Y}) \geq I(X_1; Y)$, $I(X_2; \hat{Y}) \geq I(X_2; Y)$, $I(X_1; \hat{Y}|X_2) \geq I(X_1; Y|X_2)$, and $I(X_2; \hat{Y}|X_1) \geq I(X_2; Y|X_1)$. These inequalities lead to $r_s^{TIN} \leq \hat{r}_s^{TIN}$, $\mathscr{R}^{TIN} \subseteq \hat{\mathscr{R}}^{TIN}$, and $\mathscr{R}^{SC} \subseteq \hat{\mathscr{R}}^{SC}$. ∎

APPENDIX D
PROOF OF THEOREM 15

Proof: We calculate $r_s^{SC} - r_s^{TIN}$ as follows:

$$\begin{aligned} r_s^{SC} - r_s^{TIN} &= I(X_1, X_2, X_3; Y) - \sum_{i=1}^{3} I(X_i; Y) \\ &= I(X_1; Y) + I(X_2; Y|X_1) + I(X_3; Y|X_1, X_2) \\ &\quad - \sum_{i=1}^{3} I(X_i; Y) \\ &= H(X_2|X_1) - H(X_2|Y, X_1) + H(X_3|X_1, X_2) \\ &\quad - H(X_3|Y, X_1, X_2) \\ &\quad - \Big(H(X_2) - H(X_2|Y) + H(X_3) - H(X_3|Y)\Big) \\ &\overset{(a)}{=} I(X_1; X_2|Y) + I(X_1, X_2; X_3|Y) \geq 0, \end{aligned}$$

where step (a) is due to the fact that X_1, X_2, and X_3 are independent. The last step is from that $I(X_1; X_2|Y) \geq 0$ and $I(X_1, X_2; X_3|Y) \geq 0$. Therefore, $r_s^{TIN} = r_s^{SC}$ if and only if $I(X_1; X_2|Y) = 0$ and $I(X_1, X_2; X_3|Y) = 0$.

The condition $I(X_1; X_2|Y) = 0$ will be satisfied if and only if X_1 and X_2 are conditionally independent given Y. As in the proof of Theorem 1, the condition $I(X_1; X_2|Y) = \sum_{j=0}^{q-1}(\sum_{i=0}^{7} \frac{p_{i,s_j}}{8})I(X_1; X_2|Y = s_j) = 0$ leads to the first condition (2a) in the statement of the theorem. Similarly, $I(X_1, X_2; X_3|Y) = \sum_{j=0}^{q-1}(\sum_{i=0}^{7} \frac{p_{i,s_j}}{8})I(X_1, X_2; X_3|Y = s_j) = 0$ results in the last four conditions $(2b) - (2e)$ in the statement of the theorem.

Next, we prove $\mathscr{R}^{TIN} = \mathscr{R}^{SC}$ if $r_s^{TIN} = r_s^{SC}$. Assuming $r_s^{TIN} = r_s^{SC}$, from above we have already shown that $I(X_1; X_2|Y) = 0$ and $I(X_1, X_2; X_3|Y) = 0$. Similarly, we have $I(X_1; X_3|Y)=0$, $I(X_1, X_3; X_2|Y)=0$, $I(X_2; X_3|Y)=0$, and $I(X_2, X_3; X_1|Y) = 0$. To prove $\mathscr{R}^{TIN} = \mathscr{R}^{SC}$, we need to show:

1) $I(X_1; Y|X_2, X_3) = I(X_1; Y)$; 2) $I(X_2, X_3; Y|X_1) = I(X_2; Y) + I(X_3; Y)$;

3) $I(X_2; Y|X_1, X_3) = I(X_2; Y)$; 4) $I(X_1, X_3; Y|X_2) = I(X_1; Y) + I(X_3; Y)$;

5) $I(X_3; Y|X_1, X_2) = I(X_3; Y)$; 6) $I(X_1, X_2; Y|X_3) = I(X_1; Y) + I(X_2; Y)$.

