Bounds and Constructions of Maximally Recoverable Codes for Various Topologies

Thesis submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

in

Electronics and Communication Engineering

by

Shivakrishna Dasi

201533650

d.shivakrishna@research.iiit.ac.in

International Institute of Information Technology

Hyderabad - 500 032, INDIA

December 2023

International Institute of Information Technology

Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "**Bounds and Constructions of Maximally Recoverable Codes for Various Topologies**" by Shivarkrishna Dasi, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date

Adviser: Dr. Lalitha Vadlamani

I would like to dedicate this thesis to my parents

Smt. D Radha and Sri. D Mogilaiah

My sisters

K Shivaranjani and K Sridevi

My Wife, Son and Daughter

D Harini, D Vishnu Nandan and D Srimayi

My guide

Dr. Lalitha Vadlamani

and

My Teachers, Friends, and Well wishers

Acknowledgments

First and foremost, I would like to offer my utmost praise and heartfelt thanks to the Almighty for bestowing upon me His abundant blessings.

I sincerely express my deep gratitude to my guide, Dr. Lalitha Vadlamani, for her unwavering support and patient guidance throughout my entire journey. Without her invaluable contributions, this thesis would not have been possible. I am especially grateful for the exceptional support I received from her during the most challenging phase of my life, spanning from 2019 to 2021. Her constant encouragement and constructive feedback pushed me to surpass my limits and regain focus on my research work.

I sincerely express my deep gratitude to my guide, Dr. Lalitha Vadlamani, for her unwavering support and patient guidance throughout my entire journey. Without her invaluable contributions, this thesis would not have been possible. I am especially grateful for the exceptional support I received from her during the most challenging phase of my life, spanning from 2019 to 2021. Her constant encouragement and constructive feedback pushed me to surpass my limits and regain focus on my research work.

I extend my gratitude to Dr. Prasad Krishnan for his exceptional teaching in the field of coding theory, which has provided me with a strong foundation in this area. I am also thankful to him for conducting the "Vedanta Sessions" at IIITH, from which I have greatly benefited personally. His invaluable inputs as an examiner, from the comprehensive viva to the final defense, aided significant refinements in my research, ultimately enhancing the strength and quality of my work.

I would like to express my sincere gratitude to to Prof. Adrish Banerjee of IIT Khanpur for his support from the proposal defense to the final defense. His profound insights, meticulous feedback, and scholarly expertise have greatly enriched this work, shaping it into a more robust and insightful contribution. I am truly thankful for the insightful perspectives and rigorous examination offered by Prof. Bikash Kumar Dey, which have enhanced the overall merit of my thesis.

I would like to thank my friends Prakash, Hari, Upender, Vijay, Madhuri, Tulasi, Rekha, Harshitha, Bhavana, Vamshi, Jayanth, Zakir, Srikar, Rohan, karthik and Sumanth for making my research journey memorable with their friendship and support.

My heartfelt appreciation goes to IIITH for providing me with the necessary resources, facilities, and funding to pursue my research endeavors. The support from SPCRC and the academic staff has been invaluable in creating an enriching environment for intellectual growth.

Lastly, I am indebted to my family for their unwavering love, understanding, and patience. Their constant support and belief in my abilities have been the driving force behind my academic achievements. I would like to extend special thanks to my partner for her incredible support and for standing by me during critical situations.

Abstract

In the present era of Big Data, the demand for storing vast amounts of data is rapidly increasing among companies such as Facebook, Microsoft, Google, Intel, IBM, and others, for various applications. To address this need, Distributed Storage Systems (DSSs) have been established, offering improved capabilities in terms of flexibility, scalability, speed, and cost. In DSS, data is distributed and stored on different nodes and are connected through the network. However, data loss is inevitable due to physical limitations such as hardware failures and power shutdowns. Maximum Distance Separable (MDS) codes are very efficient in terms of storage overhead. For practicality, Locally Recoverable codes (LRCs) are discovered to facilitate the low reconstruction cost for single and multiple failures (Independent and correlated), with a slight increase in storage overhead. Maximally recoverable codes are a class of codes that recover from all potentially recoverable erasure patterns given the locality constraints of the code. Our main objectives are to provide MRCs for independent failures, correlated failures with low computational complexity, and encoding complexity.

In earlier works, codes have been studied in the context of codes with the locality to handle independent failures. The notion of locality has been extended to the hierarchical locality, which allows for the locality to gradually increase in levels with the increase in the number of erasures. In one direction, we consider MRC for the case of codes with 2-level hierarchical locality for the specific topology (locality constraints) called Hierarchical Local MRC (HL-MRC). We derive a field size lower bound on HL-MRC. We also give constructions of HL-MRC for some parameters whose field size is smaller than that of earlier known constructions.

We investigate Locally Recoverable Codes (LRCs) with availability, which refers to the ability to have multiple repair sets. The presence of multiple repair sets in LRCs is beneficial

as it facilitates the distribution of the repair load among various nodes. This distribution helps prevent excessive strain on specific nodes and promotes a more balanced workload within the system. In our research, we expand on the concept of availability in Locally Recoverable Codes (LRCs) and apply it to codes with hierarchical locality. The minimum distance plays a vital role in determining the codes' capability to handle erasures effectively. Our study focuses on investigating the upper bound of the minimum distance for the specific case of LRCs with hierarchical locality.

To reduce the encoding complexity, Halbawi et al. introduced sparse and balanced generator matrices for MDS (Reed - Solomon) code and LRCs (Tamo-Barg code) for single erasure. Building upon this work, we contribute by presenting sparse generator matrices for MRC with locality for single erasure. Furthermore, we also provide sparse and balanced generator matrices for MRC with locality, specifically for single erasures where the locality value is set to 2.

In order to deal with correlated failures, Gopalan et al. initiated the study of MRCs grid-like and product topologies. In another research direction, we focus on MR codes for product topology $T_{m,n}(a, b)$. Product codes are a class of codes with generator matrices as the tensor product of the generator matrices of component codes. The codeword can be represented as an $m \times n$ array, where the component codes are referred to as the row and column codes. We derive a few properties of maximally recoverable product codes. We give a sufficient condition to characterize a certain subclass of erasure patterns as correctable and another necessary condition to characterize another subclass of erasure patterns as not correctable. We construct a certain bipartite graph based on the erasure pattern satisfying the regularity condition (a necessary condition for recoverability) for product topology and show that there exists a complete matching in this graph. We used this condition to identify a subset of recoverable erasure patterns for $a = 2$.

In earlier work, higher-order MDS codes denoted by MDS(l) have been defined in terms of generic matrices, and these codes have been shown to be constituent row codes for maximally recoverable product codes for the case of $a = 1$. We derive a certain inclusion-exclusion type principle for characterizing the dimension of intersection spaces of generic matrices. Applying this, we formally derive a relation between MDS(3) codes and points/lines of the associated projective space.

Keywords: Distributed storage, independent failures, correlated failures, maximally recoverable codes, locally recoverable codes, codes with hierarchical locality, field size bound, locally recoverable codes with availability, sparse and balanced generator matrices, product codes.

Contents

List of Figures

List of Tables

List of Notations and Abbreviations

List of Notations:

$\mathbb{N}$	–	set of natural numbers
$[n]$	–	$\{1, 2, \cdots, n\}$ where $n \in \mathbb{N}$
ϕ	–	empty set
$\mathbb{F}_q$	–	finite field with q elements
$\mathbb{F}_q^k$	–	k-dim(dimensional) vector space over $\mathbb{F}_q$
$dim(V)$	–	dimension of a vector space V
$supp(c) \subseteq [n]$	–	the set of non-zero coordinates of codeword c of the code $\mathcal{C}$
$G\|_I$	–	the restriction of the matrix G to the set of columns indexed by the set $I \subseteq [n]$
$rank(G\|_I)$	–	rank of the matrix $G\|_I$
$\| . \|$	–	cardinality of a set
$\mathcal{C}\|_I$	–	$\mathcal{C}$ is punctured to the set of co-ordinates corresponding to $I \subseteq [n]$
$dim(\mathcal{C})$	–	dimension of the code $\mathcal{C}$
MDS(.)	–	Higher order MDS code
$VS(k, q)$	–	1-dimensional subspace of k-dimensional vector space over $\mathbb{F}_q$
$PG(k-1, q)$	–	$(k-1)$-dimensional projective space over F_q

List of Abbreviations:

DSS	–	Distributed Storage System
MDS	–	Maximum Distance Separable
MRC	–	Maximally Recoverable Code
LRC	–	Locally Recoverable Codes
HL-MRC	–	Hierarchical Local MRC

SB	–	Sparse and Balanced
H-LRC	–	Hierarchical Locally Recoverable Codes
MRPC	–	Maximally Recoverable Product Codes
IEP	–	Irreducible Erasure Pattern
REP	–	Regular Erasure Pattern
IREP	–	Irreducible Regular Erasure Pattern
l-REP	–	l-partitioned REP for $T_{m,n}(a, b, 0)$

Terminology

- $\mathcal{C}[n,k,d]q$ is a linear block code over the finite field $\mathbb{F}_q$ with length n, dimension k, minimum distance d, generator matrix $G_{k\times n}$, and parity check matrix $H_{(n-k)\times n}$
- $k_{opt}^{(q)}(n,d)$ is largest possible dimension of the code
- $\mathcal{C}[n,k,d,r]_q$ LRC with locality r for single parity
- $\mathcal{C}[n,k,d,r,\delta,h]_q$ LRC with $(r,\delta\geq 2)$ locality along with h global parities
- $\mathcal{C}$ is a code with hierarchical locality characterized by the parameters $[k,r_1,r_2,h_1,h_2,\delta]$, is used to define two types of topologies: the Hierarchical Data Local Code and the Hierarchical Local Code. Here, δ represents the number of parities in the local code, and h_2 and h_1 denote the number of global parities of the middle code and the code, respectively.
- $T_{m,n}(a,b,h)$ is called grid-like topologies. In which each codeword is an array of size m $\times$ n, where a, b and h are the number of parities per column, per row, and global, respectively.
- $T_{m,n}(a,b,h=0)$ is called product topologies and also represented as $T_{m,n}(a,b)$.

Chapter 1

Introduction

In this chapter, we will explore storage systems that are well-suited to meet the demands of the present Big Data Era. Section 1.1 delves into the concept of Distributed Storage System (DSS), highlighting the significance of this system and the challenges it faces. Section 1.2 focuses on the practicality of implementing Erasure Codes in DSS. Erasure Codes are introduced as a viable solution to mitigate the risk of data loss in distributed storage environments. We explore the fundamental principles of Erasure Codes and their applicability in DSS. Preliminary codes are also introduced and discussed, which serve as the building blocks for achieving efficient data storage and recovery. In Section 1.4, we outline the specific goals and contributions of the thesis. Lastly, in Section 1.5, we provide an overview of the organization of the thesis. The structure and flow of the subsequent chapters are outlined.

1.1 Distributed Storage System (DSS) and Challenges

The demand for storing massive volumes of data is rapidly growing across various industries, with companies like Facebook, Microsoft, Google, Intel, IBM, and many others facing the need to manage and store vast amounts of information. In light of this trend, there is a pressing requirement for storage systems that can surpass the capabilities of current solutions in terms of flexibility, scalability, speed, and cost-effectiveness.

To establish such a storage system, data is distributed and stored on multiple nodes that are interconnected through a network. Despite the distribution of data, these nodes operate collectively as a unified storage system. Such a storage system is referred to as a Distributed Storage System. This arrangement allows for improved efficiency and enhanced storage capabilities. The physical model of a DSS can be visualized as depicted in Fig. 1.1.

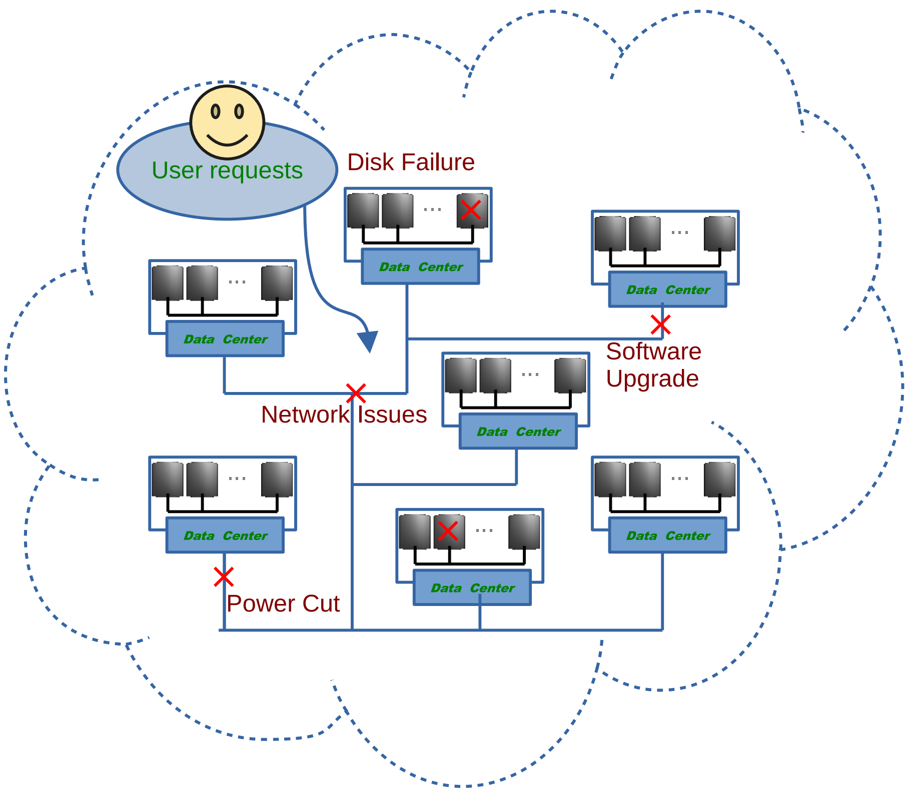

Figure 1.1: DSS Network Physical Model

However, in a DSS, data loss becomes an inevitable challenge due to the inherent limitations of hardware, such as power shutdowns and node failures resulting from device or disk malfunctions, etc. In a DSS, node failures can occur due to temporary or permanent reasons.

Temporary failures refer to situations where a node or component of the system experiences a temporary disruption or outage but can be restored to normal operation after the issue is resolved. These failures can be caused by various factors, such as network connectivity issues, power fluctuations, temporary hardware malfunctions, or software glitches. Temporary failures are generally transient and can be recovered without permanent data loss.

On the other hand, permanent failures are more severe and result in permanent data loss or the inability to recover the affected node or component. Permanent failures can occur due to catastrophic hardware failures, unrecoverable disk errors, or data center disasters. When a node in the DSS encounters a permanent failure, the data stored on that node becomes inaccessible or irretrievable. These failures are typically more challenging to handle as they require data recovery mechanisms to mitigate the loss and ensure data availability. To minimize the impact of both permanent and temporary failures, DSS integrates data protection mechanisms like erasure coding.

1.2 Erasure Codes in DSS

In DSS, the detection of failed nodes is possible. Hence, for data recovery, erasure codes are applicable. Forward Error Correction (FEC) techniques are utilized to enhance reliability. In FEC, extra redundant bits are added at the transmitter, allowing for correct data decoding even in the presence of a certain number of lost bits. Erasure codes represent a specific type of FEC where data is transmitted over an erasure channel. In this channel model, the transmitter sends the data, and the receiver either receives it correctly or experiences complete loss. For instance, the Binary Erasure Channel (BEC) is illustrated in Fig. 1.2.

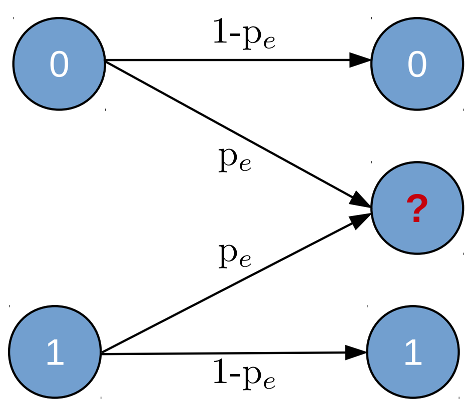

Figure 1.2: Binary Erasure Channel with error probability p_e.

Initially, to ensure reliability, data replication is commonly employed as a common and easily implemented solution. Typically, a standard approach involves three-way replication. Among the three nodes (three copies), suppose one node has failed. One of the other nodes can be used to create a new node by simply copying the data. At the same time, another node can be used to access the data, which ensures the data availability. In this three-way replication, no computations are required in data recovery. However, the three-way replication incurs a storage overhead of 200%, leading to inefficiency. Erasure coding techniques have garnered significant attention as an alternative to traditional replication methods. These techniques offer more efficient data storage and improved fault tolerance capabilities.

Erasure coding involves encoding data into a set of redundant fragments that are distributed across multiple storage nodes in the system. By leveraging mathematical algorithms, the original data can be reconstructed from a subset of these fragments, even if some of them are lost or inaccessible. This approach provides higher storage efficiency compared to replication, as it reduces the storage overhead required to ensure data reliability.

In the following, we will quickly setup some basic notation of codes and define Maximum Distance Separable (MDS) codes.

- $\mathcal{C}[n, k, d]_q$ is a linear block code with the generator matrix $G_{k\times n}$ and parity check matrix $H_{(n-k)\times n}$. Here, n, k, d and q are length, dimension, minimum distance and field size of the code. Any $c = (c_1, c_2, \ldots, c_n) \in \mathcal{C}$ is called a codeword and each c_i is a code symbol for any $i \in [n]$. $supp(c) \subseteq [n]$ denote the set of non-zero coordinates of c.

- The code can recover up to $d - 1$ erasures (recoverability). An upper bound for recoverability of the code is called *Singleton bound* given in (1.1).

$$d - 1 \leq n - k. \tag{1.1}$$

- *Maximum Distance Separable* (MDS) codes are the optimal codes that meet the Singleton bound. i.e., it can recover any $(n - k)$ erasures.

1.3 Motivation for Maximally Recoverable Codes (MRCs)

The study of erasure coding for distributed storage systems encompasses areas such as code construction, decoding algorithms, fault tolerance analysis, and system performance evaluation. It is essential to develop erasure coding techniques that strike a balance between storage efficiency, reliability, and computational complexity. MDS codes are a class of erasure codes that offer optimal data recovery capabilities in the presence of node failures with less storage overhead than replication. However, MDS codes may not be suitable for distributed storage systems due to several reasons.

Firstly, the repair process in MDS codes involves accessing all k number of surviving nodes, resulting in high network bandwidth utilization during data recovery. This can introduce performance bottlenecks and scalability challenges in distributed storage systems with a large number of nodes.

Furthermore, MDS codes are designed for scenarios where node failures are independent and occur randomly. In practice, however, failures in distributed storage systems can often be correlated, such as failures caused by power outages, hardware malfunctions, or network issues. MDS codes are not efficient when correlated failures occur.

Due to these limitations, alternative coding techniques such as Locally Recoverable Codes (LRCs) and Maximally Recoverable Codes (MRCs) have gained attention in distributed storage systems. LRCs and MRCs offer more flexible and efficient approaches to achieve fault tolerance, minimize repair bandwidth, and handle correlated failures. These coding schemes strike a better balance between storage efficiency, reliability, and computational complexity, making them more suitable for practical distributed storage environments compared to MDS codes.

Unlike traditional erasure codes, LRCs introduce a locality property that enables the reconstruction of a failed node's data by accessing a limited number of other nodes ($< k$) in the system. By minimizing the amount of data transferred during repair operations, LRCs reduce the network bandwidth usage and improve the system's overall performance. MRCs ensure that the maximum possible patterns of failures can be recovered by the code, subject the locality conditions. Both LRC and MRC play significant roles in addressing data reliability challenges, efficient repair operations, and fault tolerance in distributed storage systems.

