Kufa Journal of Engineering
Vol. 16, No. 3, July 2025, P.P. 607-623
Article history: Received 11 November 2024, last revised 24 December 2024, accepted 7 January 2025

CHANNEL CAPACITY FORMULATIONS FOR NON-VOLATILE MEMORIES WITH UNREACHABLE LEVELS

Haider Abdul Hassan Hadi Al Kim[1,2]

[1] Electronic and Communications Engineering, University of Kufa (UoK), Iraq, Email:haider.alkim@uokufa.edu.iq.

[2] Former Researcher at Institute for Communications Engineering, Technical University of Munich (TUM), Germany, Emai:haider.alkim@tum.de.

https://doi.org/10.30572/2018/KJE/160334

ABSTRACT

This paper investigates the channel capacity of unreachable memory cells (UMCs), where a cell is deemed $\tilde{s}$-unreachable if it cannot store values beyond a specific state, $\tilde{s}$. Memories with these impairments are modeled as discrete memoryless channels (DMCs), similar to those used in information theory and communications. We derive Shannon-type capacity equations for memories with unreachable levels and substitution errors. These novel equations generalize classical Shannon capacity to systems with UMCs. We also compare ideal memories (without imperfections or errors) with normal memories affected by random errors only, as well as defective and erroneous memories. Our findings corroborate previous studies, particularly regarding random distributions of defective cells. Our results highlight the impact of increasing faulty cells and substitution errors, demonstrating the necessity of greater redundancy to maintain system performance.

KEYWORDS

Non-volatile Memories; Defective Memory; Unreliable Cells; Error Correcting Codes; Shannon Capacity; Capacity-achieving Codes; Entropy.

1. INTRODUCTION

Point-to-point communication systems can operate in space, between a transmitter and receiver, or in time, where symbols are stored and later retrieved. These systems depend on accurate channel estimation and understanding of channel capacity to minimize errors and optimize performance. Many types of research highlights various approaches to improving signal reliability and data rates, and reducing error probabilities, focusing on the critical roles of SNR, antenna configurations, and advanced estimation techniques (Hameed 2021; Ali and Hreshee 2023; Al-Ja'afari et al. 2021).

Channel coding plays a crucial role in determining the maximum data rate that a channel can transfer. The foundational ideas of information theory were introduced by Claude E. Shannon in his groundbreaking 1948 work (Shannon, 1948). The capacity theorem, which is central to understanding these concepts, is revisited in (Brémaud, 2017) and serves as a primary reference for general aspects of information theory.

The necessity for reliable storage devices, namely non-volatile memories (NVMs) like flash memory and phase-change memory (PCM), is rising across various applications. NVMs are a type of memory that can keep data even when power is unplugged. These multi-level storage devices provide permanent data retention, quick access speeds, low power usage, improved durability, and scalable storage capacities, making them an essential technology in modern systems (Dolecek and Sala 2016). However, the gradual deterioration of the read channel, which causes a decline in reliability over time, is a major obstacle to NVM technology. This deterioration largely arises from repeated programming and erasing cycles in flash memory, causing issues such as charge trapping within oxide and interface states (Olivo, Ricco, and Sangiorgi 1986; Compagnoni et al. 2009). These trapped charges result in what are referred to be stuck-at or unreachable levels (see (Wachter-Zeh and Yaakobi 2016; Heegard 1983)), which can fully or partially prevent a cell from changing states even if new charges are applied, thus limiting the cell's functionality. In order to operate, PCM cells alternate between two states: amorphous and crystalline. These cells can become unreliable (or inaccessible, as described in (Gleixner, Pellizzer, and Bez 2009; Pirovano et al. 2004; Gabrys, Sala, and Dolecek 2014)) if they fail to transition states during the heating and cooling processes, leaving them fixed in a one phase. In multi-level PCM cells, defects may arise either at extreme states or within sub-states of the crystalline layer. Each multi-level cell functions as a symbol across a discrete alphabet of size q, representing one of the q possible levels. Unreachable memory cell restriction is described in (Gabrys, Sala, and Dolecek 2014) (see also Definition 1) and is seen as the dual of partially stuck memory cells (PSMC), which is defined in (Wachter-Zeh and

Yaakobi 2016; Al Kim, Puchinger, et al. 2023). Fig. 1 illustrates writing and reading in MLC memory (cf. Section 1) with two such unreliable cells, where only the lower levels are adjustable with two restricted cells. Cell0 and Cell2 can only be programmed at the first and second levels, with their most significant bits (MSBs) fixed as "0" and unable to change to "1" during encoding. In contrast, Cell1 is fully functional, allowing all four states through adjustments its MSB and LSB (least significant bit), illustrated by four colors in ascending order: blue, green, orange, and red. The decoder interprets each level index (an integer in 0, 1, 2, 3) as a pair of bits, (b1, b2), constructing symbols from the binary alphabet $\{0, 1\}^2$. Symbol0 corresponds to bits (b1, 0) from Cell0, Symbol1 to Cell1, and so on. These symbols concatenate into a codeword of length n.

Non-volatile memory (NVM) channels differ from traditional communication channels by some physical constraints, such as unreachable levels, and unique error patterns. Moreover, NVM errors are often asymmetric and magnitude-limited, unlike symmetric noise in conventional channels (Dolecek and Sala 2016). As an instant, multi-level cell (MLC) flash memory shows frequent asymmetric errors and technologies like phase-change RAM (PCRAM) face challenges such as limited write endurance and data retention. These factors necessitate the analysis of channel capacity and the development of solutions for NVM-specific defects and errors. Furthermore, emerging memory technologies like STT-MRAM, RRAM, and PCRAM encounter impediments such as limited write endurance, small read windows, and data retention issues under extreme conditions (Chih et al., 2021).

The performance degradation in NVM devices can be relieved by employing system-level channel codes, as these memories are analogous to point-to-point communication systems. The concept of using error correction techniques to improve reliability in such memories dates back to the 1970s (Kuznetsov and Tsybakov 1974).