In the following, we will prove equations 5) and 6) of the above six equations. The other four equations can be proved in a similar way. For the first, we have

$$\begin{aligned} & I(X_3; Y|X_1, X_2) - I(X_3; Y) \\ &= H(X_3|X_1, X_2) - H(X_3|Y, X_1, X_2) - \Big(H(X_3) - H(X_3|Y)\Big) \\ &\overset{(a)}{=} H(X_3|Y) - H(X_3|Y, X_1, X_2) = I(X_3; X_1, X_2|Y) = 0, \end{aligned}$$

where step (a) is due to the fact that X_1, X_2, and X_3 are independent. Similarly, for the second,

$$\begin{aligned} & I(X_1, X_2; Y|X_3) - I(X_1; Y) - I(X_2; Y) \\ &= H(X_1, X_2|X_3) - H(X_1, X_2|Y, X_3) - \Big(H(X_1) - H(X_1|Y)\Big) \\ &\quad - \Big(H(X_2) - H(X_2|Y)\Big) \\ &\overset{(a)}{=} H(X_1|Y) + H(X_2|Y) - H(X_1, X_2|Y, X_3) \\ &\overset{(b)}{=} H(X_1|Y) + H(X_2|Y) - H(X_1, X_2|Y) = I(X_1; X_2|Y) = 0, \end{aligned}$$

where step (a) is due to the fact that X_1, X_2, and X_3 are independent, and step (b) is from $I(X_1, X_2; X_3|Y) = H(X_1, X_2|Y) - H(X_1, X_2|Y, X_3) = 0$. ∎

APPENDIX E
PROOF OF THEOREM 19

Proof: To prove $\mathbb{P}_G^0 = \mathbb{P}_{NO}^1$, we first show the first coordinate $\mathbb{P}_G^0(1)$ of point $\mathbb{P}_G^0$ is equal to the first coordinate

$\mathbb{P}^1_{NO}(1)$ of point $\mathbb{P}^1_{NO}$.

$$\begin{aligned}
&\mathbb{P}^0_G(1)\\
&= I(X_1;Y) = H(Y) - H(Y|X_1)\\
&= H(\frac{\varepsilon_0}{8}, \frac{1-\varepsilon_0+\varepsilon_1}{8}, \frac{1-\varepsilon_1+\varepsilon_2}{8}, \frac{1-\varepsilon_2+\varepsilon_3}{8},\\
&\qquad \frac{1-\varepsilon_3+\varepsilon_4}{8}, \frac{1-\varepsilon_4+\varepsilon_5}{8}, \frac{1-\varepsilon_5+\varepsilon_6}{8}, \frac{1-\varepsilon_6+1}{8})\\
&\quad -\frac{1}{2}H(\frac{\varepsilon_0}{4}, \frac{1-\varepsilon_0+\varepsilon_1}{4}, \frac{1-\varepsilon_1+\varepsilon_2}{4}, \frac{1-\varepsilon_2+\varepsilon_3}{4}, \frac{1-\varepsilon_3}{4})\\
&\quad -\frac{1}{2}H(\frac{\varepsilon_4}{4}, \frac{1-\varepsilon_4+\varepsilon_5}{4}, \frac{1-\varepsilon_5+\varepsilon_6}{4}, \frac{1-\varepsilon_6+1}{4})\\
&= \mathbb{P}^1_{NO}(1).
\end{aligned}$$

Now, we show that the second coordinate $\mathbb{P}^0_G(2)$ of point $\mathbb{P}^0_G$ equals the second coordinate $\mathbb{P}^1_{NO}(2)$ of point $\mathbb{P}^1_{NO}$.