Depending upon the type of erasures, LRCs can be categorized as shown in Fig. 1.3. To address various types of failures within the Distributed Storage System (DSS), it is possible to construct codes based on specific support constraints, which we refer to as topology in the context of these support constraints. In the subsequent chapters, we provide formal definitions of codes for different topologies, specifically within the context of the problems we address.

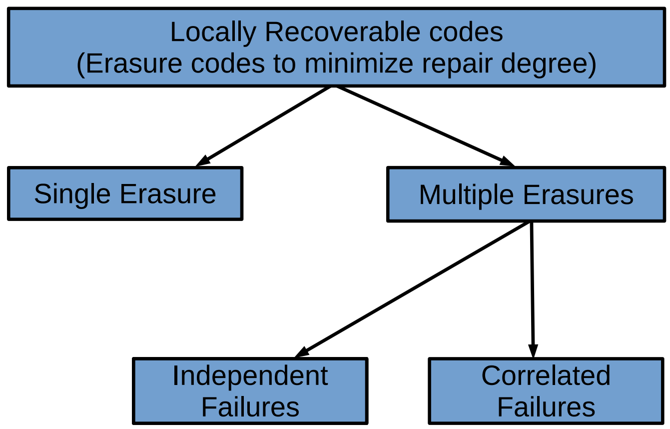

Figure 1.3: Classification of Erasures.

1.4 Objectives and Contributions of the Thesis

To enhance comprehension of the objectives and contributions of the thesis, we provide informal definitions for key concepts. The four objectives of the thesis are given as follows:

- **MRCs for independent failures:** To minimize the cost of data reconstruction, codes with (r, δ) locality are introduced, which can recover $\delta - 1$ erasures by contacting $r < k$ (locality) other nodes. This concept is further extended to codes with hierarchical locality, characterized by (r_1, δ_1) and (r_2, δ_2), where $r_2 < r_1$ and $\delta_2 < \delta_1$. This hierarchical structure provides a locality of $r_1 < k$ when erasures exceed $\delta_2 - 1$ but remain less than or equal to $\delta_1 - 1$. Consequently, one of our primary objectives is to construct maximally recoverable codes with hierarchical locality (particularly for 2-level locality). Specifically, we focus on a topology Hierarchical Local MRC (HL-MRC), which we see in detail in Chapter 3. The contributions are as follows:

  - This thesis presents the constructions of Hierarchical Local MRCs and their corresponding field sizes by utilizing algebraic properties of the field. We have presented two constructions for a few fixed parameters and two constructions for all general parameters. For a few fixed parameters, we have provided two kinds of HL-MRC constructions using the coset properties and Cauchy matrix. One is obtained with the field size of $O(n_1)$ (here, $n_1 < n$), and the other one is of $O(n^4)$. For all general parameters, one construction is based on the generator matrix of Linearized Reed-Solomon through the parity check matrix approach. It is obtained with the field size of $O((n_2)^c)$, here $n_2 < 1$ and c is some constant. Under certain parameters of the regimes, our constructions outperform the prior work; a detailed comparison is described in Subsection 3.3. Also, a random construction for all general parameters is obtained through the generator matrix approach, and its field size is $O(n^k)$, which is higher than any other construction. We have also derived lower bounds on the field size of HL-MRC for a few cases.

- **MRCs for correlated failures:** In order to address correlated failures, a specific type of topology known as grid-like topologies $T_{m,n}(a, b, h)$ is commonly employed. However, our research focus is on a subclass of these topologies referred to as product topologies $T_{m,n}(a, b, h = 0)$. Product codes belong to a class of codes where the generator matrices are obtained through the tensor product of the generator matrices of component codes. In this context, the codeword can be represented as an array of size $m \times n$, where the component codes are denoted as the row and column codes. Within this topology, one of our objectives is to investigate maximally recoverable codes for product topology known as Maximally Recoverable Product Codes (MRPCs). In this direction our contributions are as follows:

  - This thesis presents several properties of maximally recoverable product codes. We provide a sufficient condition to identify a specific subclass of erasure patterns as correctable, as well as a necessary condition to determine another subclass of erasure patterns as non-correctable.

- Additionally, a bipartite graph construction is introduced in this thesis to establish relationships between erasures and non-erasures within an erasure pattern for $T_{m,n}(a, b)$. This construction enables the identification of matching conditions applicable to regular erasure patterns (a class of ersure patterns which satisfies the regularity condition), leading to the detection of a subset of recoverable erasure patterns for $a = 2$. The thesis also explores the relationship between projective geometries over finite fields and higher-order MDS codes, establishing a correspondence between $(n, 3)$-MDS(3) codes and $PG(2, q)$ to facilitate the construction of MRPCs for $a = 1$.

- **Sparse and Balanced (SB) generator matrix of maximally recoverable LRCs:** Sparsity refers to the property of having a significant number of zero entries, specifically $n - d$, in every row of the generator matrix, which can lead to more efficient encoding and decoding algorithms. A balanced generator matrix ensures that the number of ones and zeros in each row and column is approximately equal. The balanced property guarantees an almost uniform load distribution on each node. One of our objectives is to provide SB generator matrix for maximally recoverable LRCs. In this aspect, our contributions are as follows:

  - This thesis discusses the existence of sparse generator matrices and under certain conditions, balanced generator matrices for Local MRCs with single erasures for $r = 2$.

- **LRCs with availability:** The inclusion of multiple repair sets for LRCs offers advantages by enabling the distribution of repair tasks across different nodes. This distribution aims to prevent overburdening specific nodes and promote a balanced workload distribution within the system. This class of codes is called LRCs with availability. One of our objectives is to find the upper bound on the minimum distance for Hierarchical LRCs with availability.

- This thesis provides an upper bound of the minimum distance for the specific case of LRCs with the hierarchical locality and availability.

1.5 Organization of Thesis

Rest of the thesis is organized as follows:

- In **Chapter 2**, we provide a comprehensive review of LRCs and MRCs for both independent and correlated failures. It covers the evolution of LRCs from single erasure to multiple erasures, as well as the concept of MRCs.
- In **Chapter 3** we present our constructions of maximally recoverable codes with hierarchical locality, specifically focusing on the Hierarchical Local Code known as Hierarchical Local MRC (HL-MRC). The chapter includes the field sizes required for these constructions, a comparison between the HL-MRC constructions found in previous works and those introduced in this thesis, and a lower bound on the field size for HL-MRC.
- In **Chapter 4**, we explore Maximally Recoverable Product Codes (MRPCs). It delves into our research on the properties of MRPCs, the characterization of recoverable and non-recoverable erasure patterns, and the investigation of the relationship between projective geometries over finite fields and higher-order MDS codes.
- In **Chapter 5**, we focus on constructing sparse and balanced (SB) generator matrices for Local MRCs in the case of single erasure.
- In **Chapter 6**, we review various upper bounds on the minimum distance of LRCs with availability and then present a new upper bound on the minimum distance of Hierarchical Locally Recoverable Codes (H-LRCs) with availability for specific condition.
- In **Chapter 7**, we summarize the main findings of the thesis and discuss the scope of future work.

Chapter 2

Background and Literature Review

In this chapter, an extensive review of the existing literature on LRCs and MRCs is presented. Firstly, a thorough analysis is conducted on the literature concerning independent failures, encompassing single erasures as well. This includes significant optimal code constructions, field sizes, and various bounds, including alphabet-dependent bounds. Following that, an overview of the literature on correlated failures is provided, which includes discussions on suitable topologies, code definitions, and field size bounds.

2.1 Organization of the Chapter

The chapter is organized as follows: Section 2.2 provides a review of the existing literature on LRCs and MRCs for independent failures, while Subsection 2.3 focuses on correlated failures. Specifically, Subsection 2.2.1 introduces the formal definition and upper bound on the minimum distance for LRCs for single erasure, as well as for LRCs with multiple recovery sets for single erasure. Subsection 2.2.2 examines MRCs for single erasure, and Subection 2.2.3 explores LRCs and MRCs for multiple erasures. Section 2.4 presents the conclusion of the chapter.

2.2 LRCs and MRCs for Independent Failures

In general, single node failures occur more frequently. Therefore, considering network bandwidth, it is essential to minimize the repair degree. However, in MDS codes, even for a single node recovery, the repair degree is k. As a result, MDS codes are inefficient for repair operations.

To address this issue and reduce the repair degree for single node failures, a type of erasure codes called ”Locally Recoverable Codes” (LRCs)were developed and deployed by companies like Facebook [1] and Microsoft [2]. These codes have later been extended to handle multiple erasures.

2.2.1 LRCs for Single Erasure

A code $\mathcal{C}[n,k,d]$ is said to have the locally recoverable property of single erasure if each code symbol $c_i \in \mathcal{C}$ can be uniquely determined by at most $r(< k)$ other code symbols of $\mathcal{C}$. The parameter r is called the locality of the code.

A code $\mathcal{C}[n,k,d,r]$ LRC with locality r for single erasures are formally defined in [3] by defining code symbol locality, information symbol locality and all symbol locality as follows:

Definition 1. *Code Symbol Locality (Loc(c_i)): For $c_i \in \mathcal{C}$ to be the smallest integer r for which there exists $S_i \subseteq [n]$ of cardinality r such that*

$$c_i = \sum_{j \in S_i} \lambda_j c_j.$$

Further $Loc(\mathcal{C}) = \max_{i \in [n]} \{Loc(c_i)\}$, here $Loc(\mathcal{C})$ is called locality of the code.

Definition 2. *Information Locality: A code $\mathcal{C}$ has information locality r if there exists $I \subseteq [n]$ with cardinality k and $G|_I$ is full rank such that $Loc(c_i) \leq r$ for all $i \in I$. Here $G|_I$ denotes the restriction of G to the set of columns indexed by the set I.*

If further $\bigcup_{i \in I} S_i = [n]$, then the code is said to have an all-symbol locality.

An LRC code consists of global and local parities. Each local parity symbol is the function of a few specific data symbols, while each global parity symbol is the function of all data symbols. A local group, also known as a local code, consists of a group of data symbols and their corresponding local parity. For instance, as illustrated in Fig. 2.1, the data symbols X_1, X_2 and X_3 form a local group along with local parity P_X, which is a function of all the X_i's. Similarly, another local group is formed with data symbols Y_i's and parity symbol P_Y. The global parities of the code, P_1 and P_2 are functions of all the data symbols X_i's and Y_i's.

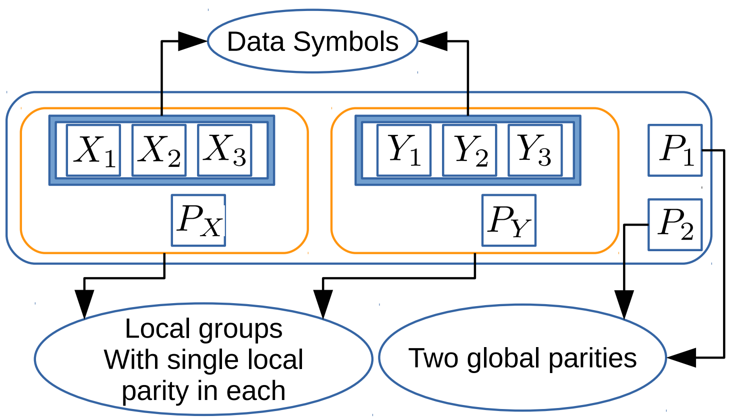

Figure 2.1: LRC Example:$(n = 10, k = 6, r = 3)$ Source [2].

For LRC with a locality parameter r, specifically for single erasure scenarios, the *Singleton-like bound* is obtained by utilizing the following Lemma, which provides the fact about the minimum distance of the code.

Lemma 1. *Let $S \subset [n]$ is the largest cardinality, such that $rank(G|_S) = k-1$ then $d = n-|S|$. Here $rank(G|_S)$ is rank of the matrix $G|_S$.*

In [3], an algorithm is provided to calculate the lower bound on the cardinality of such set S for the code $\mathcal{C}[n, k, d, r]$. Using this algorithm, the upper bound on the minimum distance of the code (Eq. (2.1)) is determined in [3].

$$d \leq n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right). \tag{2.1}$$

The bound says that as the code's locality parameter r decreases, the optimal minimum distance of the code also decreases, and (2.1) becomes (1.1) for $r = k$. However, in practical applications, the reduction of reconstruction cost takes priority of importance over storage overhead, leading to the preference for selecting a lower value of locality.

In [4], a specific type of LRCs with information locality known as Pyramid ocdes were introduced and these codes were found to be optimal later in [3]. In [5], Tamo and Barg constructed optimal codes with all-symbol locality for any value of $r \ (1 < r < k)$ and $n \mid (r+1)$. Each codeword is generated through the evaluation of specially constructed polynomials known as good polynomials over a finite field. This particular encoding scheme enables the recovery of lost symbol through polynomial interpolation at r points. The construction method for codes with multiple disjoint recovery sets for each symbol is also provided, allowing for enhanced availability of frequently accessed data, commonly referred to as hot data. Furthermore, in [6], a construction for LRCs with locality r and τ distinct recovery sets (referred to as "LRCs with (r, τ) - Availability") for each information symbol is presented using Gabidulin codes. Additionally, the paper introduces an upper bound, as given in equation (2.2), on the minimum distance for this specific class of codes, assuming each repair group consists of only one parity symbol.

$$d \leq n - k + 1 - \left(\left\lceil \frac{\tau k}{r} \right\rceil - \tau \right). \tag{2.2}$$

Also, the more general bound for LRCs with (r, τ) - Availability (6.1) is given in [6].

$$d \leq (n - k + 1) - \left(\left\lceil \frac{\tau(k-1)+1}{\tau(r-1)+1} \right\rceil - 1 \right). \tag{2.3}$$

The paper [7] presents the construction of LRCs that are optimal and almost optimal in terms of all-symbol locality. Here, "almost optimal" refers to codes whose minimum distance differs by at most one from the optimal value derived from the Singleton-like bound for LRCs (2.1). For locality $r = 2$ and $d = 2, 6, 10$, the construction of binary LRCs that are optimal in dimension is given in [8]. Optimal ternary locally repairable codes for all possible 8 classes, where the minimum distance can only be 2, 3, 4, 5, or 6, are constructed in [9]. In [10], for small locality $r = 2$ and minimum distance 6, optimal constructions are obtained using a parity check matrix and the sunflower construction. Additionally, optimal cyclic LRCs are given for distances 3 and 4 in the same paper.

2.2.2 MRCs for Single Erasures

Definition 3. *Maximally Recoverable Code (MRC): A code is said to be maximally recoverable if it can recover from all the information-theoretically recoverable erasure patterns given the locality constraints of the code.*

In [11], MRCs for data-local codes and local codes were introduced by considering the systematic code $\mathcal{C}[n,k,d,r,h]_q$ as follows.

Definition 4. *Data Local Code: Let $\mathcal{C}$ be a linear systematic $[n,k]$ code. $\mathcal{C}$ is a $[n,k,d,r,h]_q$ data-local code if it satisfies the following conditions:*

- $r \mid k$ *and* $n = k + \frac{k}{r} + h$,
- *The data symbols are divided into $\frac{k}{r}$ groups of size r. Each group has a corresponding (local) parity symbol that stores the XOR of the respective data symbols,*
- *The remaining h global parity symbols can depend on all k data symbols.*

A group of r data symbols and their local parity is referred to as a *local group*.

Definition 5. *A data-local code is said to be maximally recoverable if, for any set $E \subseteq [n]$ obtained by selecting one coordinate from each of the k/r local groups, puncturing the code $\mathcal{C}$ at the coordinates in E results in an $[k+h,k]$ MDS code.*

Definition 6. *Local Code: Let $\mathcal{C}$ be a linear systematic $[n,k]$ code. $\mathcal{C}$ is a $[n,k,d,r,h]_q$ local code if the following conditions are satisfied:*

- $r \mid (k+h)$ *and* $n = k + \frac{(k+h)}{r} + h$,
- *There are k data symbols and h global parity symbols, where each global parity may depend on all data symbols.*

- *These $k+h$ symbols are divided into $\frac{(k+h)}{r}$ groups of size r. Each group has a corresponding (local) parity symbol that stores the XOR of the respective data symbols.*

A group of r symbols and their local parity is referred to as a *local group*.

Definition 7. *The local code is called Maximally recoverable if, for any set $E \subseteq [n]$ obtained by selecting one coordinate from each of the $(k+h)/r$ local groups, puncturing the code $\mathcal{C}$ at the coordinates in E results in an $[k+h, k]$ MDS code.*

The advantages of codes with specific parameters $[n, k, d, r, h]_q$ are as follows: i) Local parities expedite the recovery process for single node failures by utilizing corresponding local groups. ii) Global parities enable the recovery of additional node failures. iii) MRCs can recover the maximum number of erasures beyond the minimum distance of the code.

For instance, let's consider the maximal erasure pattern E in the local code, which consists of erasures from each local group (one per group) and an additional h erasures from any location within the code. The size of E is $\left(\frac{k+h}{r} + h\right)$. It is important to note that E satisfies the recovery condition as defined in Definition 7. Moreover, we observe that $|E| > d - 1 = (n-k) - \left(\left\lceil \frac{k}{r} \right\rceil - 1\right) = \left(\frac{k+h}{r} + h\right) - \left(\left\lceil \frac{k}{r} \right\rceil - 1\right)$.

Table 2.1 tabulates various bounds for LRCs for single erasures, mainly the upper bound on the minimum distance, alphabet dependent bounds, and field size for certain optimal LRC and MRC constructions. Cadambe-Mazumdar bound [12] is an alphabet dependent bound derived using the quantity (i.e., the largest possible minimum distance of the code $[n, k]$ over a given alphabet size) as a parameter under locality constraints. A similar kind of technique has been applied using generalized hamming weights ($d_i^{\perp}$) and gaps ($g_i^{\perp}$) of dual code $\mathcal{C}^{\perp}$ [13] to derive the alphabet dependent bound. A study is done on the upper bound on the maximal length of optimal LRC for the given q [14]. In [11], the authors presented constructions of data-local and local MRC for all values of k, r and h and also studied improved field size constructions for $h \geq 2^r + 1$, $h = 3$ and $h = 4$.

Table 2.1: Bounds and Constructions for Single Erasures

LRC	**Type of bound / construction**	**Bound / Size of n or q**	**Reference**
$[n, k, d, r]_q$	Upper bound on d	$d \leq n - k + 1 - (\lceil \frac{k}{r} \rceil - 1)$	[3]
	Alphabet-Dependent bound on d	$k \leq \min_{t \in \mathbb{Z}_+} [tr + k_{opt}^{(q)}(n - t(r+1), d)]$ $k_{opt}^{(q)}(n, d)$ is largest possible dimension	[12]
		$k \leq \min_{1 \leq i \leq g_k^{\perp} - k} [k_{opt}^{(q)}(n - d_i^{\perp}, d) - i + d_i^{\perp}]$	[13]
	Upper bound on n for the given q	$n \leq q + k + \lceil k/r \rceil - 2$, if $r \geq 2$ and $r \mid k$ $n \leq q + k + \lceil k/r \rceil - 2$, if $r \geq 2$ and $k \geq 2 \mod r$, $n \leq 2q + k + \lceil k/r \rceil - 2$, if $r \geq 2$ and $k = 1 \mod r$, $n \leq 2q + k + \lceil k/r \rceil - 2$, if $r = 1$	[14]
	Length of the optimal LRC	for $d > 5$, $n = O(dq^{3+\frac{4}{d-4}})$ for $d = 5$, $n = O(q^2)$	[15]
	Field Size q of existing MRC construction	for data-local & loca MRC: $q = O\left(k^{h-1}\right)$ $q = O\left(k^{\lceil (h-1)(1-\frac{1}{2^r}) \rceil}\right)$, for $h \geq 2^r + 1$ local MRC: $q = O(k^{3/2})$, if $h = 3$ $q = O(k^{7/3})$, if $h = 4$	[11]

2.2.3 LRCs and MRCs for Multiple Erasures

Multiple erasures can be distinguished as independent and correlated failures based on the type of failures in DSS. Individual failures are always independent, and correlated failures in distributed storage systems refer to a situation where multiple components or nodes fail simultaneously within a short period due to a common cause. For example, it can be caused by the outage of a power source shared by multiple nodes in the network. In general, correlated failures occur as the length of the code increases [16].