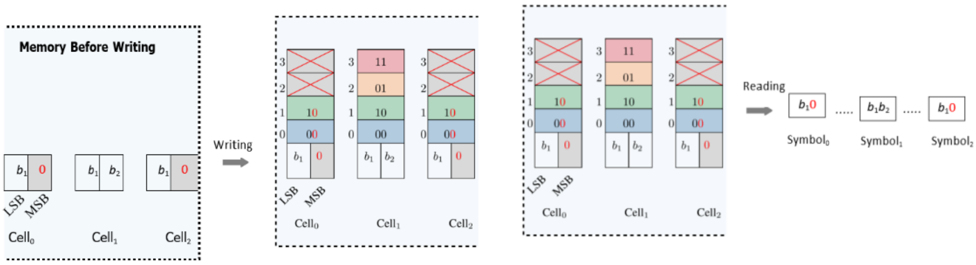
The article is organized as follows: Section 2 provides a review of related work on non-volatile memories (NVMs) with defects and errors. Section 3 clearly sates our contribution and lists our Theorems. Section 4 gives our notations. Sections 5 and 6 are to recall Shannon capacity and other related work that directly comparable to our derived channel capacity expressions in Sections 7. We discuss our simulation results in Section 8. Finally, Section 5 concludes the paper and offers suggestions for future work.

2. RELATED WORK

The first attempt to model a dual problem, i.e., stuck-at defects, mathematically and estimate some bounds on the *capacity* can be found in (Heegard and Gamal 1983). The authors in

(Heegard and Gamal 1983) studied an extreme scenario in which binary fully defective memories were considered.

In combination with a coset coding and another error-correcting code, Mahdavifar and Vardy (Mahdavifar and Vardy 2015) show how the state-of-the-art capacity-achieving codes can be used in order to *asymptotically* achieve the capacity of the *binary defective memory* (Heegard 1983). Recent work in (Wachter-Zeh and Yaakobi 2016, Section VIII) provides code constructions, without correcting substitution errors, for non-writable memories at levels $\tilde{s}$ (cf. Section 4.1). The authors (Wachter-Zeh and Yaakobi 2016, Section IX) formulate a capacity expression *without* considering any additional errors in a channel if *both* encoder and decoder know the state of the memory, i.e., the positions and states of the partially stuck-at cells.



a) Writing process by the encoder for n cells whose levels are $\{0, 1, 2, 3\}$.

b) Reading process by the decoder for n cells whose levels are $\{0, 1, 2, 3\}$.

Fig. 1. Multi-Level Cell (MLC) Device (cf. Section 1)) with Two Restricted Cells and One Normal Cell. A similar figure is in (Al Kim and Sidorenko 2023).

However, due to substitution errors caused by inter-cell interference noise or other disturbances, the *writing process* may be unsuccessful (Dolecek and Sala 2016). Additionally, the *reading process* may also fail possibly due to errors in data interpretation or signal degradation (Solomon and Y. Cassuto 2019). Thus, works demanding error correction codes and capacity consideration to tolerate both types of issues, i.e., faulty memory that suffer random errors are inevitable. The recent work in (Al Kim, Puchinger, et al. 2023) introduces diverse algorithms that are capable of overcoming the combination of substitution errors and partially stuck memory cells (PSMC). The authors examine the code rates while both problems are studied. Gabrys et al. in (Gabrys, Sala, and Dolecek 2014) suggest a coding scheme that programs the unreliable cells only at specific voltage levels, which are lower than the standard threshold, thereby reducing the likelihood of error occurrence. Their approach utilizes generalized tensor product codes to enhance reliability. Al Kim and Sidorenko (Al Kim and Sidorenko 2023) in their recent work have also suggested codes and bounds on memory with inaccessible levels. Nonetheless, they (Al Kim, Puchinger, et al. 2023; Mahdavifar and Vardy 2015; Gabrys, Sala, and Dolecek 2014; Al Kim and Sidorenko 2023) did not involve the capacity formulations regarding channels of this kind.

Ben-Hur and Cassuto paper (Ben-Hur and Yuval Cassuto 2021) addresses a similar problem with a *single* erroneous state, namely a *barrier state*. Their work as shown in (Ben-Hur and Yuval Cassuto 2021, Fig.2) explores the capacity of the *barrier channel*. However, the authors focus on a dominant error that is happening in the 0 state, i.e., a higher probability when transitioning *into* and *out* of this state.

We count on these papers as our foundation to derive our capacity formulas considering q-ary memories with unreachable levels. Our work also *generalize* the capacity expression of the barrier channel in (Ben-Hur and Yuval Cassuto 2021) since we suggest q-ary system in which any chosen state among q many could be the barrier state (cf. for example Fig. 4). We imply by our simulation results that these capacity expressions satisfy prior works that have been mentioned above. For instance, the outcomes in (Al Kim and Sidorenko 2023) align with our findings in this paper.

3. CONTRIBUTION

This paper presents Shannon-type capacity equations as reference points, enabling future and past work on channels with multiple types of errors to be directly related to our formulas. We also draw them to easily visualize the channel model of these capacities. The main derived capacity expressions that are our contribution in this work are as follows.

- Theorem 1: capacity of an unreachable memory system at *any* $\tilde{s}$ with disjoint errors,
- Theorem 2: capacity with *joint* errors in n cells with $\tilde{s}= q-2$,
- Theorem 3: generalization of Theorem 2 for *arbitrary* $\tilde{s} \in \{0,1,\ldots,q-1\}$, and
- Theorem 4: achievable capacity expressions in Theorem 1, Theorem 2, and Theorem 3.

Publications such as (Mahdavifar and Vardy 2015; Wachter-Zeh and Yaakobi 2016; Al Kim, Puchinger, et al. 2023; Al Kim and Sidorenko 2023; Gabrys, Sala, and Dolecek 2014), among others, can apply our formulations to compare their code rates with our capacity formulations We simulate our derived formulas to compare capacity across different channel regimes, as the following:

- an idealized memory with full capacity,
- a q-ary symmetric channel (qSC) modeling a discrete memoryless channel (DMC) with substitution errors only,
- defective memory without random errors, and
- a q-ary defective channel, denoted by qDC(p) (cf. Definition 3), representing defective memory.