$$\begin{aligned}
&\mathbb{P}^0_G(2) = I(X_2;Y) = H(Y) - H(Y|X_2)\\
&= H(\frac{\varepsilon_0}{8}, \frac{1-\varepsilon_0+\varepsilon_1}{8}, \frac{1-\varepsilon_1+\varepsilon_2}{8}, \frac{1-\varepsilon_2+\varepsilon_3}{8},\\
&\qquad \frac{1-\varepsilon_3+\varepsilon_4}{8}, \frac{1-\varepsilon_4+\varepsilon_5}{8}, \frac{1-\varepsilon_5+\varepsilon_6}{8}, \frac{1-\varepsilon_6+1}{8})\\
&\quad -\frac{1}{2}H(\frac{\varepsilon_0}{4}, \frac{1-\varepsilon_0+\varepsilon_1}{4}, \frac{1-\varepsilon_1}{4}, \frac{\varepsilon_6}{4}, \frac{1-\varepsilon_6+1}{4})\\
&\quad -\frac{1}{2}H(\frac{\varepsilon_2}{4}, \frac{1-\varepsilon_2+\varepsilon_3}{4}, \frac{1-\varepsilon_3+\varepsilon_4}{4}, \frac{1-\varepsilon_4+\varepsilon_5}{4}, \frac{1-\varepsilon_5}{4})\\
&\overset{(a)}{=} H(\frac{\varepsilon_0}{8}, \frac{1-\varepsilon_0+\varepsilon_1}{8}, \frac{1-\varepsilon_1+\varepsilon_2}{8}, \frac{1-\varepsilon_2+\varepsilon_3}{8},\\
&\qquad \frac{1-\varepsilon_3}{8}, \frac{\varepsilon_4}{8}, \frac{1-\varepsilon_4+\varepsilon_5}{8}, \frac{1-\varepsilon_5+\varepsilon_6}{8}, \frac{1-\varepsilon_6+1}{8})\\
&\quad -\frac{1-\varepsilon_3+\varepsilon_4}{8}H(\frac{1-\varepsilon_3}{1-\varepsilon_3+\varepsilon_4}, \frac{\varepsilon_4}{1-\varepsilon_3+\varepsilon_4})\\
&\quad -\frac{1}{2}H(\frac{\varepsilon_0}{4}, \frac{1-\varepsilon_0+\varepsilon_1}{4}, \frac{1-\varepsilon_1}{4}, \frac{\varepsilon_6}{4}, \frac{1-\varepsilon_6+1}{4})\\
&\quad -\frac{1}{2}\Big(H(\frac{\varepsilon_2}{4}, \frac{1-\varepsilon_2+\varepsilon_3}{4}, \frac{1-\varepsilon_3}{4}, \frac{\varepsilon_4}{4}, \frac{1-\varepsilon_4+\varepsilon_5}{4}, \frac{1-\varepsilon_5}{4})\\
&\qquad -\frac{1-\varepsilon_3+\varepsilon_4}{4}H(\frac{1-\varepsilon_3}{1-\varepsilon_3+\varepsilon_4}, \frac{\varepsilon_4}{1-\varepsilon_3+\varepsilon_4})\Big)\\
&\overset{(b)}{=} \frac{1}{2}H(\frac{\varepsilon_0}{4}, \frac{1-\varepsilon_0+\varepsilon_1}{4}, \frac{1-\varepsilon_1+\varepsilon_2}{4}, \frac{1-\varepsilon_2+\varepsilon_3}{4}, \frac{1-\varepsilon_3}{4})\\
&\quad +\frac{1}{2}H(\frac{\varepsilon_4}{4}, \frac{1-\varepsilon_4+\varepsilon_5}{4}, \frac{1-\varepsilon_5+\varepsilon_6}{4}, \frac{2-\varepsilon_6}{4})\\
&\quad -\frac{1}{4}H(\frac{\varepsilon_0}{2}, \frac{1-\varepsilon_0+\varepsilon_1}{2}, \frac{1-\varepsilon_1}{2}) - \frac{1}{4}H(\frac{\varepsilon_2}{2}, \frac{1-\varepsilon_2+\varepsilon_3}{2},\\
&\qquad \frac{1-\varepsilon_3}{2})\\
&\quad -\frac{1}{4}H(\frac{\varepsilon_4}{2}, \frac{1-\varepsilon_4+\varepsilon_5}{2}, \frac{1-\varepsilon_5}{2}) - \frac{1}{4}H(\frac{\varepsilon_6}{2}, \frac{2-\varepsilon_6}{2})\\
&= \mathbb{P}^1_{NO}(2)
\end{aligned}$$

where steps (a) and (b) follow from the grouping property of entropy [24].

Since the sum of the three coordinates of point $\mathbb{P}^0_G$ is the same as that of point $\mathbb{P}^1_{NO}$, the third coordinate of point $\mathbb{P}^0_G$ is also the same as that of point $\mathbb{P}^1_{NO}$. Thus, we have proved $\mathbb{P}^0_G = \mathbb{P}^1_{NO}$.

Next, the second coordinate $\mathbb{P}^2_{NO}(2)$ of point $\mathbb{P}^2_{NO}$ is 1, since $\mathbb{P}^2_{NO}(2) = I(X_2;Y|X_1,X_3) = H(X_2|X_1,X_3) - H(X_2|X_1,X_3,Y) = H(X_2) = 1$.

For point $\mathbb{P}^3_{NO}$, the first coordinate is $\mathbb{P}^3_{NO}(1) = I(X_1;Y|X_2) = H(X_1|X_2) - H(X_1|X_2,Y) = H(X_1) = 1$. For point $\mathbb{P}^4_{NO}$, the first coordinate is $\mathbb{P}^4_{NO}(1) = I(X_1;Y|X_2,X_3) = H(X_1|X_2,X_3) - H(X_1|X_2,X_3,Y) = H(X_1) = 1$. Thus, the first and second coordinates of points $\mathbb{P}^3_{NO}$ and $\mathbb{P}^4_{NO}$ are the same. Since the sum of the three coordinates of point $\mathbb{P}^3_{NO}$ is the same as that of point $\mathbb{P}^4_{NO}$, the third coordinate of point $\mathbb{P}^3_{NO}$ is also equal to that of point $\mathbb{P}^4_{NO}$.