(1) LRCs for Independent Failures:

Independent node failures are common in DSS over different networks. For such multiple failures, *Codes with* (r,δ) *Locality* are defined in [17] and [18] as follows:

Definition 8. (r,δ) *Code Symbol Locality: The* i^{th} *code symbol,* c_i, $i \in [n]$, *of* $\mathcal{C}$ *is said to have* (r,δ) *locality,* $\delta \geq 2$, *if there exists a subset* $S_i \subseteq [n]$ *such that*

- $i \in S_i, |S_i| \leq r+\delta-1$ *and*
- *minimum distance of* $\mathcal{C}|_{S_i} \geq \delta$, *where* $\mathcal{C}|_{S_i}$ *denotes the code obtained when* $\mathcal{C}$ *is punctured to the set of co-ordinates corresponding to* S_i.

Definition 9. (r,δ) *Information Locality: The code* $\mathcal{C}$ *is said to have* (r,δ) *information locality if* $\mathcal{C}$ *has a set of punctured codes* $\{\mathcal{C}_i\}_{i\in\mathcal{L}}$ *with supports* $\{S_i\}_{i\in\mathcal{L}}$, *respectively, such that, for all* $i \in \mathcal{L}$, *we have*

- $|S_i| \leq r+\delta-1$
- $d_{min}(\mathcal{C}_i) \geq \delta$ *and*
- *rank*$(G|_{\cup_{i\in\mathcal{L}} S_i}) = k$

Here $\mathcal{L}$ *denotes the index set for the local codes, and we denote the restriction of* G *to the set of columns indexed by the set* S *by* $G|_S$.

If further $\bigcup_{i\in\mathcal{L}} S_i = [n]$, then the code is said to have (r,δ) all-symbol locality. An example is shown in Figure 2.2. In this example, the number of erasures less than or equal to 3 can be recovered by contacting 6 other symbols instead of 18.

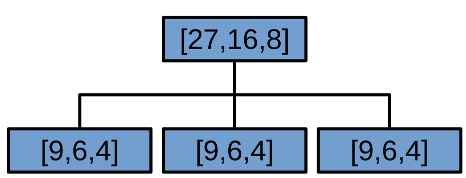

Figure 2.2: Example:$[27,16,8]$-code having $(r=6,\delta=4)$ locality.

In [17], the upper bound on minimum distance is derived as given in (2.4) and also provided construction for optimal code.

$$d \leq n-k+1-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1). \tag{2.4}$$

Such a code can tolerate the $(\delta - 2)$ additional failures for any given node failure. The upper bound on d is further reduced as the code is conditioned to offer the locality for multiple erasures and for $\delta = 2$, (2.4) becomes (2.1). These codes consist of multiple overlapped local parities through which multiple recovery sets are available for the set of node failures.

Some important bounds pertaining to the above discussed (r, δ) locality codes are tabulated in Table 2.2, which includes the upper bound and alphabet-dependent upper bound on the minimum distance and upper bound on the length of the optimal code for the given q. As an extension of the Cadambe-Mazumdar bound for the (r, δ) code, the alphabet-dependent bound is derived using the Griesmer bound on the length of the code by recursively taking the residual codes, in [19]. For the linear code, $\mathcal{C}[n, k, d]_q$ the Griesmer bound $(\mathcal{G}(k, d))$ is the lower bound on the length of the code for the given code dimension k and minimum distance d, i.e. $n \geq \sum_{i=0}^{k-1} \left\lceil \frac{d}{q^i} \right\rceil = \mathcal{G}(k, d)$. In [20], derived the upper bound on the length of the optimal code, which is super-linear in the field size and given a general construction.

Table 2.2: Bounds for multiple erasures

LRC	**Type of bound/ construction**	**Bound**	**Reference**
$[n, k, d, r, \delta, h]_q$ where $\delta \geq 2$	Upper bound on d	$d \leq n - k + 1 - (\lceil \frac{k}{r} \rceil - 1)(\delta - 1)$	[17]
	Alphabet-Dependent bound on d	replaced r with κ κ = maximum dimension of all local groups $k \leq \min_{\lambda \in \mathbb{Z}_+} \{\lambda + s\}$, here, $s = k_{opt}^{(q)}(n - (r+1)\mathcal{G}(\kappa, \delta) + \mathcal{G}(\kappa - b, \delta), d)$ where $a, b \in \mathbb{Z}$ such that $\lambda = a\kappa + b, 0 \leq b < \kappa$	[19]
	Bound on n for the optimal code	For $n = m(r + \delta - 1)$, $k = ur + v$, $u \geq 2(r - v + 1)$ or $v = 0$: $t = \lfloor \frac{(d-1)}{\delta} \rfloor$ $n \leq O\left(\frac{t(r+\delta)}{r} q^{\frac{(m-u)r-v}{\lfloor t/2 \rfloor}} - 1\right)$	[20]

Optimal codes with all-symbol locality for $\delta = 2$ has been constructed in [21] with the field size $O(n^k)$ for $r+1 \mid n$. In [22], optimal codes with all-symbol locality for $\delta \geq 2$ is constructed for $n \nmid (r+\delta-1)$ and $r \nmid k$ with the field size greater than or equal to $\binom{n-1}{k-1}$ (Algorithm 2, Theorem 16). Additionally, in the same paper for $n \mid (r+\delta-1)$ proved that the construction (Algorithm 1) with the field size greater than or equal to $\binom{n-1}{k-1}$ (Theorem 15), which is lower than the construction given by [21]. However, the construction provided in [21] is easy to implement because the codes are based on Reed-Solomon coded blocks that are re-encoded in a way that provides low repair locality. The authors constructed the optimal codes using the rank metric and subspace metric in [23] and using matrix product codes in [24]. Optimal cyclic LRCs are introduced in [25], and a class of such codes is given in [26].

Algebraic curves are used in many applications of coding theory. LRCs on algebraic curves have been extensively studied in recent years, as they offer good trade-offs between minimum distance and rate of the code. An algebraic curve is a geometric object defined by a polynomial equation in two variables, denoted by x and y. An algebraic curve in the Cartesian plane is represented by the set of points (x, y) that satisfy the polynomial equation $f(x, y) = 0.$, where f is a polynomial with coefficients in a given field. The degree of the polynomial f determines the degree of the algebraic curve. Reed-Solomon (RS) codes connected to algebraic curves. RS codes use encoding and decoding mechanisms that involve the polynomial evaluation and interpolation techniques associated with algebraic curves. The codeword is essentially the polynomial evaluated at specific points on the curve. Several families of LRCs on algebraic curves have been proposed in the literature, including LRCs on Hermitian curves, LRCs on hyperelliptic curves, and LRCs on the projective line. Hermitian curve is more specific. The Hermitian curve takes the form of $x^{q+1} = y^q + y$ considered over the field $\mathbb{F}_{q^2}$, where q is any prime power. The RS-like codes given in [5] are extended to multiple recovering sets using Hermitian curves [27]. LRCs on elliptic curves in [28], and LRCs in Hermitian function fields with certain type of divisors in [29] are studied.

The concept of locality has been extended to the hierarchical locality in [30]. In the case of (r, δ) locality, the code offers no locality if there are more than δ erasures. In the case of codes with hierarchical locality, the locality constraints are such that with the increase in the number of erasures, the locality increases in steps. The following is the definition of code with a 2-level hierarchical locality.

Definition 10. *2-level Hierarchical Locality: A $[n, k, d]$ linear code $\mathcal{C}$ is a code with hierarchical locality having locality parameters $[(r_1, \delta_1), (r_2, \delta_2)]$ if for every symbol $c_i, 1 \leq i \leq n$, there exists a punctured code C_i such that $c_i \in Supp(C_i)$ and the following conditions hold: 1) $dim(C_i) \leq r_1$, 2) $d_{min}(C_i) \geq \delta_1$, 3) C_i is a code with (r_2, δ_2) locality.*

An example of code with 2-level hierarchical locality is shown in Fig. 2.3. In this example, the code offers locality 3 for a single erasure and 8(< 14) for 2 erasures.

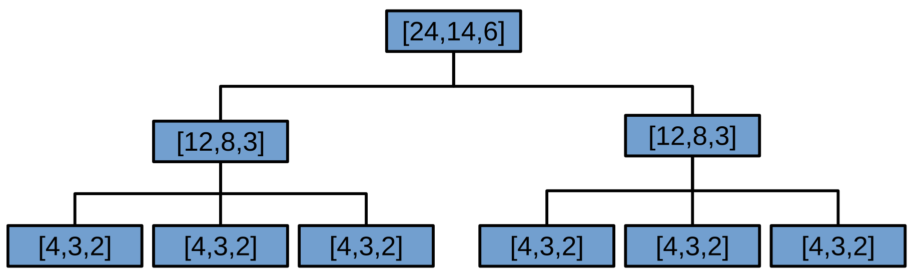

Figure 2.3: Example:$[24, 14, 6]$-code having 2-level hierarchical locality [30].

The upper bound on minimum distance is derived as

$$d \leq n - k + 1 - \left(\left\lceil \frac{k}{r_2} \right\rceil - 1 \right) (\delta_2 - 1) - \left(\left\lceil \frac{k}{r_1} \right\rceil - 1 \right) (\delta_1 - \delta_2) . \tag{2.5}$$

Constructions of optimal codes with hierarchical locality based on optimal (r, δ) LRCs have been done via generalized Reed-Solomon in [31], through matrix product codes in [32] and from covering maps of curves in [33]. Cyclic codes with hierarchical locality, which are optimal with respect to minimum distance bound, have been constructed in [34]. Constructions of codes with hierarchical locality based on maps between algebraic curves have been presented in [35]. In [36], alphabet-dependent bounds for codes with hierarchical locality have been derived (equation (2.6)), and also punctured simplex codes have been shown to be optimal w.r.t. the upper bound on the minimum distance (2.5).

$$k \leq \min_{\lambda \in \mathbb{Z}_+} [\lambda + k_{opt}^{(q)}(n - v, d)] \tag{2.6}$$

here, $v = \lambda + \lfloor \frac{\lambda}{r_2} \rfloor (\delta_2 - 1) + \lfloor \frac{\lambda}{r_1} \rfloor (\delta_2 - \delta_1)$.

Till now, we have studied codes with locality for single erasure, codes with locality for multiple erasures, codes with 2−level hierarchical locality, and codes with locality and availability for single erasure. The comparison between the upper bounds on the minimum distances of all the topologies discussed above is shown in figure (Fig. 2.4). In this, we considered fixed values for $(n = 60, r_1 = 9, r_2 = 5, \delta_1 = 4, \delta_2 = 2, \tau = 4)$. We took 10 samples by decreasing the values of k from 30 to 21. Plotted the graph code rate (k/n) versus relative distance (d/n).

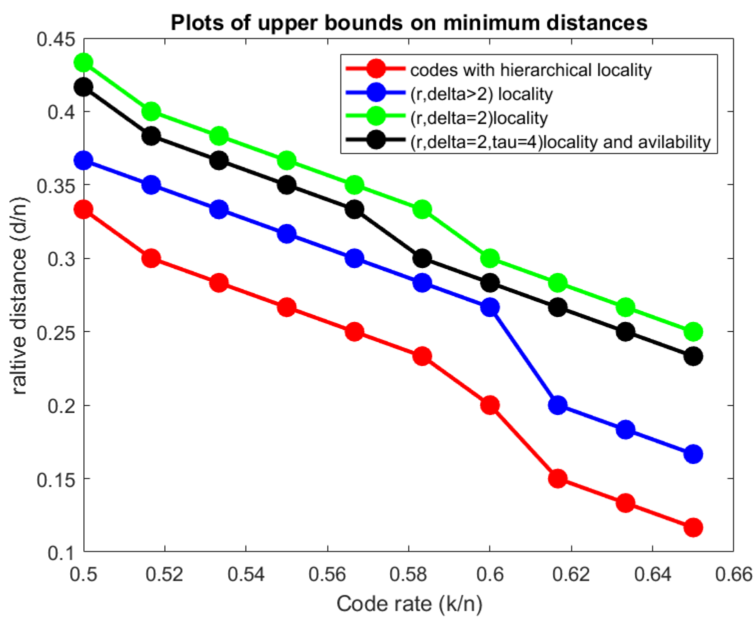

Figure 2.4: Comparison of upper bounds on d

(2) MRCs for Independent Failures:

In [37], data-local MRC and local code MRC are defined for data-local, and local codes for codes with (r, δ) locality as follows:

Definition 11. *Data Local MRC: Let $\mathcal{C}$ be a systematic $[n, k, d]$ code. We say that $\mathcal{C}$ is a $[k, r, h, \delta]$ data-local maximally recoverable code if the following conditions are satisfied:*

- *$r \mid k$ and $n = k + \frac{k}{r}\delta + h$.*
- *Data symbols are partitioned into $\frac{k}{r}$ groups of sizer r. For each such group, there are δ local parity symbols.*

- *For any set $E \subseteq [n]$ where E is obtained by picking δ coordinates from each $\frac{k}{r}$ local groups, restricting $\mathcal{C}$ to coordinates in $[n] \setminus E$ yields an $[k+h, k]$ MDS code.*

$[k, r, h, \delta]$ *data-local MRC is optimum with respect to the minimum distance bound given in* (2.4). *The minimum distance of a data-local code MRC is* $d = h + \delta + 1$.

Definition 12. *Local MRC: Let $\mathcal{C}$ be a systematic $[n, k, d]$ code. We say that $\mathcal{C}$ is a $[k, r, h, \delta]$ local maximally recoverable code if the following conditions are satisfied:*

- *$r \mid (k+h)$ and $n = k + \frac{k+h}{r}\delta + h$.*
- *There are k data symbols and h global parity symbols, where each global parity may depend on all data symbols.*
- *These $k+h$ symbols are partitioned into $\frac{k+h}{r}$ groups of size r. For each group, there are δ local parity symbols.*
- *For any set $E \subseteq [n]$ where E is obtained by picking δ coordinates from each $\frac{k+h}{r}$ local groups, restricting $\mathcal{C}$ to coordinates in $[n] \setminus E$ yields an $[k+h, k]$ MDS code.*

A group of r data symbols and their δ local parities is referred to as a *local group*. Local MRC $\mathcal{C}[k, r, h, \delta]$ is optimum with respect to the minimum distance bound given in (2.4). The minimum distance of a $[k, r, h, \delta]$ local MRC is as given below:

$$d = h + \delta + 1 + \left\lfloor \frac{h}{r} \right\rfloor \delta. \tag{2.7}$$

The lower bound on the field size for local MRC and the required field size of a general construction of local MRC codes are tabulated in Table 2.3. In [38], by exploiting the MRC property on the parity check matrix H derived the lower bound on the field size, and thus, the lower bound shows that one needs super linear size fields to instantiate H to make all non-trivial minors nonzero. Local MRCs are also known in the literature as Partial-MDS codes (PMDS) codes [39]. We interchange PMDS codes and Local MRCs with the context of interest. Let's describe the Gabidulin code before introducing the general construction of PMDS codes citecalis2017general.

Gabidulin code $\mathscr{C}_{GC}[n_1, k, d_1 = n_1 - k + 1]$ over F_{q^M} is Maximum Rank Distance (MRD) code. In this code, the codeword is obtained by evaluating the linearized polynomial $f(x) = \sum_{i=0}^{k-1} a_i x^{q^i}$ with the set of n_1 linearly independent elements from F_{q^M} over F_q, here the coefficients $a_i \in F_{q^M}$ are information symbols. Suppose $\{\alpha_1, \alpha_2, \ldots \alpha_{n_1}\} \subset F_{q^M}$ is linear independent set over F_q then the codeword $\boldsymbol{c} = (\sum_{i=0}^{k-1} a_i \alpha_1^{q^i}, \sum_{i=0}^{k-1} a_i \alpha_2^{q^i}, \ldots, \sum_{i=0}^{k-1} a_i \alpha_{n_1}^{q^i}) \in \mathscr{C}_{GC}$. MRD codes are optimal with respect to the rank distance metric, and a brief explanation of them follows below:

Rank distance metric between two vectors $X_1, X_2 \in F_q^{M \times n}$ is defined by $d_R(X_1, X_2) = \text{rank}(X_1 - X_2)$ over F_q. Minimum rank distance of any code $\mathcal{C}[n, k]$ over $F_{q^M}^n$ ($\cong F_q^{M \times n}$) with rank distance metric is $d_R(\mathcal{C}) = \min\{d_R(\boldsymbol{c_1}, \boldsymbol{c_2}) \mid \forall\, \boldsymbol{c_1} \neq \boldsymbol{c_2} \in \mathcal{C}\}$. The code is called MRD if $d_R(\mathcal{C}) = n - k + 1$.

A general construction of PMDS codes (Local MRCs) which is obtained by using the Gabidulin code and MDS codes as described in [40]. The MRD property of Gabidulin code is used to construct the generator matrix of a PMDS code, which has the following structure:

$$G = G_{GC} \left[\begin{array}{c|c|c|c} G_{MDS} & 0 & \cdots & 0 \\ \hline 0 & G_{MDS} & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & G_{MDS} \end{array} \right] \tag{2.8}$$

Here, G_{GC} is the generator matrix of Gabidulin code $\mathscr{C}_{GC}[n_1 = lr, k, d_1 = n_1 - k + 1]$ over the field F_{q^M} and G_{MDS} is the generator matrix of MDS code $\mathscr{C}_{MDS}[n_2 = r + m, r, d_2 = m + 1]$ over the field F_q. Field size for the construction of PMDS code is tabulated in Table 2.3.

Constructions of PMDS codes with two and three global parities have been discussed in [41, 42]. Improved construction of PMDS codes for all parameters over small field sizes has been presented in [43]. Construction of MRCs over small field sizes have been investigated in [44, 45]. Recently, the construction of MRCs based on linearized Reed Solomon codes via generator matrix approach is presented in [46]. The construction is given for nonuniform local codes(i.e., the size of the local code and locality are unequal). The required field size

Table 2.3: Bounds for Maximally Recoverable Codes of multiple erasures

Local MRC	**Type of bound/ construction**	**Bound / Field size**	**Reference**
$[n, k, d, r, \delta, h]_q$ where $\delta \geq 2$	Lower bound on q	$q \geq \Omega_{h,(\delta-1)}\left(n(r+\delta-1)^{\alpha}\right)$, where $\alpha = \frac{\min\left\{(\delta-1), \left(h - \left\lceil \frac{h}{t} \right\rceil\right)\right\}}{\left\lceil \frac{h}{t} \right\rceil}$ and $t = \frac{n}{(r+\delta-1)}$	[38]
	Field size for PMDS Code	$n = m(r+\delta-1)$ & $k = mr - h$ $q = (q')^{mr}$, where q' is prime power and $q' \geq (r+\delta-1)$	[40]

is $O\left(\left(\frac{k+h}{r}\right)^{r}\right)$, where r is the maximum locality among all localities of local codes. In [47], MRC is constructed through the parity check matrix approach using skew polynomials only for uniform local codes. For this construction, the field size is $O\left(\left(\frac{k+h}{r}\right)^{\min(h,r)}\right)$. The field size of the construction given in [47] is lower than the construction given in [46] for $h \leq r$. Both the codes exploit the sum rank property to get an MRC, which is explained in detail in Subsection 3.3.3. In [48], linearized Reed-Solomon codes are used to construct MR codes with field size $O(\max\left\{\frac{n}{r+\delta-1}, r+\delta-1\right\}^{h})$. The field size is equal to the construction in [47]. Whereas construction in [46] outperforms when h is relatively larger than r.