Our results are consistent with prior code constructions' rates, some of which reach our capacity limits. While we have not included these rate results here, readers can explore them further in (Kuznetsov and Tsybakov 1974; Heegard and Gamal 1983; Heegard 1983; Gabrys, Sala, and Dolecek 2014; Solomon and Y. Cassuto 2019; Mahdavifar and Vardy 2015; Wachter-Zeh and Yaakobi 2016; Al Kim, Puchinger, et al. 2023; Al Kim and Sidorenko2023). Nevertheless, we state Theorem 4 that proves these codes achieve our capacity equations.

4. PRELIMINARIES

4.1. Notations

Denote by $\mathbb{Z}/q\mathbb{Z}$ the set of integers modulo q. For $g, f \in \mathbb{Z}_{>0}$, denote $[f\,] = \{0, 1, \ldots, f-1\}$. For a prime power q, let $\mathbb{F}_q$ denote the finite field of order q. Let $\{\tilde{\boldsymbol{s}}, \boldsymbol{s} \in \mathbb{F}_q^n : \boldsymbol{s} + \tilde{\boldsymbol{s}} = \boldsymbol{q-1}\}$. The (Hamming) weight $wt(\boldsymbol{a})$ of a vector $\boldsymbol{a} \in \mathbb{F}_q^n$ equals its number of non-zero entries. We use the notation $(q-1)^n$ and $\boldsymbol{q-1} \in \mathbb{F}_q^n$ *interchangeably* to refer to a vector of length n where each element is an integer equal to $q-1$. In addition, we represent $\boldsymbol{a} \in \mathbb{F}_q^n$ as $\boldsymbol{q-1-\tilde{s}} \in \mathbb{F}_q^n$, meaning that $\boldsymbol{q-1-\tilde{s}}$ is a vector of length n where all its non-zero components indicate the positions of a memory with non-programmable levels, and $\tilde{\boldsymbol{s}}$ is a vector of length n, i.e., $\{\tilde{\boldsymbol{s}} \in \mathbb{F}_q^n : wt(\boldsymbol{q-1-\tilde{s}}) \neq \boldsymbol{0}\}$. To coincide with works considering codes for memories with unreachable levels over finite fields, we also define a total ordering "$\leq$" on the elements of $\mathbb{F}_q$ such that $0 \leq 1 \leq x$ for all $x \in \mathbb{F}_q \setminus \{0\}$. In other words, 0 is the smallest element in $\mathbb{F}_q$, and 1 is the next smallest element in $\mathbb{F}_q$. To maintain uncomplicated notation, we sometimes represent an element $x \in \mathbb{F}_q$ by an integer value, which is calculated as $q - |\{y \in \mathbb{F}_q | \, x \leq y\}|$. For example, let $q = 2^2$ where $\mathbb{F}_{2^2} = \{0, 1, \alpha, \alpha^2\}$ and α be a primitive element in $\mathbb{F}_{2^2}$, and let $x = \alpha^2$, then the integer value that associates with α^2 is $2^2 - |\{\alpha^2\}| = 3 \in \mathbb{Z}/q\mathbb{Z}$. Our theorems in this paper are defined over the set $\mathbb{Z}/q\mathbb{Z} = \{0, 1, \ldots, q-1\}$ but can also be applied to $\mathbb{F}_q \coloneqq \{0, 1, \alpha, \ldots, \alpha^{q-2}\}$ due to the one-to-one mapping between $\mathbb{Z}/q\mathbb{Z}$ and $\mathbb{F}_q$.

Definition 1. (Unreachable Memory Cells (UMC)) In a block of n q-ary memory cells, a cell is defined as unreachable at level $\tilde{s}$ if it can store only values up to $\tilde{s}$. A non-writable cell at level $q-1$ is just a normal cell as it is capable of being programmed until its $(q-1)$-th level. In contrast, an unreachable cell at level 0 is said fully non-programmable, meaning it can only store the 0 symbol.

5. STANDARD CHANNEL CAPACITY FOR DISCRETE MEMORYLESS CHANNEL (DMC)

We start by recalling the well-known channel capacity derived by Shannon (Shannon 1948).

5.1. Channel capacity

The channel coding theorem introduced by Shannon (Shannon 1948) concerns the possibility of communicating via a noisy channel with an arbitrarily small error. The channel capacity C is defined as the maximum mutual information between two discrete random variables X and Y, which can be achieved by an input distribution $P_X(\cdot)$.

$$C = \max_{P_X} I(X;\, Y\,) \tag{1}$$

The mutual information can be described by the Entropy of random variables as follows:

$$I(X;\, Y\,) = H(X) - H(X|Y\,) \;= H(XY\,) - H(X|Y\,) - H(Y\,|X) \tag{2}$$

Entropy is the logarithmic weighted uncertainty in a probability distribution, defined as:

$$H(X) = -\sum_{a\in\, supp(P_X)} P_X(a) \log_2 P_X(a) \tag{3}$$

5.2. Entropy functions and memory capacity

Definition 2. (q-ary Entropy Function) The q-ary entropy function, $h_q(\varepsilon)$, measures the uncertainty of an event occurring with probability ε in a system with an alphabet size of q. It is defined as:

$$h_q(\varepsilon) = \varepsilon\, \log_q(q-1) \,-\varepsilon\, \log_q \varepsilon - (1\,-\,\varepsilon)\, \log_q(1\,-\,\varepsilon)$$

This function generalizes the binary entropy function and is used in systems operating with q-ary symbols. For q = 2, it reduces to the binary entropy function

$$h_2(\varepsilon) = -\varepsilon \log_2 \varepsilon - (1\,-\,\varepsilon) \log_2(1\,-\,\varepsilon). \tag{4}$$

Note that to relate Eq.4 and Eq.3, we set $P_X(0) = \varepsilon$ and $P_X(1) = 1 - \varepsilon$. Then $H(X) \;=\; h_2(\varepsilon)$.

Definition 3. (q-ary Erasure Channel and q-ary Defective Channel) A q-ary erasure channel (qEC) is a communication channel model where transmitted symbols can be erased with some probability, leaving the decoder uncertain about the symbol received, which can take any of the q possible values.

Conversely, a q-ary defective channel denoted by (qDC(p)), is a channel model where stored symbols may fail to be programmed with some probability p, leaving the encoder unable to store certain symbols.