Similarly, to prove $\mathbb{P}^5_{NO} = \mathbb{P}^6_{NO}$, we only need to show that their first two coordinates are the same. For point $\mathbb{P}^5_{NO}$, the first coordinate is $\mathbb{P}^5_{NO}(1) = I(X_1;Y|X_3) = H(X_1|X_3) - H(X_1|X_3,Y) = H(X_1) = 1$, and the second coordinate is $\mathbb{P}^5_{NO}(2) = \mathbb{P}^2_{NO}(2) = 1$. For point $\mathbb{P}^6_{NO}$, the first coordinate is $\mathbb{P}^6_{NO}(1) = \mathbb{P}^4_{NO}(1) = 1$ and the second coordinate is $\mathbb{P}^6_{NO}(2) = I(X_2;Y|X_3) = H(X_2|X_3) - H(X_2|X_3,Y) = H(X_2) = 1$. ■

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers for a number of very helpful comments and suggestions. Portions of this work were performed while the third author was visiting the Center for Memory and Recording Research (CMRR) at UC San Diego. He is thankful for the hospitality and financial support provided by the Center during his stay.

REFERENCES

[1] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[2] M. Asadi, X. Huang, A. Kavcic, and N. P. Santhanam, "Optimal detector for multilevel NAND flash memory channels with intercell interference," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 825–835, May 2014.

[3] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, "Introduction to flash memory," *Proc. IEEE*, vol. 91, no. 4, pp. 489–502, Apr. 2003.

[4] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[5] Y. Cai, E. F. Haratsch, O. Mutlu, and K. Mai, "Error patterns in MLC NAND flash memory: Measurement, characterization, and analysis," in *Proc. IEEE Design, Autom., Test Eur. Conf. Exhibit. (DATE)*, Dresden, Germany, Mar. 2012, pp. 521–526.

[6] Y. Cai, E. F. Haratsch, O. Mutlu, and K. Mai, "Threshold voltage distribution in MLC NAND flash memory: Characterization, analysis, and modeling," in *Proc. IEEE Design, Autom. Test Eur. Conf. Exhibit. (DATE)*, Grenoble, France, Mar. 2013, pp. 1285–1290.

[7] G. Caire, G. Taricco, and E. Biglieri, "Capacity of bit-interleaved channels," *Electron. Lett.*, vol. 32, no. 12, pp. 1060–1061, Jun. 1996.

[8] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[9] A. R. Calderbank, "Multilevel codes and multistage decoding," *IEEE Trans. Commun.*, vol. 37, no. 3, pp. 222–229, Mar. 1989.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.

[11] G. Dong, S. Li, and T. Zhang, "Using data postcompensation and predistortion to tolerate cell-to-cell interference in MLC NAND flash memory," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 10, pp. 2718–2728, Oct. 2010.

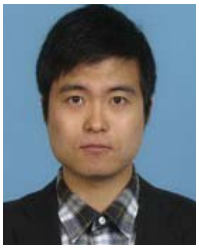
[12] G. Dong, N. Xie, and T. Zhang, "On the use of soft-decision error-correction codes in NAND flash memory," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 2, pp. 429–439, Feb. 2011.

[13] A. El Gamal and Y.-H. Kim, *Network Information Theory*. New York, NY, USA: Cambridge Univ. Press, 2011.

[14] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2005.

[15] P. Huang, P. H. Siegel, and E. Yaakobi, "Performance of flash memories with different binary labelings: A multi-user perspective," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 955–959.