In [49], authors proved a new lower bound on the field size of LRCs. Additionally, Maximally Recoverable codes are constructed, which are cyclic. They have shown that these cyclic MRCs have optimal field size in certain cases. The connection between MRCs and matroids has been established in [50] and subsequently minors of MRCs were studied in [51].

For the 2-level hierarchical locality, data-local code, local code, and their corresponding MRCs are defined in [37], which are aligned with our work. So we discuss these codes in detail in Section 3.2.

2.3 LRCs and MRCs for Correlated Failures

In practice, correlated failures may take place due to rack failure, a data center failure, or the failure of a power source shared by multiple machines. To deal with these types of failures, a different kind of topology called grid-like topologies $T_{m,n}(a,b,h)$ are proposed in [16]. In these topologies, each codeword is an array of size $m \times n$, where a, b and h are the number of parities per column, per row, and global, respectively. In this code, each codeword is like an array with multiple rows and columns, in which each row is a codeword of a certain row code and similarly each column is the codeword of a certain column code (see Fig. 2.5). In addition, there are global parities anywhere in the grid.

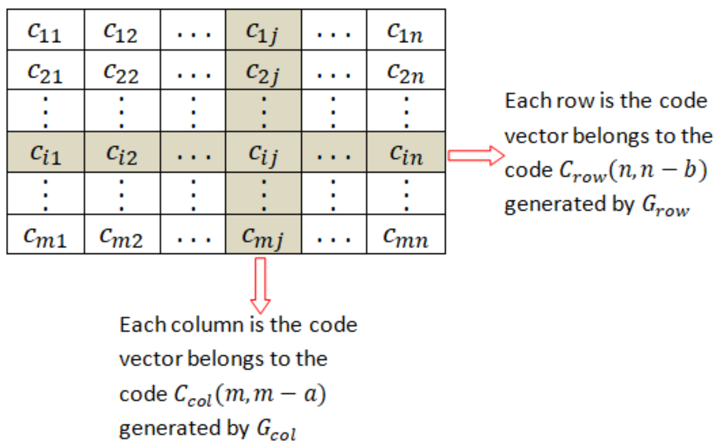

Figure 2.5: Structure of the codeword in grid-like topologies.

Definition 1 (Code Instantiating a Topology $T_{m,n}(a,b,h)$)**.** *Consider a code $\mathcal{C}$ in which each codeword is a matrix C of size $m \times n$, with c_{ij} denoting the $(i,j)^{th}$ coordinate of the codeword. The code $\mathcal{C}$ of length mn is said to instantiate a topology $T_{m,n}(a,b,h)$ if for some $b \times n$ matrix H_{row}, $a \times m$ matrix H_{col} and $h \times n$ matrix H_{glob}, it satisfies the following conditions:*

1. *$\mathcal{C}$ punctured to a row i satisfies a set of 'b' parity equations given by*

$$H_{row}\,[c_{i1}, c_{i2}, \ldots, c_{in}]^t = \mathbf{0}, \;\; \forall\, i \in [m].$$

The b parity equations given by H_{row} need not be linearly independent and hence the code whose parity-check matrix is H_{row} has parameters $[n, \geq n-b]$ code and is denoted by $\mathcal{C}_{row}$.

2. *$\mathcal{C}$ punctured to a column j satisfies a set of 'a' parity equations given by*

$$H_{col}\,[c_{1j}, c_{2j}, \ldots, c_{mj}]^t = \mathbf{0}, \;\; \forall\, j \in [n].$$

Similar to the first condition, the code whose parity-check matrix is H_{row} has parameters $[m, \geq m-a]$ code and is denoted by $\mathcal{C}_{col}$.

3. *In addition, every codeword in $\mathcal{C}$ satisfies a set of 'h' parity equations (referred to as global parities) given by*

$$H_{glob}\, Vec(C) = \mathbf{0},$$

where Vec(C) is obtained by vectorizing the codeword C (matrix of size $m \times n$) by reading row after row.

A topology $T_{m,n}(a,b,h)$ with $h = 0$ will be referred to as product topology.

MRCs for Correlated failures:

Let's begin by examining the definition of recoverable erasure patterns, which will allow us to define MRCs for the code instantiating a topology $T_{m,n}(a,b,h)$.

Definition 13. *Recoverable Erasure Pattern [16]: An erasure pattern $E \subseteq [m] \times [n]$ is a set of symbols that are erased. The pattern E is recoverable for the topology $T_{m,n}(a,b,h)$ if there exists a code instantiating the topology where the variables $\{x_{ij}\}_{(i,j)\in E}$ can be recovered from the parity check equations.*

Definition 14. *[16]: A code $\mathcal{C}$ that instantiates the topology $T_{m,n}(a,b,h)$ is Maximally Recoverable (MR) if it recovers every failure pattern that is recoverable for the topology.*

Proposition 1. *[16]: Let $\mathcal{C}$ be an MR instantiation of the topology $T_{m,n}(a,b,h)$. We have*

(a) the dimension of $\mathcal{C}$ is given by

$$dim(\mathcal{C}) = (m-a)(n-b) - h. \tag{2.9}$$

Moreover,

$$dim(\mathcal{C}_{col}) = (m-a) \text{ and } dim(\mathcal{C}_{row}) = (n-b). \tag{2.10}$$

(b) Let $U \subseteq [m]$, $|U| = m-a$ and $V \subseteq [n]$, $|V| = n-b$ be arbitrary. Then $\mathcal{C}|_{U\times V}$ is an $[(m-a)(n-b), (m-a)(n-b)-h, h+1]$ MDS code. Any subset $S \subseteq U \times V$, $|S| = (m-a)(n-b)-h$ is an information set.

(c) Assume $h \leq (m-a)(n-b) - max(m-a), (n-b)$, then the code $\mathcal{C}_{col}$ is an $[m, m-a, a+1]$ MDS code and the code $\mathcal{C}_{col}$ is an $[n, n-b, b+1]$ MDS code. Moreover, for all $j \in [n]$, $\mathcal{C}$ restricted to column j is the code $\mathcal{C}_{col}$ and for all $i \in [m]$, $\mathcal{C}$ restricted to row i is the code $\mathcal{C}_{row}$

In order to find such MR code for $T_{m,n}(a, b, h)$, finding an MR code for product topology $T_{m,n}(a, b, h = 0)$ is itself crucial. As a result, knowing the set of recoverable erasure patterns for the given topology $T_{m,n}(a, b, h = 0)$ is an intriguing question to resolve.

Characterization of recoverable erasure patterns for the $T_{m,n}(a, b, h = 0)$

Definition 15. *Irreducible Erasure Pattern (IEP): An erasure pattern is an IEP, in which the number of erasures in any row is either 0 or greater than or equal to $b+1$ and in any column is 0 or greater than or equal to $a+1$.*

An Example of IEP is given in Figure 2.6 for the topology $T_{6,10}(1, 2, 0)$. In which the first 5 rows have at least 3 erasures in each, and the last row has no erasure. Also, the first two columns have zero erasure and the remaining ones have at least 2 erasures.

Note 1. *Suppose some column has a or fewer erasures, or some row has b for fewer erasures. Those erasures can be decoded by using iterative row-column decoding. This is because row code and column code are MDS codes (see Proposition 1(c)).*

The characterization of recoverable erasure patterns is to see which IEPs are recoverable. To address this problem, the authors presented a necessary condition for a pattern to be recoverable. Also, conjecture (Conjecture 1) that this condition is also sufficient, and proved this conjecture for $a = 1$.

	1	2	3	4	5	6	7	8	9	10
1							×	×	×	×
2						×	×	×		
3			×						×	×
4				×	×	×				
5			×	×	×					
6										

Figure 2.6: An example of IEP for $(m, n) = (6, 10)$ and $(a, b, h) = (1, 2, 0)$.

Definition 16 (Regular Erasure Pattern (REP) for Topology $T_{m,n}(a, b)$)**.** *An erasure pattern $E \subseteq [m] \times [n]$ is said to be regular for topology $T_{m,n}(a, b)$ if*

$$|E \cap (U \times V)| \leq uv - \max(u - a, 0)\max(v - b, 0), \tag{2.11}$$

for all $U \subseteq [m]$, $V \subseteq [n]$ and $|U| = u$, $|V| = v$.

We can check an Example of IEP given in Figure 2.6 satisfies the regularity condition given in (2.11) for every $U \subseteq [6]$, $V \subseteq [10]$ as shown in Figure 2.7. Therefore, it is REP.

	1	2	3	4	5	6	7	8	9	10
1							×	×	×	×
2						×	×	×		
3			×						×	×
4				×	×	×				
5			×	×	×					
6										

■ $U = \{4,5\}, V = \{3,4,5\}$
$5 \leq 3 * 1 + 2 * 2 - 2 = 5$

■ $U = \{2,3,4,5\}, V = \{3,4,5,6\}$
$8 \leq 4 * 1 + 4 * 2 - 2 = 10$

Figure 2.7: An example of REP for $(m, n) = (6, 10)$ and $(a, b, h) = (1, 2, 0)$.

Conjecture 1. *An erasure pattern E is recoverable for $T_{m,n}(a, b, h = 0)$ if and only if it is regular.*

However, this conjecture is disproved by providing the counter example in [52]. So for $a > 1$, the characterization of recoverable erasure patterns needs to be addressed.

MRC for grid-like topologies have been studied in [16] and a super-polynomial lower bound on the field size of these MRCs has been derived. MRC for grid-like topologies which can recover from all bounded erasures (bounded by a constant) have been investigated in [53]. Constructions of MRPCs for topologies $T_{4,n}(1, 2)$ and $T_{3,n}(1, 3)$ and lower bounds on the field size for these specific topologies have been presented in [54]. In [55], higher order MDS codes have been introduced (denoted by MDS(l) and will be defined in a later chapter). It has been shown that it is necessary and sufficient for the row codes to be higher order MDS codes for MRPCs of topology $T_{m,n}(a = 1, b)$. This has been in turn used to derive a lower bound on the field size of MRPC for $T_{m,n}(a, b)$, given as $q \geq \Omega_l(n^{\min\{m-a+1,b,n-b\}-1})$.

2.4 Conclusion

In this chapter, an introduction to the concepts of LRCs and MRCs in the context of independent failures and correlated failures is provided as the foundation for the thesis. A summary of the bounds and constructions of codes for independent failures is presented for LRCs with single parity and (r, δ) locality. As an extension to (r, δ) locality codes, we also introduced 2-level hierarchical locality codes. Based on this, we are going to discuss our work on MRCs for 2-level hierarchical locality codes in Chapter 3. Furthermore, literature pertaining to codes with grid-like topologies and product topologies, which are specifically defined to handle correlated failures, has been discussed. Chapter 4 will discuss our work specifically related to product topologies.

Chapter 3

Hierarchical Local Maximally Recoverable Codes

In this chapter, we explore the concept of maximally recoverable codes with hierarchical locality. In Section 2.2.3, we have seen that codes with hierarchical locality have been proposed to provide multiple levels of locality as the number of erasures increases. The presence of multiple levels of locality enables faster data recovery as the number of erasures increases. Maximally recoverable codes (MRCs) tolerate the maximum number of erasures given the locality constraints. The primary objective of this chapter is to investigate the design and properties of HL-MRC, which combine the benefits of hierarchical locality and maximally recoverable codes.

Within the scope of our thesis, we assume that hierarchical locality implies a 2-level hierarchical locality. We know that the construction of the code can be obtained through the generator matrix or the parity check matrix. We define the topology (support constraints) that determine the structure of the parity check matrix and define as Hierarchical Local Maximally Recoverable Codes (HL-MRCs). Using this structure, we characterize the correctable erasure patterns set, establishing conditions the parity check matrix must satisfy to obtain HL-MRC. We derive the lower bound on the field size of HL-MRC. We give constructions of MRC with hierarchical locality for certain parameter values whose field size is smaller than earlier known constructions. We also present a general construction via parity check matrix, for which we use generator matrices of linearized Reed-Solomon (LRS) codes, and the field size is better than the earlier known construction under certain parameter regimes. In addition, we give a random construction of MRC with hierarchical locality through the generator matrix approach and characterize the field size required. For a few cases, we also provided lower bounds on the field sizes for HL-MRC.

3.1 Organization of the Chapter

This chapter is organized as follows: Section 3.2 provides an in-depth explanation of hierarchical local codes, their topology, and the conditions that the parity check matrix of hierarchical local codes must satisfy to obtain HL-MRC. Section 3.3 provides constructions of HL-MRC, along with the corresponding required field size, for the following scenarios: i) For $h_1 = 1$ and $h_2 = 1$, ii) For $h_1 = 2$ and $h_2 = 1$, general constructions iii) Based on Linearized Reed-Solomon Codes via parity check matrix, and iv) Random Construction via generator matrix. Section 3.4 presents lower bounds on the field size of HL-MRC. Finally, conclusion of this chapter are presented in Section 3.5.

3.2 Maximally Recoverable Codes with Hierarchical Locality

In this section, we recall the definitions of hierarchical data-local and local MRCs and illustrate the definitions through an example [37]. First, we define the topology corresponding to hierarchical data local code by specifying the number and kind of parities the code must satisfy. Then, we will define hierarchical data local MRC.

Definition 2 (Hierarchical Data Local Code)**.** *We define a $[k, r_1, r_2, h_1, h_2, \delta]$ hierarchical data local (HDL) code of length*

$$n = k + h_1 + \frac{k}{r_1}(h_2 + \frac{r_1}{r_2}\delta)$$

as follows:

- *The code symbols $c_1, \ldots, c_n$ satisfy h_1 global parities given by $\sum_{j=1}^{n} u_j^{(\ell)} c_j = 0, \;\; 1 \leq \ell \leq h_1$.*

- *The first $n-h_1$ code symbols are partitioned into $t_1 = \frac{k}{r_1}$ groups $A_i = \{(i-1)n_1+j \mid j \in [n_1]\}, 1 \leq i \leq t_1$ such that $|A_i| = r_1 + h_2 + \frac{r_1}{r_2}\delta = n_1$. The code symbols in the i^{th} group, $1 \leq i \leq t_1$ satisfy the following h_2 mid-level parities $\sum_{j=1}^{n_1} v_{i,j}^{(\ell)} c_{(i-1)n_1+j} = 0,\ \ 1 \leq \ell \leq h_2$.*

- *The first $n_1 - h_2$ code symbols of the i^{th} group, $1 \leq i \leq t_1$ are partitioned into $t_2 = \frac{r_1}{r_2}$ groups $B_{i,s}, 1 \leq i \leq t_1, 1 \leq s \leq t_2$ such that $|B_{i,s}| = r_2 + \delta = n_2$. The code symbols in the $(i,s)^{th}$ group, $1 \leq i \leq t_1, 1 \leq s \leq t_2$ satisfy the following δ local parities $\sum_{j=1}^{n_2} w_{i,s,j}^{(\ell)} c_{(i-1)n_1+(s-1)n_2+j} = 0,\ \ 1 \leq \ell \leq \delta$.*

Definition 3 (Hierarchical Data Local MRC)**.** *Let $\mathcal{C}$ be an $[k, r_1, r_2, h_1, h_2, \delta]$ HDL code. Then C is maximally recoverable if for any set $E \subset [n]$ such that $|E| = k + h_1$ and*

1. *$E \bigcap B_{i,s} \leq r_2\ \forall\, i, s,$*
2. *$E \bigcap A_i = r_1\ \forall\, i,$*

the punctured code $\mathcal{C}|_E$ is an $[k + h_1, k, h_1 + 1]$ MDS code.

The above definition characterizes the set of all erasure patterns which are information-theoretically correctable by any hierarchical data local code. Justification for this is similar to one provided for Definition 5 below, and hence omitted. Replacing the information locality with all-symbol locality, we have the following definition of the topology of hierarchical local codes and their corresponding MRCs.

Definition 4 (Hierarchical Local Code)**.** *We define a $[k, r_1, r_2, h_1, h_2, \delta]$ hierarchical local (HL) code of length $n = k + h_1 + \frac{k+h_1}{r_1}(h_2 + \frac{r_1+h_2}{r_2}\delta)$ as follows:*

- *The code symbols $c_1, \ldots, c_n$ satisfy h_1 global parities given by $\sum_{j=1}^{n} u_j^{(\ell)} c_j = 0,\ \ 1 \leq \ell \leq h_1$.*

- *The n code symbols are partitioned into $t_1 = \frac{k+h_1}{r_1}$ groups $A_i, 1 \leq i \leq t_1$ such that $|A_i| = r_1 + h_2 + \frac{r_1+h_2}{r_2}\delta = n_1$. The code symbols in the i^{th} group, $1 \leq i \leq t_1$ satisfy the following h_2 mid-level parities $\sum_{j=1}^{n_1} v_{i,j}^{(\ell)} c_{(i-1)n_1+j} = 0,\ \ 1 \leq \ell \leq h_2$.*

- *The n_1 code symbols of the i^{th} group, $1 \leq i \leq t_1$ are partitioned into $t_2 = \frac{r_1+h_2}{r_2}$ groups $B_{i,s}, 1 \leq i \leq t_1, 1 \leq s \leq t_2$ such that $|B_{i,s}| = r_2 + \delta = n_2$. The code symbols in the $(i,s)^{th}$ group, $1 \leq i \leq t_1, 1 \leq s \leq t_2$ satisfy the following δ local parities $\sum_{j=1}^{n_2} w_{i,s,j}^{(\ell)} c_{(i-1)n_1+(s-1)n_2+j} = 0, \;\; 1 \leq \ell \leq \delta$.*

From the definition of HL code (Definition 4), supports of local parities are disjoint for any two different groups of (i,s) ($1 \leq i \leq t_1, 1 \leq s \leq t_2$) as well as supports of h_2 mid-level parities for any two different groups of i ($1 \leq i \leq t_1$) are disjoint. Therefore, the parity check matrix of the code follows a certain type of structure, which is illustrated with the following example (Example 1).

Example 1. *We demonstrate the structure of the parity check matrix for an $[k = 5, r_1 = 3, r_2 = 2, h_1 = 1, h_2 = 1, \delta = 2]$ HL code. The length of the code is $n = k + h_1 + \frac{k+h_1}{r_1}(h_2 + \frac{r_1+h_2}{r_2}\delta) = 16$. The parity check matrix of the code is given below. In H, $M_{i,j}$'s are the parity check constraints of $\delta = 2$ local parities of $(i,j)^{th}$ group, N_i's are the parity check constraints of $h_2 = 1$ mid-level parities of i^{th} group and O is parity check constraints of global parities of the code.*

$$H = \begin{bmatrix} M_{1,1} & & & & \\ & M_{1,2} & & & \\ & N_1 & & & \\ & & M_{2,1} & & \\ & & & M_{2,2} & \\ & & & N_2 & \\ & & O & & \end{bmatrix}$$

where,

$$M_{i,j} = \begin{bmatrix} w_{i,j,1}^{(1)} & w_{i,j,2}^{(1)} & w_{i,j,3}^{(1)} & w_{i,j,4}^{(1)} \\ w_{i,j,1}^{(2)} & w_{i,j,2}^{(2)} & w_{i,j,3}^{(2)} & w_{i,j,4}^{(2)} \end{bmatrix}$$

$$N_i = \begin{bmatrix} v_{i,1}^{(1)} & v_{i,2}^{(1)} & \dots & v_{i,8}^{(1)} \end{bmatrix}$$

$$O = \begin{bmatrix} u_1^{(1)} & u_2^{(1)} & \dots & u_{16}^{(1)} \end{bmatrix}$$

Definition 5 (Hierarchical Local MRC)**.** *Let $\mathcal{C}$ be an $[k, r_1, r_2, h_1, h_2, \delta]$ HL code. Then C is maximally recoverable if for any set $E \subset [n]$ such that $|E| = k + h_1$ and*

1. $E \bigcap B_{i,s} \leq r_2 \; \forall \, i, s$,

2. $E \bigcap A_i = r_1 \; \forall \, i$,

the punctured code $\mathcal{C}|_E$ is an $[k + h_1, k, h_1 + 1]$ MDS code.