5.3. Channel models

In information theory, several channel models exist, with discrete memory-less channels. One of the most important models is the binary symmetric channel (BSC), which has a binary input alphabet X and a binary output alphabet Y.

The general formula for the capacity of this channel can be obtained by evaluating Eq.1 as given in (Brémaud 2017):

$$C = \max_{P_X} I(X;\, Y\,) = 1 - h_2(\varepsilon)\,, \tag{5}$$

where $h_2(\varepsilon)$ is the binary entropy function, representing the uncertainty of a BSC with crossover probability ε.

Thus, we one can generalize Eq.5 into q-ary entropy that is expressed by:

$$C = \max_{P_X} I(X;\, Y\,) = 1 - h_q(\varepsilon)\,. \tag{6}$$

6. CHANNEL CAPACITY FOR PARTIALLY STUCK-AT CELLS

According to (Wachter-Zeh and Yaakobi 2016, Section IX), if both the encoder and decoder are aware of the memory state, including the locations and levels of the partially stuck-at cells, the capacity $C_q(p,s)$ (in the authors' notations $s = s_{\phi} i\ \forall i$) without considering any additional errors in a channel is

$$C_q(p,s) = 1 - p + p\, \log_q(q\, -\, s)\,. \tag{7}$$

It is assumed that a cell is partially stuck at level s with some probability p, where $u = pn$. Accordingly, there are $n - u$ normal cells, i.e. non-defect memory cells, with probability $(1-p)$ and thus $n - u = (1-p)n$.

It is worth noting that referring to Eq.7, the finite field size q, critically impacts the NVM channel capacity $C_q(p,s)$. It represents the number of distinct levels in memory and influences $\log_q(q - s)$ term in Eq.7, which quantifies usable levels amidst s unreachable ones. Larger q reduces the impact of s, increasing capacity, while smaller q amplifies effect of s, and thus reducing capacity.

Remark 1. If error does not be considered as in (Wachter-Zeh and Yaakobi 2016, Section IX) and we assume a memory without any probability of partially stuck at some levels $(p = 0, s = 0)$. Then, Eq. 7 becomes $(C_q(0,0) = 1)$ which is an optimal storage memory of a full capacity to store any vector of information of length n.

Remark 2. If error does not be considered as in (Wachter-Zeh and Yaakobi 2016, Section IX) such that $\varepsilon = 0$ and we assume a memory with a partially stuck at $s = q-1$ (a cell can not store anything). Then, Eq. 7 becomes $C_q(p, q-1) = 1 - p$ which agrees with (Heegard and Gamal 1983, Eq. (1.2)).

7. CHANNEL CAPACITY FOR MEMORY WITH UNREACHABLE LEVELS AND SUBSTITUTION ERRORS

We assume that the encoder is aware of the locations and values of defective cells through techniques such as "read voltage threshold" checks, while the decoder is not. This setup mirrors a realistic scenario in flash devices, where these checks help prevent the encoder from writing to worn-out cells. Conversely, flash memories rely on the decoder's Error Correction Codes (ECC) capabilities to maintain data integrity and manage cell wear. In this context, we derive capacity expressions for various scenarios, finding that capacity reaches its maximum (cf. (Heegard and Gamal 1983, Eq. (1.2))) when both the encoder and decoder are informed about the locations of the faulty cells.

We do not specify the source of substitution errors or the reasons a cell becomes unreachable, as these errors can occur anywhere within the system of n cells following the uniform distribution.

Definition 4. (Entropy in Memory Cells) Assume a system of a memory with n cells that can be programmable at $\tilde{\boldsymbol{s}} = (\tilde{s}_0, \tilde{s}_1, \ldots, \tilde{s}_n - 1) \in \{0, 1, \ldots, q - 1\}^n$, where the Hamming weight of $(q - 1)^n - \tilde{\boldsymbol{s}}$ is at most u, i.e., $wt((q - 1)^n - \tilde{\boldsymbol{s}}) \leq u$. Consider $n - u$ are normal cells, and u cells are in a state of having inaccessible levels $\tilde{\boldsymbol{s}}$, meaning they cannot reach certain states. Errors occurring in the $n - u$ normal cells introduce uncertainty, measured by the q-ary entropy function $h_q(\varepsilon)$ (cf. Definition 2).

7.1. The capacity while disjoint errors happening in $n - u$ cells

Theorem 1 (Capacity of an Unreachable Memory System at any $\tilde{s}$ with Disjoint Errors). Let Definition 4 hold. The capacity of this memory system, disturbed by random errors in the $n - u$ normal cells, is:

$$C_q(p, \varepsilon, \tilde{s}) = 1 - p + p \log_q(\tilde{s} + 1) - (1 - p)\, h_q(\varepsilon) \qquad (8)$$

where p is the proportion of cells without errors, $\tilde{s}$ is the number of inaccessible levels in the u cells, and $h_q(\varepsilon)$ is the q-ary entropy function from Definition 2.

Proof. In this memory system, u cells have inaccessible levels with probability p, meaning they are assumed to not subjecting to further substitution errors. The $(n - u)$ remaining cells are normal and may experience random errors with crossover probability ε.

The uncertainty (or entropy) in the $n - u$ normal cells is measured by the q-ary entropy function $h_q(\varepsilon)$. The contribution of this entropy to the overall system capacity is weighted by $(1 - p)$, as the errors occur only in the $n - u$ normal cells.

On the other hand, the u cells with inaccessible levels $\tilde{s}$ contribute a capacity of $p \log_q(\tilde{s} + 1)$, i.e., only one level is not programmable. Remember that $s + \tilde{s} = q - 1$, where s (cf. Section 6).

The total capacity is the sum of the contributions from both u and $n - u$ cells, leading to the formula:

$C_q(p, \varepsilon, \tilde{s}) = 1 - p + p\log_q(\tilde{s} + 1) - (1 - p)\, h_q(\varepsilon)$.

Thus, the capacity is reduced due to the entropy arising from the errors in the n − u nor- mal cells, while the u cells contribute a reduced capacity due to their restricted levels. This concludes the proof.

Fig. 2 depicts the channel model for the capacity as stated in Theorem 1. The upper part of the figure shows the q-ary symmetric channel (qSC) with a crossover probability of ε representing the possibility that a normal cell experiences a substitution error.