[16] X. Huang, A. Kavcic, X. Ma, G. Dong, and T. Zhang, "Optimization of achievable information rates and number of levels in multilevel flash memories," in *Proc. 12th Int. Conf. Netw. (ICN)*, Seville, Spain, Jan. 2013, pp. 125–131.
[17] J. Huber and U. Wachsmann, "Capacities of equivalent channels in multilevel coding schemes," *Electron. Lett.*, vol. 30, no. 7, pp. 557–558, Mar. 1994.
[18] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 23, no. 3, pp. 371–377, May 1977.
[19] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 1. Jun. 1995, pp. 331–335.
[20] Y. Kofman, E. Zehavi, and S. Shamai (Shitz), "Performance analysis of a multilevel coded modulation system," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 299–312, Feb. 1994.
[21] Q. Li, A. Jiang, and E. F. Haratsch, "Noise modeling and capacity analysis for NAND flash memories," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun./Jul. 2014, pp. 2262–2266.
[22] J. Moon, J. No, S. Lee, S. Kim, and J. Yang, "Statistical analysis of flash memory read data," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Houston, TX, USA, Dec. 2011, pp. 1–6.
[23] T. Parnell, N. Papandreou, T. Mittelholzer, and H. Pozidis, "Modelling of the threshold voltage distributions of sub-20 nm NAND flash memory," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Austin, TX, USA, Dec. 2014, pp. 2351–2356.
[24] A. Paszkiewicz and T. Sobieszek. (2012). "Additive entropies of partitions." [Online]. Available: https://arxiv.org/abs/1202.4591
[25] B. Peleato, R. Agarwal, J. M. Cioffi, M. Qin, and P. H. Siegel, "Adaptive read thresholds for NAND flash," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3069–3081, Sep. 2015.
[26] W. J. Reed, "The normal-Laplace distribution and its relatives," in *Advances in Distribution Theory, Order Statistics, and Inference*. Boston, MA, USA: Birkhäuser, 2006, pp. 61–74.
[27] C. Schoeny, F. Sala, and L. Dolecek, "Analysis and coding schemes for the flash normal-Laplace mixture channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 2101–2105.
[28] V. Taranalli, H. Uchikawa, and P. H. Siegel, "Error analysis and inter-cell interference mitigation in multi-level cell flash memories," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 271–276.
[29] S. ten Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," *AEU Int. J. Electron. Commun.*, vol. 54, no. 6, pp. 389–398, 2000.
[30] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping and decoding for multilevel modulation," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 1. Nov. 1998, pp. 579–584.
[31] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge Univ. Press, 2005.
[32] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.
[33] U. Wachsmann and J. Huber, "Conditions on the optimality of multilevel codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 1997, p. 266.
[34] J. Wang, T. Courtade, H. Shankar, and R. D. Wesel, "Soft information for LDPC decoding in flash: Mutual-information optimized quantization," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Houston, TX, USA, Dec. 2011, pp. 1–6.
[35] J. Wang *et al.*, "Enhanced precision through multiple reads for LDPC decoding in flash memories," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 880–891, May 2014.
[36] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1713–1727, Nov. 1994.



Pengfei Huang received the B.E. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2010, and the M.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2013. He is currently pursuing the Ph.D. degree with the Center for Memory and Recording Research, Department of Electrical and Computer Engineering, University of California at San Diego. His current research interests are coding for distributed storage systems and non-volatile memories.

Paul H. Siegel (M'82–SM'90–F'97) received the S.B. and Ph.D. degrees in mathematics from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1975 and 1979, respectively. He was a Chaim Weizmann Post-Doctoral Fellow with the Courant Institute, New York University, New York, NY, USA. He was with the IBM Research Division, San Jose, CA, USA, from 1980 to 1995. He joined the Faculty of the University of California at San Diego, CA, USA, in 1995, where he is currently a Distinguished Professor of Electrical and Computer Engineering with the Jacobs School of Engineering. He is currently with the Center for Memory and Recording Research, where he is an Endowed Chair and served as the Director from 2000 to 2011. His research interests include information theory and communications, particularly coding and modulation techniques, with applications to digital data storage and transmission. He was a member of the Board of Governors of the IEEE Information Theory Society from 1991 to 1996, and from 2009 to 2014. He served as Co-Guest Editor of the 1991 Special Issue on Coding for Storage Devices of the IEEE TRANSACTIONS ON INFORMATION THEORY. He served as an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY for Coding Techniques from 1992 to 1995, and an Editor-in-Chief from 2001 to 2004. He was a Co-Guest Editor of the 2001 two-part issue on The Turbo Principle: From Theory to Practice and the 2016 issue on Recent Advances in Capacity Approaching Codes of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He is a member of the National Academy of Engineering. He was the 2015 Padovani Lecturer of the IEEE Information Theory Society. He was a co-recipient of the 2007 Best Paper Award in Signal Processing and Coding for Data Storage from the Data Storage Technical Committee of the IEEE Communications Society. He was a co-recipient of the 1992 IEEE Information Theory Society Paper Award and the 1993 IEEE Communications Society Leonard G. Abraham Prize Paper Award.

Eitan Yaakobi (S'07–M'12) received the B.A. degree in computer science and mathematics, 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 at San Diego in 2011. He is currently an Assistant Professor with the Computer Science Department, Technion Israel Institute of Technology. From 2011 to 2013, he was a Post-Doctoral Researcher with the Department of Electrical Engineering, California Institute of Technology. His research interests include information and coding theory with applications to non-volatile memories, associative memories, data storage and retrieval, and voting theory. He received the Marconi Society Young Scholar Award in 2009, and the Intel Ph.D. Fellowship in 2010 and 2011.