We will prove in the following that the above definition gives the set of all erasure patterns which are information-theoretically correctable. To characterize the set of all erasure patterns which can be corrected by any hierarchical local code, it is enough to consider maximal erasure patterns, i.e., erasure patterns of size $n - k$.

Lemma 3.2.1. *Suppose $\overline{E} \subseteq [n]$ be the subset of any maximal erasure pattern (as mentioned above) of size $|\overline{E}| = t_1 t_2 \delta + t_1 h_2$. If E violates any of the following conditions,*

1) $E \bigcap B_{i,s} \leq r_2 \; \forall \, i, s$,

2) $E \bigcap A_i = r_1 \; \forall \, i$,

then $\overline{E}$ is not correctable.

Proof. Since h_1 parities are global, these erasures can be placed anywhere in the code, we will look at the subset of maximal erasure patterns $\overline{E}$ (E in Definition 5 refers to non-erased locations) of size $|\overline{E}| = t_1 t_2 \delta + t_1 h_2$.

First, we will show that Condition 2) is necessary. Suppose for some i, $E \cap A_i > r_1$ and $\overline{E}$ is a subset of correctable maximal erasure pattern say $\overline{E} \cup \{e_{h_1}\}$ (here $\{e_{h_1}\}$ is the set of additional h_1 number of erasures). Since $dim(\mathcal{C}|_{A_i}) \leq r_1$, we can add extra $(|E \cap A_i| - r_1)$ erasures in A_i, and let the extended erasure pattern be denoted by $\overline{E}^{'}$. If $\overline{E} \cup \{e_{h_1}\}$ is correctable, then so is $\overline{E}^{'} \cup \{e_{h_1}\}$. However, $|E^{'}| + h_1$ is greater than the number of parities of the code $(n - k)$ and hence we demonstrated a correctable erasure pattern of size greater than the number of parities of the code. This is a contradiction and $\overline{E}$ is not correctable. Using similar steps, we can also prove that condition 1) is necessary. □

The Lemma 3.2.1 characterizes the set of erasure patterns which are not correctable. The fact that all other maximal erasure patterns are correctable. To summarize, any correctable maximal erasure pattern can have δ erasures anywhere in every local code $B_{i,s}$, additional h_2 erasures anywhere in each middle code A_i and remaining other h_1 erasures anywhere in the code $\mathcal{C}$.

In [37], derived the properties that the middle codes of an HDL/HL-MRC have to be data-local and local MRC respectively. Also given a method to derive any HDL-MRC from an HL-MRC. Hence, we will discuss the constructions of HL-MRC.

In an independent parallel work [46], a class of MRCs known as multi-layer MRCs have been introduced. We would like to note that hierarchical local MRCs (given in Definition 5) form a subclass of these multi-layer MRCs. One key difference between the codes constructed in [46] and our work is that the authors in [46] take the generator matrix based approach, and we take the parity-check matrix based approach. The constructions in [46] are based on maximum sum-rank distance codes. Also, in [46], a generalized framework of hierarchical LRCs with unequal locality parameters is considered. We believe that our framework of HL-MRC and the subsequent constructions can be extended to the case of unequal locality parameters as well, but we leave it for future work.

A comparison of the HL-MRC constructions known from earlier works and those in this thesis is given in Table 3.1. In the following, we provide parameter regimes where the field size of Construction 1 in [46] is lower than that of Construction IV.2 in [37] and vice versa.

- For $r_1 \leq h_1$, we have $((\delta+1)(h_2+1)h_1 - 1) > r_1$. Since $n > \frac{k+h_1}{r_1}$, Construction 1 in [46] has a lower field size than that of Construction IV.2 in [37].

- For the fixed values of δ, h_1, h_2, if asymptotically $k = \Theta(n)$ and $r_1 = \sqrt{n}$, then the field size required for Construction 1 in [46] is $q = \Theta(n^{\sqrt{n}/2})$. Construction IV.2 in [37] has a field size given by $q = \Theta(n^c)$ (here c is constant, since δ, h_1 and h_2 are $O(1)$). Hence, in this parameter regime, Construction IV.2 in [37] has a lower field size.

Under certain parameter regimes, if $r_1 > h_1 h_2$ and $t_1 \geq \max(t_2 + 1, n_2)$, the field size of Construction 3.3.7 is lower than the Construction 1 given in [46]. The random construction provided in Section 3.3.4 can be constructed if the field size is $\geq 3(n-k)\binom{n-1}{k-1}$. When the field size is exactly $3(n-k)\binom{n-1}{k-1}$, it is $O(n^k)$, which is higher than any other construction.

Reference	Parameters of HL-MRC	Field Size
Construction 1 in [46]	$[k, r_1, r_2, h_1, h_2, \delta]$	$O\left(\left(\frac{k+h_1}{r_1}\right)^{r_1}\right)$
Construction IV.2 in [37]	$[k, r_1, r_2, h_1, h_2, \delta]$	$O(n_2 n_1^{(\delta+1)h_2-1} n^{(\delta+1)(h_2+1)h_1-1})$
Construction V.1 in [37]	$[k, r_1, r_2, h_1 = 1, h_2, \delta]$	$O(n_2 n_1^{(\delta+1)(h_2+1)-1})$
Construction 3.3.1 in the current thesis	$[k, r_1, r_2, h_1 = 1, h_2 = 1, \delta]$	$O(n_1)$
Construction 3.3.4 in the current thesis	$[k, r_1, r_2, h_1 = 2, h_2 = 1, \delta]$	$O(n^4)$
Construction 3.3.7 LRS based (parity check)	$[k, r_1, r_2, h_1, h_2, \delta]$	$O((\max(t_2+1, n_2))^{h_2 h_1})$
Construction given in Section 3.3.4 Randomized	$[k, r_1, r_2, h_1, h_2, \delta]$	$\geq 3(n-k)\binom{n-1}{k-1}$

Table 3.1: Summary of HL-MRC constructions. Recall that $n_2 = r_2 + \delta$, $n_1 = r_1 + h_2 + \frac{r_1+h_2}{r_2}\delta$,$n = k + h_1 + \frac{k+h_1}{r_1}(h_2 + \frac{r_1+h_2}{r_2}\delta)$.

3.3 Constructions of HL-MRCs and its Field Size

In this section, we discuss the constructions of HL-MRC for $h_1 = 1\,\&\,h_2 = 1$, that are constructed using Vandermode matrices and for $h_1 = 2\,\&\,h_2 = 1$ using Cauchy matrices. Additionally, we present two more general constructions: one through the parity check matrix approach using the generator matrix of the Linearized Reed-Solomon code and the other involving a Random Construction approach through the generator matrix.

3.3.1 HL-MRC for $h_1 = 1$ and $h_2 = 1$

Now we will describe the construction for the case when there is one mid-level parity per mid-level code ($h_2 = 1$) and one global parity ($h_1 = 1$). This construction is based on the construction of local MRC with 2 global parities in [56].

Construction 3.3.1. *We give a construction of the code $\mathcal{C}$, which is specified by the following matrix H:*

$$H = \begin{bmatrix} H_0 & & & \\ & H_0 & & \\ & & \ddots & \\ & & & H_0 \\ H_1 & H_2 & \dots & H_{t_1} \end{bmatrix} \quad H_0 = \begin{bmatrix} M_0 & & & \\ & M_0 & & \\ & & \ddots & \\ & & & M_0 \\ M_1 & M_2 & \dots & M_{t_2} \end{bmatrix}$$

$$M_0 = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_{n_2} \\ \alpha_1^2 & \alpha_2^2 & \dots & \alpha_{n_2}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^\delta & \alpha_2^\delta & \dots & \alpha_{n_2}^\delta \end{bmatrix} \quad M_i = \begin{bmatrix} \lambda_i & \lambda_i & \dots & \lambda_i \end{bmatrix}$$

$$H_i = \begin{bmatrix} H_{i,1} & H_{i,2} & \dots & H_{i,t_2} \end{bmatrix}$$

$$H_{i,s} = \begin{bmatrix} \alpha_1^{\delta+1} & \alpha_2^{\delta+1} & \dots & \alpha_{n_2}^{\delta+1} \end{bmatrix}$$

is a parity check matrix for an $[k, r_1, r_2, 1, 1, \delta]$ HL-MRC if the following conditions are satisfied:

- *q is a prime power such that there exists a subgroup G of $\mathbb{F}_q^*$ of size at least n_2 and with at least t_2 cosets.*

- $\alpha_1, \alpha_2, \ldots \alpha_{n_2} \in G$ *and* $\alpha_i \neq \alpha_j$.

- $\lambda_1, \lambda_2, \ldots, \lambda_{t_2} \in \mathbb{F}_q^*$ *be elements from distinct cosets of* G.

We make use of the following determinantal identity to show that the matrix formed by the columns of the parity check matrix corresponding to the erased positions are invertible and hence can be recovered.

Lemma 3.3.2 ([56]). *Let* $C_1, \cdots, C_h$ *be* $a \times (a+1)$ *dimensional matrices and* $D_1, \cdots, D_h$ *be* $h \times (a+1)$ *dimensional matrices over a field and let* $D_i^{(j)}$ *be the* j^{th} *row of* D_i. *Then,*

$$\det \left[\begin{array}{c|c|c|c} C_1 & 0 & \cdots & 0 \\ \hline 0 & C_2 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & C_h \\ \hline D_1 & D_2 & \cdots & D_h \end{array}\right]$$

$$= (-1)^{\frac{ah(h-1)}{2}} \det \begin{bmatrix} \det \begin{pmatrix} C_1 \\ D_1^{(1)} \end{pmatrix} & \cdots & \det \begin{pmatrix} C_h \\ D_h^{(1)} \end{pmatrix} \\ \vdots & \ddots & \vdots \\ \det \begin{pmatrix} C_1 \\ D_1^{(h)} \end{pmatrix} & \cdots & \det \begin{pmatrix} C_h \\ D_h^{(h)} \end{pmatrix} \end{bmatrix}.$$

Theorem 3.3.3. *The code* C *given by Construction 3.3.1 is a* $[k, r_1, r_2, h_1 = 1, h_2 = 1, \delta]$ *HL-MRC. The field size required for the Construction 3.3.1 is* $O(n_1)$.

Proof. To show that the code is a $[k, r_1, r_2, 1, 1, \delta]$ HL-MRC, we consider erasure patterns where there are δ erasures per local code, one erasure per mid-level code [1] and one more erasure anywhere in the global code. We will show that any such erasure pattern is recoverable. Since

[1]This terminology is somewhat non-standard but convenient for description. δ erasures per local code refers to erasures whose coordinates fall within the support of parities corresponding to a local and similar is the case for mid-level code.

there is only one global erasure and it can be in one mid-level code, we consider that the mid-level code which has additional global erasure has index l and for all $j \neq l$, there are no global erasures associated with these mid-level codes. We will show that erasures in each mid-level code can be recovered.

Case 1: Consider the j^{th} mid-level code ($j \neq l$). Let the local code within the mid-level code where the erasure occurs be j'. The submatrix B_j of the parity-check matrix which is used to recover the erasures within the j^{th} mid-level code is given by,

$$B_j = \begin{bmatrix} \alpha_{j'_1} & \alpha_{j'_2} & \dots & \alpha_{j'_\delta} & \alpha_{j'_{\delta+1}} \\ \alpha^2_{j'_1} & \alpha^2_{j'_2} & \dots & \alpha^2_{j'_\delta} & \alpha^2_{j'_{\delta+1}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha^\delta_{j'_1} & \alpha^\delta_{j'_2} & \dots & \alpha^\delta_{j'_\delta} & \alpha^\delta_{j'_{\delta+1}} \\ \lambda_{j'} & \lambda_{j'} & \dots & \lambda_{j'} & \lambda_{j'} \end{bmatrix}$$

where $\{j'_1, \dots j'_{\delta+1}\}$ denote the $\delta+1$ erased coordinates in the local group j'. We can clearly see that this matrix is a Vandermonde matrix after scaling and permuting rows. Hence $\det(B_j) \neq 0$.

Case 2: After all the erasures corresponding to $j \neq l$ mid-level codes are recovered, for the l^{th} mid-level code, we will also involve the global parity. This case can again be divided into two sub cases depending on the local group where the extra erasure happens.

Case 2a: Both the mid-level erasure and the global erasure occur in the same local code, l'. The matrix B_l which multiplies the erased symbols in the l^{th} mid-level code is given by

$$B_l = \begin{bmatrix} \alpha_{l'_1} & \alpha_{l'_2} & \dots & \alpha_{l'_\delta} & \alpha_{l'_{\delta+1}} & \alpha_{l'_{\delta+2}} \\ \alpha^2_{l'_1} & \alpha^2_{l'_2} & \dots & \alpha^2_{l'_\delta} & \alpha^2_{l'_{\delta+1}} & \alpha^2_{l'_{\delta+2}} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \alpha^\delta_{l'_1} & \alpha^\delta_{l'_2} & \dots & \alpha^\delta_{l'_\delta} & \alpha^\delta_{l'_{\delta+1}} & \alpha^\delta_{l'_{\delta+2}} \\ \lambda_{l'} & \lambda_{l'} & \dots & \lambda_{l'} & \lambda_{l'} & \lambda_{l'} \\ \alpha^{\delta+1}_{l'_1} & \alpha^{\delta+1}_{l'_2} & \dots & \alpha^{\delta+1}_{l'_\delta} & \alpha^{\delta+1}_{l'_{\delta+1}} & \alpha^{\delta+1}_{l'_{\delta+2}} \end{bmatrix}.$$

This is similar to above where B_l after scaling and permuting the rows is also a Vandermonde matrix. Hence $\det(B_l) \neq 0$.

Case 2b: The mid-level and global erasure occur in different local codes. Let those local codes be l' and l''. The matrix B_l has the determinant as follows,

$$\det(B_l) =$$

$$\det \begin{bmatrix} \alpha_{l'_1} & \dots & \alpha_{l'_{\delta+1}} & & & \\ \alpha^2_{l'_1} & \dots & \alpha^2_{l'_{\delta+1}} & & & \\ \vdots & \ddots & \vdots & & & \\ \alpha^\delta_{l'_1} & \dots & \alpha^\delta_{l'_{\delta+1}} & & & \\ & & & \alpha_{l''_1} & \dots & \alpha_{l''_{\delta+1}} \\ & & & \alpha^2_{l''_1} & \dots & \alpha^2_{l''_{\delta+1}} \\ & & & \vdots & \ddots & \vdots \\ & & & \alpha^\delta_{l''_1} & \dots & \alpha^\delta_{l''_{\delta+1}} \\ \lambda_{l'} & \dots & \lambda_{l'} & \lambda_{l''} & \dots & \lambda_{l''} \\ \alpha^{\delta+1}_{l'_{\delta+1}} & \dots & \alpha^{\delta+1}_{l'_{\delta+1}} & \alpha^{\delta+1}_{l''_{\delta+1}} & \dots & \alpha^{\delta+1}_{l''_{\delta+1}} \end{bmatrix}$$

$$= \det \begin{bmatrix} \det V_1 & \det V_2 \\ \det V_3 & \det V_4 \end{bmatrix}, \text{ where}$$

$$V_1 = \begin{pmatrix} \alpha_{l'_1} & \dots & \alpha_{l'_{\delta+1}} \\ \alpha^2_{l'_1} & \dots & \alpha^2_{l'_{\delta+1}} \\ \vdots & \ddots & \vdots \\ \alpha^{\delta}_{l'_1} & \dots & \alpha^{\delta}_{l'_{\delta+1}} \\ \lambda_{l'} & \dots & \lambda_{l'} \end{pmatrix}, V_2 = \begin{pmatrix} \alpha_{l''_1} & \dots & \alpha_{l''_{\delta+1}} \\ \alpha^2_{l''_1} & \dots & \alpha^2_{l''_{\delta+1}} \\ \vdots & \ddots & \vdots \\ \alpha^{\delta}_{l''_1} & \dots & \alpha^{\delta}_{l''_{\delta+1}} \\ \lambda_{l''} & \dots & \lambda_{l''} \end{pmatrix},$$

$$V_3 = \begin{pmatrix} V_1(1:\delta,:) \\ \alpha^{\delta+1}_{l'_{\delta+1}} \ \dots \ \alpha^{\delta+1}_{l'_{\delta+1}} \end{pmatrix}, V_4 = \begin{pmatrix} V_2(1:\delta,:) \\ \alpha^{\delta+1}_{l''_{\delta+1}} \ \dots \ \alpha^{\delta+1}_{l''_{\delta+1}} \end{pmatrix},$$

where $V_1(1:\delta,:)$ and $V_2(1:\delta,:)$ are the first δ rows of matrices V_1 and V_2 respectively.

Based on the above equations, $\det(B_l) = 0$ if and only if

$$\det \begin{bmatrix} \lambda_{l'} & \lambda_{l''} \\ \prod_{i=1}^{\delta+1} \alpha_{l'_i} & \prod_{i=1}^{\delta+1} \alpha_{l''_i} \end{bmatrix} = 0,$$

where we factored out the non-zero Vandermonde determinants from each column. Since $\alpha_{l'_i}, \alpha_{l''_i} \in G$ and $\lambda_{l'}, \lambda_{l''}$ are in different cosets of G, the last determinant cannot be zero. Hence, we proved that the code can recover from all possible erasure patterns specified by the definition of HL MRC and hence it is an HL MRC with the corresponding parameters.

In Construction 3.3.1, we require to consider a finite field $\mathbb{F}_q$, such that there exists a subgroup G of $\mathbb{F}_q^*$ of size at least n_2 and with at least t_2 cosets. Based on Lemma 4 in [56], there exists a finite field $\mathbb{F}_q$ of size $O(n_2 t_2) = O(n_1)$. □

3.3.2 HL-MRC for $h_1 = 2$ and $h_2 = 1$

Now, we will provide the construction of a HL-MRC with 2 global parities ($h_1 = 2$) and 1 mid-level parity per mid-level code ($h_2 = 1$).

Construction 3.3.4. *We give a construction of code $\mathcal{C}$, which is specified by the following parity-check matrix H:*

$$H = \begin{bmatrix} H_0 & & & \\ & H_0 & & \\ & & \ddots & \\ & & & H_0 \\ H_1 & H_2 & \dots & H_{t_1} \end{bmatrix} \quad H_0 = \begin{bmatrix} M_0 & & & \\ & M_0 & & \\ & & \ddots & \\ & & & M_0 \\ M_1 & M_2 & \dots & M_{t_2} \end{bmatrix}$$

$$M_0 = \begin{bmatrix} \frac{1}{\alpha_1-\beta_1} & \frac{1}{\alpha_2-\beta_1} & \dots & \frac{1}{\alpha_{n_2}-\beta_1} \\ \frac{1}{\alpha_1-\beta_2} & \frac{1}{\alpha_2-\beta_2} & \dots & \frac{1}{\alpha_{n_2}-\beta_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\alpha_1-\beta_\delta} & \frac{1}{\alpha_2-\beta_\delta} & \dots & \frac{1}{\alpha_{n_2}-\beta_\delta} \end{bmatrix}$$

$$M_i = \begin{bmatrix} \frac{1}{\alpha_1-\beta_{\delta+1}} & \frac{1}{\alpha_2-\beta_{\delta+1}} & \dots & \frac{1}{\alpha_{n_2}-\beta_{\delta+1}} \end{bmatrix}$$

$$H_i = \begin{bmatrix} H_{i,1} & H_{i,2} & \dots & H_{i,t_2} \end{bmatrix}$$

$$H_{i,s} = \begin{bmatrix} \frac{\lambda_{s+(i-1)t_2}}{\alpha_1-\beta_{\delta+2}} & \frac{\lambda_{s+(i-1)t_2}}{\alpha_2-\beta_{\delta+2}} & \dots & \frac{\lambda_{s+(i-1)t_2}}{\alpha_{n_2}-\beta_{\delta+2}} \\ \frac{\mu_{s+(i-1)t_2}}{\alpha_1-\beta_{\delta+3}} & \frac{\mu_{s+(i-1)t_2}}{\alpha_2-\beta_{\delta+3}} & \dots & \frac{\mu_{s+(i-1)t_2}}{\alpha_{n_2}-\beta_{\delta+3}} \end{bmatrix}$$

The parameters described in the above parity-check matrix are picked as follows:

- $q_0 \geq 2(n_2 + \delta) + 3$ *is a prime power.*
- *There exists a subgroup G of $\mathbb{F}_{q_0}^*$ of size at least $n_2 + 2$ with atleast $t_1 t_2$ cosets, where $\mathbb{F}_{q_0}^*$ is the multiplicative group of the field $\mathbb{F}_{q_0}$.*
- *$\mathbb{F}_q$ is an extension field of $\mathbb{F}_{q_0}$.*
- *$\mu_1, \dots, \mu_{t_1 t_2}$ are picked from distinct cosets of G.*

- *Choose distinct* $\beta_{\delta+1}, \beta_{\delta+2}, \beta_{\delta+3} \in \mathbb{F}_{q_0}$.