During the writing (or reading) process, a value is stored (or retrieved) in $\mathbb{Z}/q\mathbb{Z}$ or $\mathbb{F}$q, differing from the intended value with probability $\frac{\varepsilon}{q-1}$. The lower part, on the other hand, illustrates the unreliable cells, which are assumed to not experience any random errors during the reading and writing operations. We call the channel in the lower part with qDC(p) a q-ary defective channel, (cf. Definition 3).

Corollary 1 (Capacity in Theorem 1 fulfilling Eq.7 and Eq.8). If $\tilde{s} = q - 1$, i.e., all cells are fully reachable (cf. Definition 1) and thus $p = 0$, then Eq.8 coincides with:

$C_q(0, \varepsilon, q - 1) = 1 - h_q(\varepsilon)$.

If $\varepsilon = 0$, i.e., substitution errors are not considered, the Eq.8 becomes

$$C_q(p, 0, \tilde{s}) \;=\; 1 - p + p\log_q(q - s).$$

Proof. When $\tilde{s} = q - 1$, all cells are fully reachable, meaning there are no inaccessible levels. Thus, $p = 0$, eliminating the p terms in Eq.8. The capacity formula simplifies to:

$$C_q(0, \varepsilon, q - 1) = 1 - h_q(\varepsilon),$$

which matches the general formula for the capacity, Eq.6. Furthermore, if the system has no substitution errors, i.e., $h_q(0) \;=\; 0$, Eq. 8 reduces to Eq. 7.

7.2. The capacity while joint errors happening in n cells

Theorem 2 (Capacity with Joint Errors in n Cells and $\tilde{s}$= $q - 2$). Let Definition 4 hold, but now assume that errors can occur in any of the n cells and $(\tilde{s}_0, \tilde{s}_1, \ldots, \tilde{s}_n - 1) \in \{q - 2, q - 2, \ldots, q - 2\}^n$. The channel capacity considering joint errors in any of the n cells is given by:

$$C_q\big(p, \varepsilon, \varepsilon_{q-2}, q - 2\big) = 1 - p + p\log_q\,(\tilde{s} + 1) - h_q(\varepsilon)(1 - p) - p\log_q(\tilde{s} + 1)h_q(\varepsilon_{q-2}), \qquad (9)$$

where $h_q(\varepsilon)$ and $h_q(\varepsilon_{q-2})$ are the q-ary entropy functions for $n - u$ normal cells and u-UMCs, respectively.

Proof. In the joint case, random errors due to, e.g., a failure in a writing process can occur in both the $n - u$ normal cells and the u cells that are inaccessible at $\tilde{s}$= $q - 1$. The entropy in

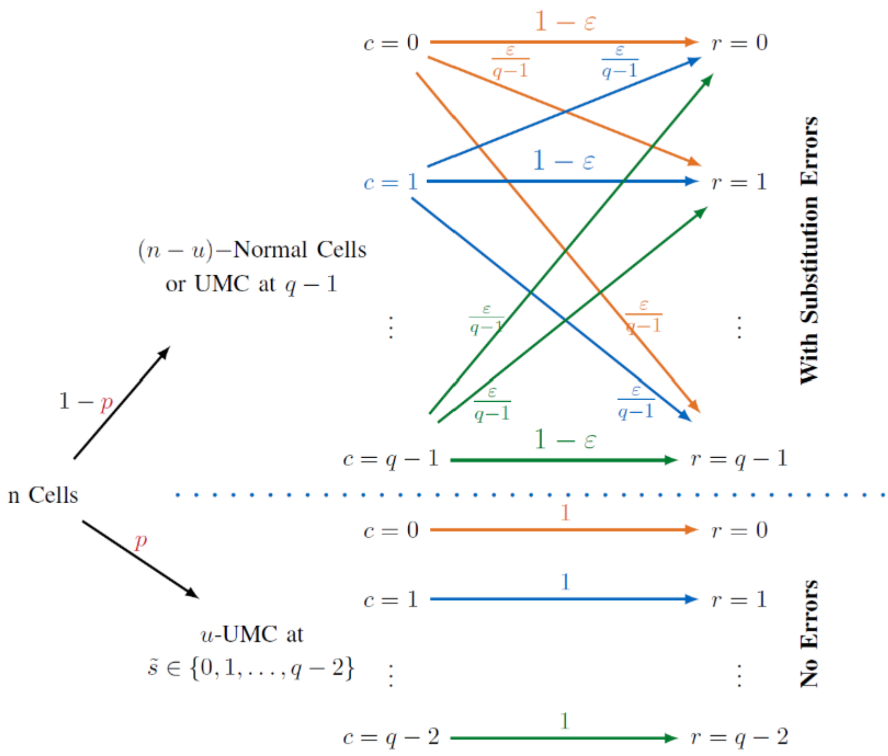


Fig. 2. Channel model for the capacity in a disjoint case. The state $\tilde{s} = 0$ means that a cell is wholly unreachable and thus only the value 0 can be stored there.

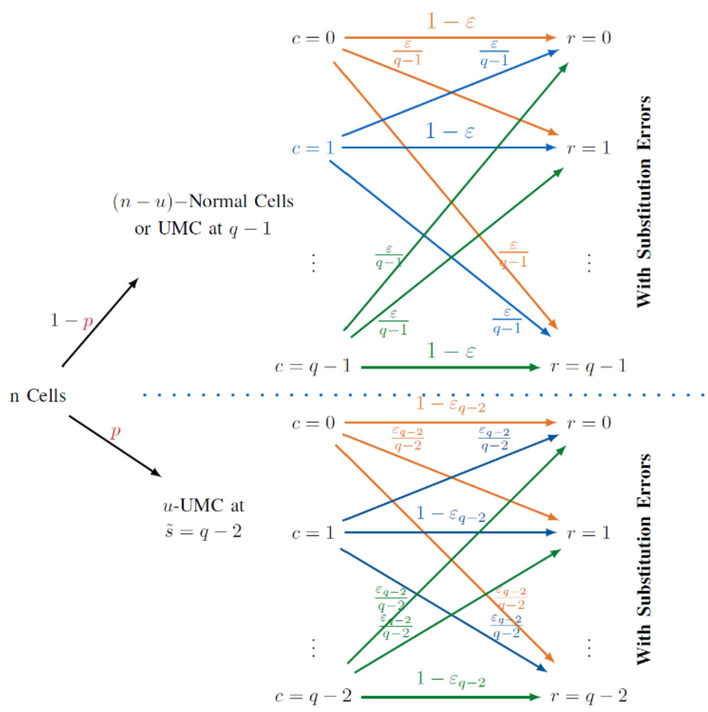


Fig. 3. Channel model for the capacity in a joint case. The state $\tilde{s}$=$q-2$ means that a cell is not programmable at only the state $q-1$ thus the latter value cannot be stored in that memory cell.