- *Pick* $\alpha_1, \dots \alpha_{n_2} \in \mathbb{F}_{q_0}$ *such that,* $\frac{\alpha_i - \beta_{\delta+2}}{\alpha_i - \beta_{\delta+3}}, \frac{\alpha_i - \beta_{\delta+1}}{\alpha_i - \beta_{\delta+3}} \in G$.

- *Pick distinct* $\beta_1, \dots, \beta_\delta \in \mathbb{F}_{q_0} \setminus \{\alpha_1, \dots, \alpha_{n_2}, \beta_{\delta+1}, \beta_{\delta+2}, \beta_{\delta+3}\}$.

- $\lambda_1, \lambda_2, \dots, \lambda_{t_1 t_2} \in \mathbb{F}_q$ *are picked 4 wise-independent over* $\mathbb{F}_{q_0}$.

Theorem 3.3.5. *The code* C *given by Construction 3.3.4 is a* $[k, r_1, r_2, h_1 = 2, h_2 = 1, \delta]$ *HL-MRC. The field size required for the Construction 3.3.4 is* $O(n^4)$.

Proof. Similar to the previous proof, we consider the case when there are δ erasures per local code, one erasure per mid-level code and two more global erasures anywhere in the code. We again look at the erasure patterns within each mid-level code. The following distinct patterns are possible with respect to the mid-level codes.

1. No global erasures occur in that mid-level code.

2. Either one or both of the global erasures occur in the mid-level code.

We prove that all the above erasure patterns are recoverable.

Case 1: When no global erasures occur in the mid-level code, there are δ erasures per local code and one more erasure per mid-level code. In this scenario, we involve the mid-level parities. Let l be the affected mid-level code and l' be the local code within the mid-level code where the erasure occurs. Let $\gamma_{i,j} = \frac{1}{\alpha_j - \beta_i}$. For this case based on erased coordinates, The matrix B_l is defined as follows:

$$B_l = \begin{bmatrix} \gamma_{1,l'_1} & \gamma_{1,l'_2} & \cdots & \gamma_{1,l'_{\delta+1}} \\ \gamma_{2,l'_1} & \gamma_{2,l'_2} & \cdots & \gamma_{2,l'_{\delta+1}} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{\delta,l'_1} & \gamma_{\delta,l'_2} & \cdots & \gamma_{\delta,l'_{\delta+1}} \\ \gamma_{\delta+1,l'_1} & \gamma_{\delta+1,l'_2} & \cdots & \gamma_{\delta+1,l'_{\delta+1}} \end{bmatrix},$$

where $\{l'_1, l'_2, \ldots, l'_{\delta+1}\}$ are the erased coordinates in local code l'. This is a Cauchy matrix and hence $\det(B_l) \neq 0$.

Case 2: For this case, we assume that all the erasures corresponding to mid-level codes where no global parities are involved are recovered. Then, we can have either one or two mid-level codes to be considered which can have global erasures. When there are global erasures, there are δ erasures per local code, one erasure per mid-level code and two more erasures anywhere in the code. We will only list the cases (6 in total) of erasure patterns here. We refer the reader to Appendix A for details of the proof, where we derive that in each of the following cases, the parity-check matrix restricted to the erased columns is full rank.

1. Two global erasures are in the same local code as the mid-level erasure.
2. Two global erasures are in one local code and the mid-level erasure is in a different local code for that mid-level code.
3. Two global and the one mid-level erasure are all in three different local codes within the same mid-level code.
4. Two global erasures are in different mid-level codes. In both the mid-level codes, the mid-level erasure is in the same local code as the global erasure.
5. Two global erasures are in two different mid-level codes and in each mid-level code, they are in a different local code as compared to the mid-level erasure for that mid-level code.
6. Two global erasures are in two different mid-level codes. In one of the mid-level code, the mid-level erasure is in the same local code as the global erasure, while in the other they are in different local codes.

In Construction 3.3.4, we require to consider a finite field $\mathbb{F}_{q_0}$, such that there exists a subgroup G of $\mathbb{F}^*_{q_0}$ of size at least n_2 and with at least $t_1 t_2$ cosets. Based on Lemma 4 in [56], there exists a finite field $\mathbb{F}_{q_0}$ of size $O(n_2 t_1 t_2) = O(n)$. Consider a field $\mathbb{F}_q$ which is a degree 4 extension of $\mathbb{F}_{q_0}$ and let v_0, v_1, v_2, v_3 form a basis of $\mathbb{F}_q$ over $\mathbb{F}_{q_0}$. Consider $\lambda_1, \lambda_2, \ldots, \lambda_{t_1 t_2}$ such that $\lambda_i = v_0 + v_1\mu_i + v_2\mu_i^2 + v_3\mu_i^3$, where μ_i are all picked from distinct cosets of G in

$\mathbb{F}_{q_0}$. Then, we can easily prove that $\lambda_1, \lambda_2, \ldots, \lambda_{t_1 t_2}$ are all 4-wise independent over $\mathbb{F}_{q_0}$. Thus, the required field size is $q_0^4 = O(n^4)$. For detailed proof, please refer to Appendix A.1. $\square$

3.3.3 HL-MRC based on Linearized Reed-Solomon Codes

There were two approaches considered to construct MRCs based on generator matrices of linearized Reed-Solomon codes - the first is generator matrix based [46] and the second is parity check matrix based [48], [47]. We give the construction of an HL-MRC based on the generator matrices of linearized Reed-Solomon codes via parity check matrix. The field size of this construction is better than that of previously known construction.

We will briefly describe the constructions of linearized Reed-Solomon codes and a certain sum rank property of these codes [57], which facilitates the construction of MRCs with locality.

Consider a finite field $\mathbb{F}_{q^m}$ and the following field automorphism given by $\sigma : \mathbb{F}_{q^m} \to \mathbb{F}_{q^m}$, where $\sigma(a) = a^q, a \in \mathbb{F}_{q^m}$. An $\mathbb{F}_q$ linear operator $\mathcal{D} : \mathbb{F}_{q^m} \to \mathbb{F}_{q^m}$ is given by

$$\mathcal{D}_\alpha^i(\beta) = \sigma^i(\beta) \prod_{j=0}^{i-1} \sigma^j(\alpha). \tag{3.1}$$

Using the above linearized operator, we define Vandermonde-like matrices as follows:

$$D_q(\alpha, \mathcal{B}, k, l) = \begin{bmatrix} \beta_1 & \beta_2 & \ldots & \beta_l \\ \mathcal{D}_\alpha^1(\beta_1) & \mathcal{D}_\alpha^1(\beta_2) & \ldots & \mathcal{D}_\alpha^1(\beta_l) \\ \vdots & \vdots & & \vdots \\ \mathcal{D}_\alpha^{k-1}(\beta_1) & \mathcal{D}_\alpha^{k-1}(\beta_2) & \ldots & \mathcal{D}_\alpha^{k-1}(\beta_l) \end{bmatrix}, \tag{3.2}$$

where $\mathcal{B} = \{\beta_1, \beta_2, \ldots, \beta_l\} \subset \mathbb{F}_q^m$.

Definition 6 (Linearized Reed-Solomon Code)**.** *Consider a code of length N, dimension k and $N = l_1 + \ldots + l_g$. Consider a finite field $\mathbb{F}_q$ and its extension $\mathbb{F}_{q^M}$. Let $\mathcal{B} = \{\beta_1, \ldots, \beta_M\}$ be the basis of $\mathbb{F}_{q^M}$ over $\mathbb{F}_q$. Also let γ denote the primitive element of $\mathbb{F}_{q^M}$. Let $\mathcal{B}_j$ denote the subset $\{\beta_1, \ldots, \beta_{l_j}\}$. Now, a linearized Reed-Solomon code has the following generator* *matrix:*

$$\begin{aligned} G_{LRS} &= \left[D_q(\gamma^0, \mathcal{B}_1, k, l_1) | D_q(\gamma^1, \mathcal{B}_2, k, l_2) | \ldots | \right. \\ & \qquad \left. D_q(\gamma^{g-1}, \mathcal{B}_g, k, l_g) \right]. \end{aligned}$$

The linearized Reed-Solomon code defined above is a maximum sum-rank distance code, i.e., it achieves the equivalent of Singleton bound in the sum-rank metric. We will now state a property satisfied by these linearized Reed-Solomon codes.

Theorem 3.3.6. *Consider a linearized Reed-Solomon code with parameters* $(n, k, l_1, \ldots, l_g)$ *with generator matrix* G_{LRS} *as given in the above definition. Consider matrices* $W_1, W_2, \ldots, W_g$ *where* W_i *is of size* $l_i \times n_i$*, then we have that*

$$\textit{rank}(G_{LRS}\ \textit{diag}(W_1, W_2, \ldots, W_g)) \tag{3.3}$$

$$= \min\left(k, \sum_{j=1}^{g} \textit{rank}(W_j)\right). \tag{3.4}$$

There were two approaches considered to construct MRCs based on generator matrices of linearized Reed-Solomon codes - the first is generator matrix based [46] and the second is parity check matrix based [48].

Construction 3.3.7. *We give a construction of the code* $\mathcal{C}$*, which is specified by the following matrix* H*:*

$$H = \begin{bmatrix} H_0 & & & \\ & H_0 & & \\ & & \ddots & \\ & & & H_0 \\ H_1 & H_2 & \ldots & H_{t_1} \end{bmatrix} \quad H_0 = \begin{bmatrix} M_0 & & & \\ & M_0 & & \\ & & \ddots & \\ & & & M_0 \\ M_1 & M_2 & \ldots & M_{t_2} \end{bmatrix}$$

$$M_0 = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_{n_2} \\ \alpha_1^2 & \alpha_2^2 & \dots & \alpha_{n_2}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^\delta & \alpha_2^\delta & \dots & \alpha_{n_2}^\delta \end{bmatrix}$$

$$M_0 = \textit{Vandermonde}(\alpha_1, \alpha_2, \dots, \alpha_{n_2}, 0 : \delta - 1), \quad (3.5)$$

where M_0 is a matrix of size $\delta \times n_2$. We used the terminology Vandermonde(.) to denote a matrix whose columns are identified by $\alpha_1, \alpha_2, \dots, \alpha_{n_2}$ and starting and ending powers are indicated as the last two entries.

$$\Lambda = \textit{Vandermonde}(\alpha_1, \alpha_2, \dots, \alpha_{n_2}, \delta : \delta + h_2 - 1), \quad (3.6)$$

where Λ is a matrix of size $h_2 \times n_2$. Let $\Gamma = (\gamma_1, \gamma_2, \dots, \gamma_{h_2})$ denote the basis of $\mathbb{F}_{q_0^{h_2}}$ over $\mathbb{F}_{q_0}$. Define a set of vectors $\mathcal{B}$ as follows:

$$\mathcal{B} = [\beta_1, \beta_2, \dots, \beta_{n_2}] = [\gamma_1, \gamma_2, \dots, \gamma_{h_2}]\Lambda.$$

Consider γ to be a primitive element of $\mathbb{F}_{q_0^{h_2}}$. For $0 \le j \le t_2 - 1$, let γ^j denote elements belonging to distinct conjugacy classes, which requires that $q_0 \ge t_2 + 1$. The matrices M_i are defined as follows:

$$M_i = D_{q_0}(\gamma^i, \mathcal{B}, h_2, n_2). \quad (3.7)$$

Consider another matrix of size $h_1 \times n_2$ given by the following, where we assume that $h_1 \le r_2 - h_2$

$$\Omega = \textit{Vandermonde}(\alpha_1, \alpha_2, \dots, \alpha_{n_2}, \delta + h_2 : \delta + h_1 + h_2 - 1) \quad (3.8)$$

Let $\Psi = (\psi_1, \psi_2, \dots, \psi_{h_1})$ denote the basis of $\mathbb{F}_{q_1^{h_1}}$ over $\mathbb{F}_{q_1}$. Define a set of vectors Φ as follows:

$$\Phi = [\phi_1, \phi_2, \dots, \phi_{n_2}] = [\psi_1, \psi_2, \dots, \psi_{h_1}]\Omega.$$

Consider λ to be a primitive element of $\mathbb{F}_{q_1^{h_1}}$. For $0 \leq j \leq t_2 t_1$, let λ^j denote elements belonging to distinct conjugacy classes, which requires that $q_1 \geq t_1 t_2 + 1$. The matrices $H_{i,j}$ are defined as follows:

$$H_{i,j} = D_{q_1}(\lambda^{(i,j)}, \Phi, h_1, n_2), \tag{3.9}$$

where $(i, j) = (i - 1)t_2 + j$.

Theorem 3.3.8. *Let $q_0 \geq \max(t_2 + 1, n_2)$ be any prime power and $q_0^{h_2} = q_1 \geq t_1 t_2 + 1$ where $t_1 = \frac{(k+h_1)}{r_1}$ is the number of middle codes and $t_2 = \frac{(r_1+h_2)}{r_2}$ is the number of local codes in each middle code. Then there exists an explicit $[n, k, r_1, r_2, h_1, h_2, \delta]$ HL-MRC over a field of size $O((\max(t_2 + 1, n_2))^{h_2 h_1})$.*

Proof. Let the erasure pattern E be of size $|E| = t_1 t_2 \delta + t_1 h_2 + h_1$ composed of δ erasures in each local code, h_2 additional erasures in each middle code and h_1 additional erasures in the global code. We have to prove that $H|_E$ is full rank, in order to prove that we can recover the erasure pattern E. WLOG assume that h_1 additional erasures occur in middle codes indexed by $1, 2, \ldots, t_1^*$, where $t_1^* \leq h_1$. In each i^{th} middle code, the additional erasures occur in local codes indexed by $1, 2, \ldots, j_i$, where $j_i \leq h_2$. Let $E_{i,j}$ be the set of erasures in the j^{th} local code of the i^{th} middle code. Let $\Delta_{i,j} \subset E_{i,j}$ be an arbitrary subset of size $|\Delta_{i,j}| = \delta$ and $\bar{\Delta}_{i,j} = E_{i,j} \setminus \Delta_{i,j}$. Note that $M_0|_{\Delta_{i,j}}, \forall i \in [t_1^*], j \in [j_i]$ are all $\delta \times \delta$ matrices of full rank. Applying $M_0|_{\Delta_{i,j}}^{-1}$ to column reduce the rest of the columns in $E_{i,j}$, we have

$$\begin{bmatrix} M_0|_{\Delta_{i,j}} & 0 \\ M_j|_{\Delta_{i,j}} & A_{i,j} = M_j|_{\bar{\Delta}_{i,j}} - M_j|_{\Delta_{i,j}} M_0|_{\Delta_{i,j}}^{-1} M_0|_{\bar{\Delta}_{i,j}} \\ H_{i,j}|_{\Delta_{i,j}} & B_{i,j} = H_{i,j}|_{\bar{\Delta}_{i,j}} - H_{i,j}|_{\Delta_{i,j}} M_0|_{\Delta_{i,j}}^{-1} M_0|_{\bar{\Delta}_{i,j}} \end{bmatrix}$$

Note that all the entries in $M_0|_{\Delta_{i,j}}^{-1} M_0|_{\bar{\Delta}_{i,j}}$ are in the base field $\mathbb{F}_{q_0}$. Hence, the column operations on the M_j and $H_{i,j}$ with $\mathbb{F}_{q_0}$ coefficients results in the same structure with β's and ϕ's replaced by their corresponding $\mathbb{F}_{q_0}$ linear combinations and are denoted by β' and ϕ'. Consider the following matrix obtained by stacking $A_{i,j}$ and $B_{i,j}$.

$$C = \left[\begin{array}{ccc|c|ccc} A_{1,1} & \ldots & A_{1,j_1} & \ldots & A_{t_1^*,1} & \ldots & A_{t_1^*,j_{t_1^*}} \\ B_{1,1} & \ldots & B_{1,j_1} & \ldots & B_{t_1^*,1} & \ldots & B_{t_1^*,j_{t_1^*}} \end{array}\right]$$

$$= \left[\begin{array}{c|c|c} \underbrace{A_1}_{h_2\times(h_2+h_{1,1})} & \ldots & \underbrace{A_{t_1^*}}_{h_2\times(h_2+h_{1,t_1^*})} \\ \underbrace{B_1}_{h_2\times(h_2+h_{1,1})} & \ldots & \underbrace{B_{t_1^*}}_{h_2\times(h_2+h_{1,t_1^*}}) \end{array}\right],$$

where $\sum_{i=1}^{t_1^*} h_{1,i} = h_1$. In every i^{th} middle code, pick h_2 number of columns say Υ_i as follows from $\cup_{j=1}^{j_i}\bar{\Delta}_{i,j}$ such that $|\Upsilon_i| = h_2$ and $\bar{\Upsilon}_i = \cup_{j=1}^{j_i}\bar{\Delta}_{i,j} \setminus \Upsilon_i$. WLOG , we pick first h_2 columns from $\cup_{j=1}^{j_i}\bar{\Delta}_{i,j}$. Here, we just state two quick properties:

P1: We note that the matrix $\begin{bmatrix} M_0|_{\Delta_{i,j}} & M_0|_{\Upsilon_i} & M_0|_{\bar{\Upsilon}_i} \\ \mathcal{B}|_{\Delta_{i,j}} & \mathcal{B}|_{\Upsilon_i} & \mathcal{B}|_{\bar{\Upsilon}_i} \\ \Phi|_{\Delta_{i,j}} & \Phi|_{\Upsilon_i} & \Phi|_{\bar{\Upsilon}_i} \end{bmatrix}$ is of Vandermonde structure of size $(\delta + h_2 + h_1) \times n_2$ and hence full rank.

P2: Similarly, the matrix $\begin{bmatrix} M_0|_{\Delta_{i,j}} & M_0|_{\Upsilon_{i,j}} \\ \mathcal{B}|_{\Delta_{i,j}} & \mathcal{B}|_{\Upsilon_{i,j}} \end{bmatrix}$ is also full rank, since it is also a Vandermonde matrix of size $(\delta + h_2) \times (\delta + \Upsilon_{i,j})$, where $\Upsilon_{i,j}$ are the column indices of Υ_i restricted to the j^{th} local code.

We know that in each A_i, there are h_2 columns which are linearly independent. So, the column indices of Υ_i can be either of the following two cases: (a) All $j_i^* \leq j_i$ columns, i.e., $\Upsilon_{i,j} = \bar{\Delta}_{i,j}, \forall j \leq j_i^*$. (b) All columns upto $j < j_i^*$ and some columns of $j = j_i^*$. In this case, $\Upsilon_{i,j} = \bar{\Delta}_{i,j}, \forall j < j_i^*$, $\Upsilon_{i,j} \subset \bar{\Delta}_{i,j}, j = j_i^*$ and $\bar{\Upsilon}_{i,j} = \bar{\Delta}_{i,j} \setminus \Upsilon_{i,j}, j = j_i^*$. Let the first row of

$[A_{i,1}|A_{i,2}|\dots|A_{i,j_i^*}]$ be denoted by D_i. Then, D_i can be written as follows:

$$\begin{aligned} D_i &= \Big[\mathcal{B}|_{\bar{\Delta}_{i,j}} - \mathcal{B}|_{\Delta_{i,j}} M_0|_{\Delta_{i,j}}^{-1} M_0|_{\bar{\Delta}_{i,j}}|\dots| \\ &= (\mathcal{B}|_{\bar{\Delta}_{i,j_1^*}} - \mathcal{B}|_{\Delta_{i,j_1^*}} M_0|_{\Delta_{i,j_1^*}}^{-1} M_0|_{\bar{\Delta}_{i,j}})|_{\Upsilon_{i,j_i^*}}\Big]. \end{aligned}$$

where $\mathcal{B} = [\beta_1, \beta_2, \dots, \beta_{n_2}]$ matrix over $\mathbb{F}_{q_0}$. So column operations on the , $D_i = [D_{i,1}|D_{i,2}|\dots|D_{i,j_i^*}]$, since each $D_{i,j}$ in D_i is full rank from P2. It follows that each matrix $A_{i,j}|_{\Upsilon_{i,j}}$ is full rank $\forall j \le j_i^*$. Applying Theorem 3.3.6, we have that $\text{rank}(A_i|_{\Upsilon_i}) = \sum_{j=1}^{j_i^*} \text{rank}(A_{i,j}|_{\Upsilon_{i,j}}) = h_2$. Now, we can write $C = \begin{bmatrix} A_i|_{\Upsilon_i} & A_i|_{\bar{\Upsilon}_i} \\ B_i|_{\Upsilon_i} & B_i|_{\bar{\Upsilon}_i} \end{bmatrix}$. We can use the columns of $A_i|_{\Upsilon_i}$ to reduce the other columns and after column reduction, we get the following matrix:

$$\begin{bmatrix} A_i|_{\Upsilon_i} & 0 \\ B_i|_{\Upsilon_i} & B_i|_{\bar{\Upsilon}_i} - B_i|_{\Upsilon_i} A_i|_{\Upsilon_i}^{-1} A_i|_{\Upsilon_i} \end{bmatrix} = \begin{bmatrix} A_i & 0 \\ B_i & F_i \end{bmatrix}.$$

Let $F = [F_1|F_2|\dots|F_{t_1^*}]$. Recall that the first row of elements in $B_i|_{\bar{\Upsilon}_i}$ as follows:

$$\begin{aligned} &\Big[\Big(\Phi|_{\bar{\Delta}_{i,j_i^*}} - \Phi|_{\Delta_{i,j_i^*}} M_0|_{\Delta_{i,j_i^*}}^{-1} M_0|_{\bar{\Delta}_{i,j_i^*}}\Big)|_{\bar{\Upsilon}_{i,j_i^*}}|\dots| \\ &\Phi|_{\bar{\Delta}_{i,j_i}} - \Phi|_{\Delta_{i,j_i}} M_0|_{\Delta_{i,j_i^*}}^{-1} M_0|_{\bar{\Delta}_{i,j_i}}\Big] \\ = &\ [\Theta_{i,j_i^*}|\Theta_{i,j_i^*+1}|\dots|\Theta_{i,j_i}] = \Theta|_{\bar{\Upsilon}_i} \end{aligned}$$

From the above, we can calculate that the first row of elements of F_i are given by $G_i = [\Theta|_{\bar{\Upsilon}_i} - \Theta|_{\Upsilon_i} A_i|_{\Upsilon_i}^{-1} A_i|_{\Upsilon_i}]$. Recall that ϕ''s retains its structure as ϕ. Let ϕ'' are the elements obtained after column operations correspond to the ϕ', which are also retains the same structure. Therefore each G_i is full rank from P1. It follows that each matrix F_i is full rank $\forall i \le t_1^*$. Applying Theorem 3.3.6, we have that $\text{rank}(F) = \sum_{j=1}^{t_i^*} \text{rank}(F_j) = h_1$. Hence, the erasure pattern E can be recovered and the theorem follows. □

Note 2. *The field size of HL-MRC required for the above construction given by $O((\max(t_2 + 1, n_2))^{h_2 h_1})$ is lower than the field size required for the construction in [58] ($\ge n^{\delta h_1 h_2}$). This is* *because the construction in [58] is based on the generator matrices of the Gabidulin codes as opposed to those of linearized Reed-Solomon codes used.*

3.3.4 Random Construction of HL-MRC

In this section, we redefine HL-MRC in terms of the generator matrix and identify the standard form of the generator matrix. We use this notion for the random construction of HL-MRC and subsequently we provide a lower bound on the field size.

Theorem 3.3.9 (HL-MRC standard form)**.** *Let $\mathscr{C}[n,k,r_1,r_2,h_1,h_2,\delta]$ HL-MRC over a finite field q. Then $\mathscr{C}$ has a generator matrix of the form*

$$G = [B_1|B_2|\dots|B_{t_1}] \in F_q^{k\times n}, \tag{3.10}$$

where $B_i = \Big[B_{i1}|B_{i2}|\dots|B_{it_2}\Big] \in F_q^{k\times n_1}$, $\forall i \in [t_1]$. Here $B_{ij} = [E_{ij}|F_{ij}] \in F_q^{k\times n_2}$, $E_i = [E_{i1}|\dots|E_{it_2}] \in F_q^{k\times t_2 r_2}$ and $F_i = [F_{i1}|\dots|F_{it_2}] \in F_q^{k\times t_2\delta}$ and $E_i = [C_i|D_i]$, $C_i \in F_q^{k\times r_1}$ and $D_i \in F_q^{k\times h_2}$ and the submatrix $G_C = [C_1|\dots|C_{t_1}]$ is of the form $G_C = [I_k|A]$, with $A_{k\times h_1}$ being superregular.

Proof. The proof technique is similar to the Theorem 15 in [59]. Let $\tilde{G}$ be the generator matrix of any HL-MRC in the form of equation (3.10)

$$\tilde{G} = [\tilde{B}_1|\tilde{B}_2|\dots|\tilde{B}_{t_1}] \in F_q^{k\times n},$$

where $\tilde{B}_i = [\tilde{B}_{i1}|\tilde{B}_{i2}|\dots|\tilde{B}_{it_2}] \in F_q^{k\times n_1}$, $\forall i \in [t_1]$, $\tilde{B}_{ij} = [\tilde{E}_{ij}|\tilde{F}_{ij}] \in F_q^{k\times n_2}$, $\tilde{E}_i = [\tilde{E}_{i1}|\dots|\tilde{E}_{it_2}] \in F_q^{k\times t_2 r_2}$ and $\tilde{F}_i = [\tilde{F}_{i1}|\dots|\tilde{F}_{it_2}] \in F_q^{k\times t_2\delta}$. Here $\tilde{E}_i = [\tilde{C}_i|\tilde{D}_i]$, $\tilde{C}_i \in F_q^{k\times r_1}$ and $\tilde{D}_i \in F_q^{k\times h_2}$.

By puncturing the last δ columns of each $\tilde{B}_{ij}$ and last h_2 columns of each $\tilde{E}_i$, we obtain $\tilde{G}_C = [\tilde{C}_1|\dots|\tilde{C}_{t_1}]$. Suppose the obtained matrix $\tilde{G}_C$ is not in the systematic form. Given the definition of HL-MRC, it is evident that $\tilde{G}_C$ is the generator matrix of $[k+h_1,k]$ MDS code. Therefore, there exists an invertible matrix P such that $P\tilde{G}_C$ will be in systematic form. Thus, the matrix $G = P\tilde{G}_C$ is the generator matrix in the required systematic form for the code $\mathscr{C}$.

□

Proposition 3.3.1. *A matrix $G \in F_q^{k\times n}$ generates $\mathscr{C}[n,k,r_1,r_2,h_1,h_2,\delta]$ HL-MRC code if and only if the determinant of matrices belongs to T_{k,r_1,r_2} are non-zero, here*

$$T_{k,r_1,r_2} = \left\{ S \in F_q^{k\times k} \;\middle|\; \begin{array}{l} S \text{ is a submatrix of } G \text{ with} \\ \text{at most } r_2 \text{ columns per } B_{ij}^{th} \\ \text{block and at most } r_1 \text{ columns} \\ \text{per } B_i^{th} \text{ block,.} \end{array} \right\}.$$

We now consider the entries $a_{v,w}$ for $v = [k], w = [h_1]$. We know that the column space of D_i is inside the column space of C_i by the parameters of the block MDS code and column space of F_{ij} is inside the column space of E_{ij}, by the parameters of the block Local MRC. If we denote by $C_i^{(l)}$, $D_i^{(l)}$, $E_{ij}^{(l)}$ and $F_{ij}^{(l)}$ the l^{th} column of C_i, D_i, E_{ij} and F_{ij} respectively, then $D_i^{(l)} = \sum_{t=1}^{r_1} y_{t,i,l} C_i^{(t)}$ and $F_{ij}^{(l)} = \sum_{t'=1}^{r_2} z_{t',ij,l} E_{ij}^{(t')}$, for some $y_{t,i,l}$ and $z_{t',ij,l}$ which we also consider variables. This way we can consider a $k \times n$ generator matrix in the variable form by $G(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ as a matrix in $F_q[x_{v,w}, y_{t,i,l}, z_{t',ij,l}]^{k\times n}$.

By imposing the condition of HL-MRC using 3.3.1 on the variable form $G(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$, we obtain the polynomial

$$p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \operatorname{lcm}\{\det S \mid S \in T_{k,r_1,r_2}(G(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}))\} \tag{3.11}$$

which belongs to $F_q[x_{v,w}, y_{t,i,l}, z_{t',ij,l}]$, then we have that $G(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ generates $\mathscr{C}[n,k,r_1,r_2,h_1,h_2,\delta]$ HL-MRC over F_q if and only if $p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ is non zero. By selecting random vector $\boldsymbol{\alpha} \in F_q^{h_1 k}$, $\boldsymbol{\beta} \in F_q^{R_1 r_1}$ and $\boldsymbol{\gamma} \in F_q^{R_2 r_2}$, we can obtain the generator matrix $G(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ as described above, here $R_1 = t_1 h_2$ and $R_2 = t_1 t_2 \delta$. If the corresponding polynomial $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$, defined in (3.11), is non-zero, then the row space of $G(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ is a $\mathscr{C}[n,k,r_1,r_2,h_1,h_2,\delta]$ HL-MRC.

Note 3. *Existence of such $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \in F_q^{h_1 k} \times F_q^{R_1 r_1} \times F_q^{R_2 r_2}$ such that $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \neq 0$ can be verified using Proposition 3.3.1 as done in [59].*

Lower bound on the field size of Random HL-MRC:

We now derive the lower bound on the field size using the Schwartz-Zippel Lemma, for which we need to find the upper bound on the degree of $p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$.

Lemma 3.3.10. *The total degree of the polynomial* $p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$*, defined as in (3.11), satisfies the inequality*

$$deg\, p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \leq \sum_{j_0=0}^{k} \binom{k}{j_0} L. \tag{3.12}$$

where, $j_2^{'} = k - j_0 - j_1$ *and*$j_3 = k - j_0 - j_1 - j_2$ *in the following equation*

$$L = \begin{cases} \sum_{j_1=0}^{h_1} \sum_{j_2=0}^{j_2^{'}} (j_1 + 2j_2 + 3j_3) \binom{h_1}{j_1} \binom{R_1}{j_2} \binom{R_2}{j_3} & \textit{if } j_3 \leq R_2 \\ 0 & \textit{otherwise} \end{cases}$$

Proof. Let $\mathcal{M}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ is the set of all possible $k \times k$ submatrices of $G(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$. Hence $T_{k,r_1,r_2}(G(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}))$ is a subset of $\mathcal{M}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$, the polynomial $p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ divides the polynomial

$$q(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \operatorname{lcm}\{\det S \mid S \in \mathcal{M}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}).\}$$

Therefore $\deg p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \leq \deg q(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$. Now, consider the generator matrix $G(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ as defined in Theorem 3.3.9. Observe that the entries of the first k columns of the submatrix G_C have degree 0, h_1 columns of the matrix A have degree atmost 1, h_2 columns of the E_i for every i have degree atmost 2 and δ columns of the matrix F_{ij} for every i and j has degree at most 3. So, we can partition n columns of the generator matrix in to four groups of degree atmost 0,1,2 and 3. Therefore

$$\deg q(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \leq \sum_{S \in \mathcal{M}_k(G)} \deg \det\, S \leq \sum_{j_0=0}^{k} \binom{k}{j_0} L$$

The last inequality holds depending upon the number of ways of selecting k columns from the different partitions of n columns.

□

Theorem 3.3.11. *Let the entries of $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ be uniformly and independently chosen at random in F_q. Then*

$$Pr\{\text{ row space } (G(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})) \text{ is HL-MRC}\} \geq 1 - \frac{\sum_{j_0=0}^{k} \binom{k}{j_0} L}{q}$$

Proof.

$$\begin{aligned} LHS = 1 - \Pr\{p(\alpha, \beta, \gamma) = 0\} &\geq 1 - \frac{\deg p(x, y, z)}{q} \\ &\geq 1 - \frac{\sum_{j_0=0}^{k} \binom{k}{j_0} L}{q} \end{aligned}$$

where the last two inequalities follow from Schwartz-Zippel Lemma and Lemma 3.3.10.

□

Corollary 1. *If $q > \sum_{j_0=0}^{k} \binom{k}{j_0} L$ then there exists an $\mathscr{C}[n, k, r_1, r_2, h_1, h_2, \delta]$ HL-MRC over the finite field F_q.*

This bound is quantitatively similar to the known existence result of MRC of any locality configuration given in [60] Theorem 1, i.e. $q > \binom{n-1}{k-1}$, and also random construction of Local MRC given in [59] Corollary 25, i.e. $q > 2(n-k)\binom{n-1}{k-1}$.

However, considering step-by-step construction as done for Local MRC in [59] can be extended to HL-MRC to obtain the better result.

Example 2. *Construction of $[n = 18, k = 5, r_1 = 3, r_2 = 2, h_1 = 1, h_2 = 1, \delta = 1]$ HL-MRC code and determine the required field size in the process of step by step construction.*

We start with $[6, 1]$ MDS code with all ones parity column, let G^* is the corresponding generator matrix. We divide the columns of the generator matrix G^* in to two groups of each $r_1 = 3$ columns.

- *By adding parity column ($h_2 = 1$) in variable form, we get set of determinant polynomials with all possible $k \times k$ submatrix of the obtained generator matrix in variable form.*

- *By imposing the condition on lcm of set of polynomials not equal to zero, we can find the values for the variables.*

We will repeat this process step by set to obtain the generator matrix $G^{'}$ for the code $\mathcal{C}^{'}[n^{'} = 8, k = 5, h_1 = 1, r_1 = 3, h_2 = 1]$. Again we divide the columns of $G^{'}$ in to 4 groups of each $r_2 = 2$ columns. By repeating the same process we can obtain the local parity ($\delta = 1$) of the four local codes one by one.

Table 3.2: Lcm of determinant polynomials of parity columns and field size

parity column	***LCM of set of det polynomials***	***Values of variables***	q
$D_1^{(1)}$	$(y_{1,1,1})(y_2,1,1)(y_{3,1,1})$ $(y_{1,1,1} - y_{2,1,1})$ $(y_{1,1,1} - y_{3,1,1})$ $(y_{2,1,1} - y_{3,1,1})$	$y_{1,1,1} = 1$ $y_{2,1,1} = 2$ $y_{3,1,1} = 3$	*5*
$D_2^{(1)}$	$6(y_{1,2,1})(y_{2,2,1})(y_{3,2,1})$ $(y_{2,2,1} - y_{1,2,1})$ $(y_{1,2,1} + y_{3,2,1})$ $(y_{1,2,1} + y_{3,2,1})$	$y_{1,2,1} = 2$ $y_{2,2,1} = 3$ $y_{3,2,1} = 1$	*5*
$F_{11}^{(1)}$	$36(z_{1,11,1})(z_2,11,1)$ $(z_{2,11,1} - 2z_{1,11,1})$ $(2z_{2,11,1} - z_{1,11,1})$	$z_{1,11,1} = 1$ $z_{2,11,1} = 4$	*5*
$F_{12}^{(1)}$	$48(z_{1,12,1})(z_{2,12,1})$ $(z_{2,12,1} + z_{1,12,1})$ $(7z_{2,12,1} + 3z_{1,12,1})$ $(3z_{2,12,1} + z_{1,12,1})$	$z_{1,12,1} = 1$ $z_{2,12,1} = 1$	*7*
$F_{21}^{(1)}$	$(240)(7)(z_{1,21,1})(z_{2,21,1})$ $(z_{1,21,1} - z_{2,21,1})$ $(3z_{2,21,1} - 4z_{1,21,1})$ $(2z_{2,21,1} - 3z_{1,21,1})$	$z_{1,21,1} = 1$ $z_{2,21,1} = 2$	*11*
$F_{22}^{(1)}$	$(240)(7)(z_{1,22,1})(z_{2,22,1})$ $(z_{1,22,1} + z_{2,22,1})$ $(z_{1,22,1} + 4z_{2,22,1})$ $(z_{1,22,1} + 3z_{2,22,1})$ $(z_{1,22,1} + 2z_{2,22,1})$	$z_{1,22,1} = 1$ $z_{2,22,1} = 1$	*11*

Table 3.2 consists of set of lcm of polynomials obtained while computing the generator matrix $G = [B_{11}|B_{12}|B_{21}|B_{22}]$ *of the code for the given example. We Obtained*

$$B_{11}^{\top} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 & 0 \end{bmatrix}, \; B_{12}^{\top} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 2 & 3 & 0 & 0 \\ 1 & 2 & 4 & 0 & 0 \end{bmatrix},$$

$$B_{21}^{\top} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix}, \; B_{22}^{\top} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 3 & 4 \\ 2 & 2 & 2 & 3 & 4 \end{bmatrix},$$

which generates an $[n = 18, k = 5, r_1 = 3, r_2 = 2, h_1 = 1, h_2 = 1, \delta = 1]$ *HL-MRC code over* F_{11}.

3.4 Lower Bound on the Field Size of HL-MRC

In this section, we will derive lower bounds on the field size of HL-MRC. The proof technique is similar to the one developed in [56], with the difference being that in this case, there are mid-level codes as well and hence while performing shortening in the parity check matrix, this has to be taken into account. The following Lemma derived in [56] will be useful in deriving the lower bounds on field size.