the normal cells follows the standard probability ε, contributing the term $h_q(\varepsilon)(1-p)$. In the u-UMCs, errors occur in the range between the lowest level 0 and $\tilde{s}$= $q-2$, with crossover probability of error denoted by ε_{q-2}. The uncertainty in these cells is reflected by the entropy term $p\log_q(\tilde{s}+1)h_q(\varepsilon_{q-2})$, which reduces the overall capacity. Combining these two contributions, the total capacity is given by Eq.9, which accounts for the entropy in both normal and defective cells.

One can visualize in Fig. 3 the capacity of a channel model as proved in Theorem 2.

Corollary 2. If $\varepsilon = 0$ and $\varepsilon_{q-2} = 0$, the capacity $C_q(p, 0, 0, q-2)$ in Eq.9 coincides with the capacity $C_q(p, s)$ as given in Eq.7. Furthermore, if $p = 0$, Eq.9 reduces to the q-ary channel capacity:

$$C_q(\varepsilon) = 1 - h_q(\varepsilon),$$

which matches the general formula for the capacity in Eq.6.

Proof. For $\varepsilon = 0$ and $\varepsilon_{q-2} = 0$, the channel behaves as if no errors occur, thus reducing the capacity expression to $C_q(p, s)$ as in Eq.7. When $p = 0$, there are no UMCs, leading to the classical q-ary channel capacity, $C_q(p, s) = 1 - h_q(\varepsilon)$, as shown in Eq.6.

7.3. A generalization of Theorem 2

Theorem 3 (Generalization of Theorem 2). Let $\tilde{\boldsymbol{s}} = \{0,1, \dots, q-1\}^n$ with $wt((q-1)^n - \tilde{\boldsymbol{s}}) \leq u$. A generalization of the capacity from Theorem 2 is given by:

$$C_q(p, \varepsilon, \varepsilon_0, \varepsilon_1, \dots, \varepsilon_{q-2}, \tilde{\boldsymbol{s}}) = 1 - p + \sum_{\tilde{s}=0}^{q-2} p_{\tilde{s}} \left[(\tilde{s}+1) - log_q(\tilde{s}+1)h_q(\varepsilon_{\tilde{s}})\right] - (1-p)h_q(\varepsilon), \quad (10)$$

where $p_{\tilde{s}}$ denotes the probability of a cell to become unreachable at level $\tilde{s}$ that has a crossover probability $\varepsilon_{\tilde{s}}$.

Proof. The expression in Eq.10 is a generalization of the capacity from Eq.9. This means that u cells cannot be reached at arbitrary levels $\tilde{s}$. The probability p for Unreachable Memory Cells (UMCs) splits into sub-probabilities according to the number of defective levels, $\tilde{\boldsymbol{s}} = (\tilde{s}_0, \tilde{s}_1, \dots, \tilde{s}_n - 1) \in \{0,1, \dots, q-1\}^n$. The union bound on $p_0, p_1, \dots, p_{q-2}$ represent the probabilities at levels s˜ = 0, 0,1, … , $q-2$, respectively, with $\sum p_{\tilde{s}} = p$.

Fig. 4 shows the channel model for the generalization of Theorem 2. Since substitution errors occur randomly and can affect any cell in the memory (both $n - u$ and u cells), the q-ary symmetric channel (qSC) with a crossover probability of $\frac{\varepsilon}{(q-1)}$ represents the likelihood that an $n - u$ normal cell experiences a substitution error. The lower part illustrates the u unreachable cells, which also experience random errors during the reading and writing operations, but with $\frac{\varepsilon}{(q-\tilde{s})}$, corresponding to the probability $p_{\tilde{s}}$ of being not reachable in the $\tilde{s}$ state.

Theorem 4. The capacity expressions in Theorem 1, Theorem 2, and Theorem 3 are achievable.

Proof. We begin by noting the work in (Mahdavifar and Vardy 2015), which explicitly provides Hamming and convolutional coding schemes that achieve capacity. Additionally, Heegard demonstrated in (Heegard 1983) that random partitioned linear codes reach capacity, although this required solving an implicit optimization problem within the encoder. Further, Wachter-Zeh and Yaakobi verified capacity-achieving codes for $s = 1$ and larger values of q in (Wachter-Zeh and Yaakobi 2016, Figs 2 and 3). By applying Corollary 1 and Corollary 2, our capacity expressions align with those in (Heegard and Gamal 1983; Mahdavifar and Vardy 2015; Wachter-Zeh and Yaakobi 2016). Consequently, we infer that our capacities are indeed achievable.

8. RESULTS AND DISCUSSIONS

In Fig. 5, 6, and 7, we plot the channel capacity across different configurations, examining various values of q, p, ε, and u with $\tilde{s} = q - 1$. The plots illustrate capacities for multiple scenarios: "No UMCs" and "No errors" ($C_q(0,0) = 1$) representing ideal memory, "Only errors" as standard Shannon capacity for Discrete Memoryless Channels (DMCs) from Eq.5, "Only UMCs" without substitution errors from Eq.7, and "UMCs and errors" under non-overlapping and overlapping conditions, as defined by Eq.8 and Eq.9, respectively. These figures reveal that an increase in UMCs results in greater capacity degradation, as seen in the long-dashed curve. Additionally, the most severe case occurs when errors overlap with UMC locations, as depicted by the dashed-dotted line, which is closest to real-world conditions. For instance, under error-free conditions, our results match the expressions in (Wachter-Zeh and Yaakobi 2016, Section IX), and for $q = 2$ and erroneous system, our formulas coincide with those in (Heegard and Gamal 1983).Observing the sequence of these figures, we note that when ε is low, the capacity from Eq.5 approaches optimal levels. However, as ε increases, the probability of errors rises, leading to significant capacity deterioration. This is because, as the number of u cells and substitution errors in the system increases, more redundancy is required to compensate for these problems, resulting in lower capacity in the qDC(p) channels.

9. CONCLUSIONS

In this paper, we developed comprehensive capacity expressions for multi-level non-volatile memory (NVM) channels, particularly those experiencing varied error types due to degraded read channels and restricted programmability. We generalized previous channel capacity formulations to accommodate cases with both disjoint and jointly occurring cells with unreachable levels and substitution errors, addressing the practical needs of modern memory storage technologies like flash memory and phase-change memory (PCM). Our derived formulas offer a foundational reference for evaluating and comparing the rates of various coding schemes developed for NVMs under different channel conditions. Through simulations, we demonstrated the applicability of these capacity expressions across diverse memory conditions, including ideal, substitution-error-only, and defective memory channels with and without substitution error, aligning well with established capacity benchmarks.

Furthermore, we demonstrate that these capacity expressions can be achieved with appropriate coding schemes, thus contributing a robust theoretical framework for future code designs in error-resilient memory storage.

Future research might examine the usage of erasure coding to analyze a channel capacity with three considerations: unreliable states, erasures, and random errors. This extension would yield capacity formulas for a more comprehensive channel model, enhancing applicability in real-world memory systems.

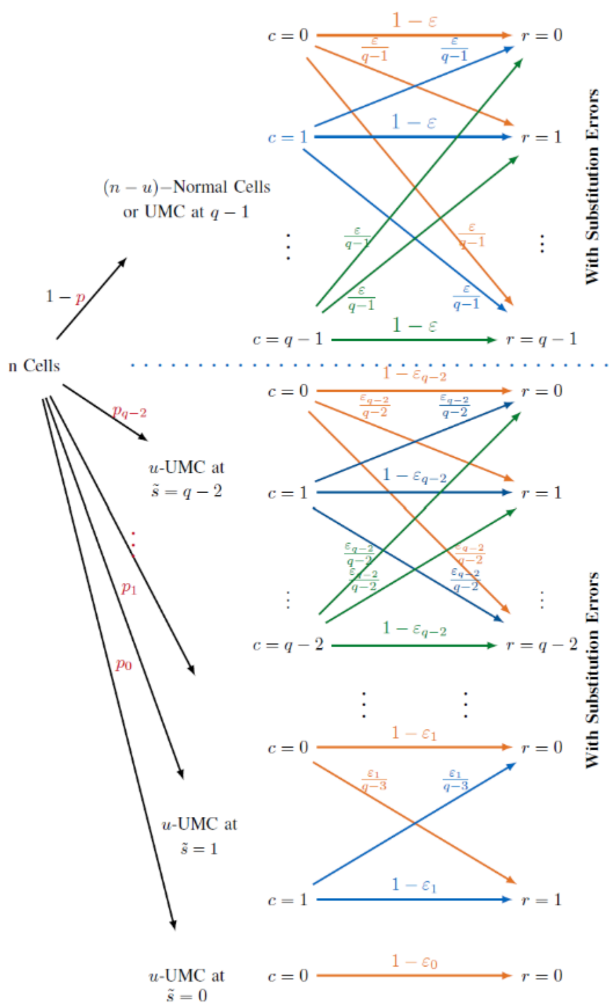


Fig. 4. A Generalization of Theorem 2. The summation of the probabilities $p_{q-2}+\cdots+p_1+p_1=p$. The crossover probabilities respective to each non-writable state $\tilde{s}\in\{q-2,\dots,1,0\}$ are denoted by $\varepsilon_{\tilde{s}}$.

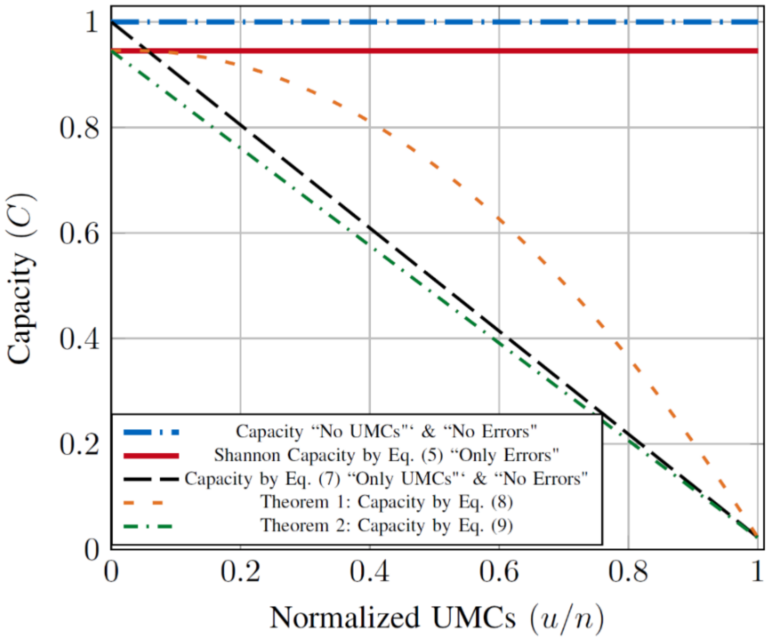


Fig. 5. The achievable capacity by Eq. 5, Eq. 7, Eq. 8, and Eq. 10. The parameters are $\varepsilon = 0.02, n = 200$ and alphabet size 2^4, where the unreachable state $\tilde{s} = 14$ is considered in $0 \leq wt(\tilde{s}) = u \leq n$ cells for the probability $0 \leq p \leq 1$.

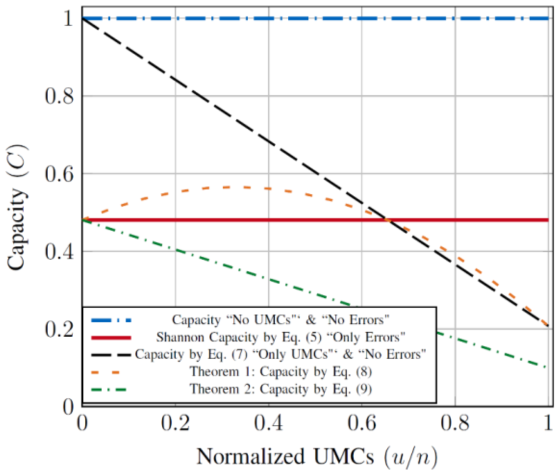


Fig. 6. Same as in Fig. 5. The changes in the parameters are $\varepsilon = 0.2, n = 15$ and alphabet size 2^2, where the unreachable state $\tilde{s} = 2$.

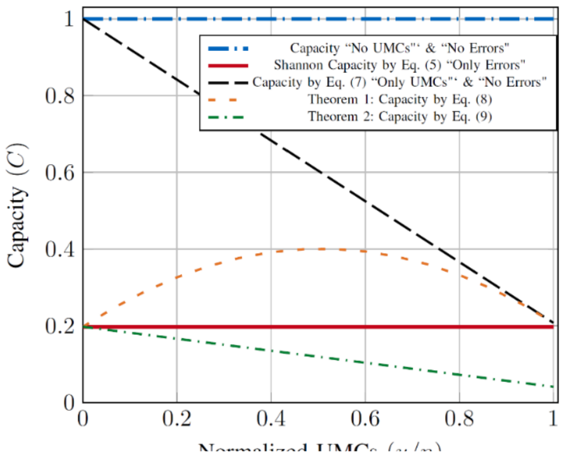


Fig. 7. Same as in Fig. 5. The changes in the parameters are $\varepsilon = 0.4$, $n = 15$ and alphabet size 2^2, where the unreachable state $\tilde{s} = 2$.

10. REFERENCES

Al Kim, H., S. Puchinger, et al. (Aug. 2023). "Coding and bounds for partially defective memory cells". In: Designs, Codes and Cryptography. DOI:10.1007/s10623-023-01270-0.

Al Kim, H. and V. Sidorenko (Oct. 2023). "Codes for Memory Cells with Unreachable Levels". In: International Symposium on Problems of Redundancy in Information and Control Systems 2023. Moscow, Russia.

Ben-Hur, Yuval and Yuval Cassuto (2021). “Coding on Dual-Parameter Barrier Channels beyond Worst-Case Correction”. In: 2021 IEEE Global Communications Conference (GLOBECOM), pp. 01–06. DOI: 10.1109/GLOBECOM46510.2021.9685188.

Brémaud, P. (2017). Discrete Probability Models and Methods. 78th. Switzerland: Springer International.

Compagnoni, C. Monzio et al. (2009). “Random Telegraph Noise Effect on the Programmed Threshold-Voltage Distribution of Flash Memories”. In: IEEE Electron Device Letters 30.9, pp. 984–986. DOI: 10.1109/LED.2009.2026658.

Dolecek, L. and F. Sala (2016). Channel Coding Methods for Non-Volatile Memories.

Gabrys, R., F. Sala, and L. Dolecek (Sept. 2014). “Coding for unreliable flash memory cells”. In: IEEE Commun. Lett. 18.9, pp. 1491–1494.

Gleixner, B., F. Pellizzer, and R. Bez (2009). “Reliability Characterization of Phase Change Memory”. In: 2009 10th Annual NVMTS. IEEE, pp. 7–11.

Heegard, C. (1983). “Partitioned Linear Block Codes for Computer Memory with ’Stuck-at’ De- fects”. In: IEEE Trans. Inf. Theory 29.6, pp. 831–842.

Heegard, C. and A. A. El Gamal (Sept. 1983). “On the Capacity of Computer Memory with De- fects”. In: IEEE Trans. Information Theory 29.5, pp. 731–739.

Hameed, Ashwaq (2021). “RAYLEIGH FADING - SHADOWING OF OUTDOOR CHANNELS ANALYSIS BASED ON SNR-PDF MODEL”. Kufa Journal of Engineering, vol. 10, no. 2, pp. 114-25, doi:10.30572/2018/KJE/100209.

Ali, Ameer H. and Saad S. Hreshee (2023). “GFDM TRANSCEIVER BASED ON ANN CHANNEL ESTIMATION”. Kufa Journal of Engineering, vol. 14, no. 1, pp. 33-49, doi:10.30572/2018/KJE/140103.

Al-Ja’afari, Mohannad, et al. (2021). “COMPARISON BETWEEN MLD AND ZF ALGORITHMS FOR MIMO WIRELESS SYSTEM AT RAYLEIGH CHANNEL”. Kufa Journal of Engineering, vol. 9, no. 4, pp. 13-22, doi:10.30572/2018/KJE/090402.

Kuznetsov, A. and B. Tsybakov (1974). “Coding for memories with defective cells”. In: Problems Inf. Transmiss. 10.2, pp. 52–60.

Mahdavifar, H. and A. Vardy (2015). “Explicit capacity achieving codes for defective memories”. In: 2015 IEEE International Symposium on Information Theory (ISIT), pp. 641–645. DOI: 10.1109/ISIT.2015.7282533.

Olivo, P., B. Ricco, and E. Sangiorgi (1986). “High-field-induced voltage-dependent oxide charge”. In: Applied Physics Letters. ISSN: 0003-6951.

Pirovano, A. et al. (2004). “Reliability Study of Phase-Change Nonvolatile Memories”. In: IEEE Trans Device Mater Reliab 4.3, pp. 422–427.

Shannon, C. E. (Oct. 1948). “A Mathematical Theory of Communication”. In: Bell System Technical Journal 27, pp. 623–656.

Solomon, A. and Y. Cassuto (Sept. 2019). “Error-correcting WOM codes: Concatenation and joint design”. In: IEEE Trans. Inf. Theory 65.9, pp. 5529–5546.

Wachter-Zeh, A. and E. Yaakobi (2016). “Codes for Partially Stuck-at Memory Cells”. In: IEEE Trans. Inf. Theory 62.2, pp. 639–654 .

Y. -D. Chih et al., “Design Challenges and Solutions of Emerging Nonvolatile Memory for Embedded Applications, ” 2021 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2021, pp. 2.4.1-2.4.4, doi: 10.1109/IEDM19574.2021.9720557.