Lemma 3.4.1 ([56]). *Let $X_1, X_2, \ldots, X_g \in \mathbb{P}^d(\mathbb{F}_q)$ be mutually disjoint subsets each of size t with $g \geq d+1$ of the projective space $\mathbb{P}^d(\mathbb{F}_q)$* [2]*. If $q < \left(\frac{g}{d} - 1\right) t - 4$, then there exists a hyperplane which intersects $d+1$ distinct subsets among $X_1, X_2, \ldots, X_g$.*

Theorem 3.4.2. *Consider an $[k, r_1, r_2, h_1, h_2, \delta]$ HL-MRC. If $(\delta + 2) \leq h_1 + h_2$, $h_1 \leq \frac{n}{n_1}$ and $h_2 \leq \frac{n_1}{n_2} - 1$, then the field size q is lower bounded as follows:*

$$q \geq \left(\frac{\frac{n}{n_2}}{h_1 h_2 + h_1 - 1} - 1 \right) \binom{r_2 + \delta}{\delta + 1} - 4. \tag{3.13}$$

Proof. Consider an arbitrary $[k, r_1, r_2, h_1, h_2, \delta]$ HL-MRC with $t_1 = \frac{n}{n_1}$ mid-level codes and $t_2 = \frac{n_1}{n_2}$ local codes per mid-level code. The code has a parity-check matrix of the form

$$H = \begin{bmatrix} M_1 & & \\ & \ddots & \\ & & M_{t_1} \\ P_1 & \ldots & P_{t_1} \end{bmatrix}, \quad M_i = \begin{bmatrix} M_{i,1} & & \\ & \ddots & \\ & & M_{i,t_2} \\ N_{i,1} & \ldots & N_{i,t_2} \end{bmatrix},$$

where $M_1, \ldots, M_{t_1}$ are $\delta' \times (r_1 + \delta')$ matrices over $\mathbb{F}_q$, where $\delta' = h_2 + \left(\frac{r_1 + h_2}{r_2}\right)\delta$. $P_1, \ldots, P_{t_1}$ are $h_1 \times (r_1 + \delta')$ matrices over $\mathbb{F}_q$. $M_{i,1}, \ldots, M_{i,t_2}$ are $\delta \times (r_2 + \delta)$ matrices over $\mathbb{F}_q$. $N_{i,1}, \ldots, N_{i,t_2}$ are $h_2 \times (r_2 + \delta)$ matrices over $\mathbb{F}_q$. For every subset $S_{i,j} \subseteq [r_2 + \delta]$ of size

[2] $\mathbb{P}^d(\mathbb{F}_q)$ is the set of all one dimensional spaces in $\mathbb{F}_q^{d+1}$

$|S_{i,j}| = \delta + 1$, $M_{i,j}(S_{i,j})$ is a $\delta \times (\delta + 1)$ full rank matrix. Let $M_{i,j}(S_{i,j})^{\perp} \in \mathbb{F}_q^{\delta+1}$ be a nonzero vector orthogonal to the row space of $M_{i,j}(S_{i,j})$. We know that $M_{i,j}(S_{i,j})M_{i,j}(S_{i,j})^{\perp} = 0$. We denote $q_{i,j}^2(S_{i,j}) = N_{i,j}(S_{i,j})M_{i,j}(S_{i,j})^{\perp}$ where $q_{i,j}^2(S_{i,j})$ is a $h_2 \times 1$ vector and $q_{i,j}^1(S_{i,j}) = P_{i,j}(S_{i,j})M_{i,j}(S_{i,j})^{\perp}$ where $q_{i,j}^1(S_{i,j})$ is a $h_1 \times 1$ vector.

We now consider the maximal erasure pattern, which has $h_1 + t_1 h_2 + t_1 t_2 \delta$ erasures and is correctable. We assume that there are $h_1(h_2 + 1)$ local codes in which there are $\delta + 1$ erasures in each. WLOG, we assume that the indices of the local codes are given by $(i, j), 1 \leq i \leq h_1$, $1 \leq j \leq h_2 + 1$. For every $1 \leq i \leq h_1$, $1 \leq j \leq h_2 + 1$, consider $S_{i,j} \subseteq [r_2 + \delta]$ of size $|S_{i,j}| = \delta + 1$, let $S_i = \cup_{j=1}^{h_2+1} S_{i,j}$ and $S = \cup_{i=1}^{h_1} S_i$. Since the code is HL-MRC, $H|_S$ is full rank.

$M_i(S_i)^{\perp}$ is defined as follows:

$$M_i(S_i)^{\perp} = \begin{bmatrix} M_{i,1}(S_{i,1})^{\perp} & & & \\ & M_{i,2}(S_{i,2})^{\perp} & & \\ & & \ddots & \\ & & & M_{i,h_3}(S_{i,h_3})^{\perp} \end{bmatrix}.$$

where $h_3 = h_2 + 1$. Let

$$H|_S = D = \begin{bmatrix} M_1(S_1) & & & \\ & M_2(S_2) & & \\ & & \ddots & \\ & & & M_{h_1}(S_{h_1}) \\ P_1(S_1) & P_2(S_2) & \dots & P_{h_1}(S_{h_1}) \end{bmatrix}.$$

We form the matrix

$$Q = D \, diag(M_1(S_1)^{\perp}, M_2(S_2)^{\perp}, \dots, M_{t_1}(S_{h_1})^{\perp}).$$

After removing the zero rows in the Q matrix, the structure of the resulting matrix is Q' as mentioned in (3.14):

$$\begin{aligned} Q' &= [q_{1,1}(S_{1,1})\ q_{1,2}(S_{1,2}) \dots\ q_{h_1,h_3}(S_{h_1,h_3})] \\ &= \begin{bmatrix} q^2_{1,1}(S_{1,1}) & \dots & q^2_{1,h_3}(S_{1,h_3}) & & & & \\ & & \ddots & & & & \\ & & & & & & \\ & & & \dots & q^2_{h_1,1}(S_{h_1,1}) & \dots & q^2_{h_1,h_3}(S_{h_1,h_3}) \\ q^1_{1,1}(S_{1,1}) & \dots & q^1_{1,h_3}(S_{1,h_3}) & \dots & q^1_{h_1,1}(S_{h_1,1}) & \dots & q^1_{h_1,h_3}(S_{h_1,h_3}) \end{bmatrix} \end{aligned} \tag{3.14}$$

Recall that D is full rank, and also notice that $diag(M_1(S_1)^\perp, M_2(S_2)^\perp, \dots, M_{t_1}(S_{h_1})^\perp)$ is full rank ($\because$ $M_i(S_i)$'s are full rank). Therefore, $Q^{'}$ is full rank. From this, we can state the following Lemma, which is used to prove the current theorem (Theorem 3.4.2).

Lemma 3.4.3. *In $[k, r_1, r_2, h_1, h_2, \delta]$ HL-MRC, for any $I \subset [t_1]$, $J \subset [t_2]$ with $|I| = h_1$, $|J| = h_2 + 1$ and subsets $S_{i,j} \subseteq (r_2 + \delta)$ of size $\delta + 1$ for all $i \in I \,\&\, \forall j \in J$. Then $(h_1h_2 + h_1) \times (h_1h_2 + h_1)$ the matrix Q' is full rank.*

From the Lemma 3.4.3, we can claim that $q_{i,j}(S_{i,j})$ and $q_{i',j'}(T_{i',j'})$ are not multiples of each other unless either $i = i^{'}$ and $j = j^{'}$. Now, we show that even within a local group, these vectors are distinct.

Lemma 3.4.4. *For every $i, j \in [t_1, t_2]$, no two vectors in $\{q_{i,j}(S_{i,j}) : S \subseteq \binom{[r_2+\delta]}{\delta+1}\}$ are multiples of each other.*

Proof. Suppose $q_{i,j}(S_{i,j}) = \lambda q_{i,j}(T_{i,j})$ for some distinct $S_{i,j}, T_{i,j} \subseteq [r_2 + \delta]$ of size $\delta + 1$ each and some nonzero $\lambda \in \mathbb{F}_q$.

$$\begin{bmatrix} M_{i,j}(S_{i,j}) \\ N_{i,j}(S_{i,j}) \\ P_{i,j}(S_{i,j}) \end{bmatrix} M_{i,j}(S_{i,j})^{\perp} - \lambda \begin{bmatrix} M_{i,j}(T_{i,j}) \\ N_{i,j}(T_{i,j}) \\ P_{i,j}(T) \end{bmatrix} M_{i,j}(T_{i,j})^{\perp}$$
$$\begin{bmatrix} 0 \\ q^2_{i,j}(S_{i,j}) \\ q^1_{i,j}(S_{i,j}) \end{bmatrix} - \lambda \begin{bmatrix} 0 \\ q^2_{i,j}(T_{i,j}) \\ q^1_{i,j}(T_{i,j}) \end{bmatrix} = 0.$$

Note that every coordinate of $M_{i,j}(S_{i,j})^{\perp}$ is nonzero. Otherwise, it implies a linear dependence among δ columns of $M_{i,j}(S_{i,j})$. Thus, we have a linear combination of $\begin{bmatrix} M_{i,j}(S_{i,j} \cup T_{i,j}) \\ N_{i,j}(S_{i,j} \cup T_{i,j}) \\ P_{i,j}(S_{i,j} \cup T_{i,j}) \end{bmatrix}$. However, $|S_{i,j} \cup T_{i,j}| \leq 2\delta + 2 \leq \delta + h_1 + h_2$. By the MR property, any set of columns of the matrix $\begin{bmatrix} M_{i,j} \\ N_{i,j} \\ P_{i,j} \end{bmatrix}$ of size $\delta + h_1 + h_2$ has to be full rank. Thus, we arrive at a contradiction. Thus, no two vectors in $\{q_{i,j}(S_{i,j}) : S_{i,j} \subseteq \binom{[r_2+\delta]}{\delta+1}\}$ are multiples of each other. $\square$

We utilize Lemma 3.4.3 and Lemma 3.4.4 to complete our proof of the current theorem (Theorem 3.4.2). Now, define set $X_{i,j}$ as follows: $X_{i,j} = \{q_{i,j}(S_{i,j}) : S_{i,j} \subseteq \binom{[r_2+\delta]}{\delta+1}\}$. The sets $\{X_{i,j} : i \in [t_1], j \in [t_2]\}$ are all mutually disjoint (from Lemma 3.4.3) and $|X_{i,j}| = \binom{[r_2+\delta]}{\delta+1}$ (from Lemma 3.4.4). we can think of $\{q_{i,j}(S_{i,j}) : S_{i,j} \subseteq \binom{[r_2+\delta]}{\delta+1}\}$ as distinct points in $\mathbb{P}^{(h_1 h_2 + h_1 - 1)}(\mathbb{F}_q)$. Since $t_1 \geq h_1$ and $t_2 \geq h_2 + 1$, it follows that $t_1 t_2 \geq h_1 h_2 + h_1$. Based on Lemma 3.4.4, there is no hyperplane in $\mathbb{P}^{(h_1 h_2 + h_1 - 1)}(\mathbb{F}_q)$ which contains $h_1 h_2 + h_1$ points from distinct subsets of $\{X_{i,j} : i \in [t_1], j \in [t_2]\}$. By Lemma 3.4.1, we have a lower bound on the field size.

□

Theorem 3.4.5. *Consider an* $[k, r_1, r_2, h_1, h_2, \delta]$ *HL-MRC.*

(a) If $4 \leq h_1 + h_2 \leq (\delta + 2)$*,* $h_1 \leq \frac{n}{n_1}$ *and* $h_2 \leq \frac{n_1}{n_2} - 1$*, then the field size* q *is lower bounded as follows:*

$$q \geq \left(\frac{\frac{n}{n_2}}{h_1 h_2 + h_1 - 1} - 1 \right) \binom{r_2 + h_1 + h_2 - 2}{h_1 + h_2 - 1} - 4. \tag{3.15}$$

(b) If $(\delta + 2) \leq h_1 + h_2$*,* $h_1 > \frac{n}{n_1}$ *and* $h_2 \leq \frac{n_1}{n_2} - \lceil \frac{h_1}{t_1} \rceil$*, then the field size* q *is lower bounded as follows:*

$$q \geq \left(\frac{\frac{n}{n_2}}{\frac{n}{n_1} h_2 + h_1 - 1} - 1 \right) \binom{r_2 + \delta}{\delta + 1} - 4. \tag{3.16}$$

Proof. We will give the proof sketch by highlighting the differences between the proof of Theorem 3.4.2 and these cases.

- For proving part (a) of the theorem, we do not take arbitrary S and T as in the case of proof of Theorem 3.4.2 but consider subsets S that have size $\delta + 1$ but constrained to contain the subset $\{1, 2, \ldots, (\delta + 2 - h_1 - h_2)\}$. By picking the sets in this way, we still ensure that the pairwise unions have size at most $\delta + h_1 + h_2$. The total number of such sets is given by $\binom{r_2+h_1+h_2-2}{h_1+h_2-1}$. Based on this counting, the field size bound in (a) follows.

- For proving part (b) of the theorem, let $f_1 \geq f_2 \geq \ldots \geq f_{t_1}$ be such that $f_i = \lceil \frac{h_1}{t_1} \rceil$ or $\lfloor \frac{h_1}{t_1} \rfloor$ and $\sum_{i=1}^{t_1} f_i = h_1$. From each i^{th} middle code, pick $h_2 + f_i$ local codes and $\delta + 1$ columns from every local code. By applying the row reduction similar to the proof of Theorem 3.4.2 and applying appropriately modified versions of Lemmas 3.4.3 and 3.4.4, the result follows.

□

Note 4. *Please note that we derive the field size bounds for the cases when (i)* $(\delta + 2) \leq h_1 + h_2$*,* $h_1 \leq \frac{n}{n_1}$ *and* $h_2 \leq \frac{n_1}{n_2} - 1$ *(ii)* $4 \leq h_1 + h_2 \leq (\delta + 2)$*,* $h_1 \leq \frac{n}{n_1}$ *and* $h_2 \leq \frac{n_1}{n_2} - 1$ *(ii)*

$(\delta+2) \leq h_1 + h_2$, $h_1 > \frac{n}{n_1}$ *and* $h_2 \leq \frac{n_1}{n_2} - \lceil \frac{h_1}{t_1} \rceil$. *There are other cases of parameters* δ, h_1, h_2 *for which field-size bounds need to be derived and their proofs are similar to those discussed above.*

Comparative Analysis of Field Size Bounds of LRCs, MRCs and HL-MRCs We will compare the techniques used for deriving field size bounds in the papers mentioned below with those in the present thesis.

- In [19], dimensional locality is defined for (r, δ) locality and is the dimension of the repair set. The dimension locality differs from the conventional notion of locality when the repair sets are non-MDS. Alphabet-dependent bounds for codes with dimension locality have been derived in [37]. These bounds are obtained on the maximum possible dimension of the code, given the length, field size and dimension locality. The bound itself is derived by puncturing an LRC based on the repair sets and applying the Singleton bound to the shortened code which remained after few iterations of puncturing codes. The dimensional locality is used to find the Griesmer bound in order to find length of the shortened code. Note that the bounds is derived assuming general repair sets, which could be potentially intersecting too. In our case of MRCs and HL-MRCs, we have disjoint repair sets and also local codes are MDS. Thus, the notion of dimension locality reduces to that of conventional locality. Also the derivation of field size for an MRC has to take into account the condition that there are many ways of puncturing the code to result in an MDS code. For an LRC, only the puncturing patterns defined by the repair sets can be applied. Thus, the lower bound on field size for an MRC is in general much higher than that for an LRC.

- In [36], upper bound on the dimension of codes with hierarchical locality have been derived (dimensional locality is not considered here). We would like to note that in hierarchical local codes, only a limited number of repair sets are present to ensure locality of the code. Whereas in the case of HL-MRCs, all the erasure patterns have to be corrected, which is a much stronger condition. Thus, the lower bound on field size for a HL-MRC is in general much higher than that for a hierarchical local code.

- To make the field size comparison with [19] and [36] more concrete, we would like to note that alphabet-dependent bounds in [19] and [36] are tighter than the Singleton-like bound. The field size required to construct codes which are optimal with respect to Singleton-like bound is $\Theta(n)$. Hence, the lower bound on the field size resulting from the alphabet-dependent bound will be less than $\Theta(n)$. The field size lower bound on the HL-MRC is given by $q \geq \Omega(\frac{n}{n_2} r_2^{\min(\delta+1, h_1+h_2-1)})$. This is polynomial in n, if r_2 also increases as n^c for some c.

- In [51], MRCs for $\delta = 2$ are considered and by using matroid theory, the list of all possible minors have been derived. These minors have been in turn used to derive non-asymptotic lower bound on the field size of MRC, better than the naive bound known earlier. Using similar arguments, we believe that we can derive the list of all possible minors of HL-MRCs too (since all the correctable erasure patterns are precisely characterized). This can give improved non-asymptotic bounds for HL-MRCs. We leave it for future work.

- In [15], under the assumptions of disjoint repair sets for LRCs and for $d_{\min} \geq 5$ upper bound on the length of the code for an optimal LRCs of sufficiently large length n over an alphabet size q is given by when $4 \nmid d$, $n \leq O(dq^3)$. Note that this bound is also applicable only for LRCs and loose for MRCs.

3.5 Conclusion

In this chapter, we have presented explicit constructions of HL-MRCs: i) For $h_1 = 1, h_2 = 1$ with the field size is $O(n_1)$ and ii) For $h_1 = 2, h_2 = 1$ with the field size is $O(n^4)$. We have also provided the construction of HL-MRCs using the parity check matrix approach, employing Linearized Reed-Solomon Codes with the field size equal to $O((\max(t_2{+}1, n_2))^{h_2 h_1})$. The random construction approach utilizing the generator matrix is obtained with $q > \sum_{j_0=0}^{k} \binom{k}{j_0} L$. Furthermore, we have provided lower bounds on the field size required for HL-MRCs to achieve the desired properties of maximum recoverability and hierarchical locality. We have also compared these bounds with those derived from other papers.

Chapter 4

Maximally Recoverable Product Codes

In this chapter, we study Maximally Recoverable Product Codes (MRPCs) that are explicitly designed for product topologies $T_{m,n}(a, b, 0)$. Product codes are a class of codes which have generator matrices as the tensor product of the component codes and the codeword itself can be represented as an $(m \times n)$ array, where the component codes themselves are referred to as the row and column codes. MRPCs are a class of codes which can recover from all information theoretically recoverable erasure patterns, given the a column and b row constraints imposed by the code.

In this work, we derive puncturing and shortening properties of maximally recoverable product codes. We give a sufficient condition to characterize a certain subclass of erasure patterns as correctable and another necessary condition to characterize another subclass of erasure patterns as not correctable.

We construct a certain bipartite graph based on the erasure pattern satisfying the regularity condition for product topology (any $a, b,\ h = 0$) and show that there exists a complete matching in this graph. We then present an alternate direct proof of the sufficient condition for recoverability of an erasure pattern when $a = 1$. We later extend our technique to study the topology for $a = 2$, and characterize a subset of recoverable erasure patterns in that case. For both $a = 1, 2$, our method of proof is uniform, i.e., by constructing tensor product $G_{\text{col}} \otimes G_{\text{row}}$ of generator matrices of column and row codes such that certain square sub-matrices retain full rank. The full-rank condition is proved by resorting to the matching identified earlier.

In an earlier work, higher order MDS codes denoted by MDS(l) have been defined in terms of generic matrices and these codes have been shown to be constituent row codes for maximally recoverable product codes for the case of $a = 1$. We derive a certain inclusion-exclusion type principle for characterizing the dimension of intersection spaces of generic matrices. Applying this, we formally derive a relation between MDS(3) codes and points/lines of the associated projective space.

4.1 Organization of the Chapter

The rest of the chapter is organized as follows: Section 4.2 presents the derivation of several properties related to MRPCs. Section 4.3 focuses on providing a certain class of correctable and non-correctable erasure patterns. Additionally, it presents a subclass of recoverable erasure patterns specifically for the case when $a = 2$. Section 4.4 introduces generic matrices and higher-order MDS codes, along with discussing certain properties associated with them. Lastly, Section 4.5 concludes the chapter

4.2 Properties of MRPCs

In this section, we show that puncturing on the rows of an MRPC results in an MRPC. We also prove that shortening on the rows of an MRPC also results in an MRPC. Finally, we derive that transpose of an MRPC is also an MRPC. We would like to note that though the shortening and puncturing have been discussed in [52], the results are about incorrectable erasure patterns and not MRPCs.

Definition 7 (Punctured code)**.** *Let $J \subset [n]$. The punctured code $\mathcal{C}|_J$ of the code $\mathcal{C}[n, k]$ is the linear code generated by the rows of the $k \times |J|$ submatrix of G, i.e. $G|_J$.*

Definition 8 (Shortened Code)**.** *The shortened code $\mathcal{C}^J$ of the linear code $\mathcal{C}$ is obtained by puncturing the set of codewords that have zeros in the J-locations,*

$$\mathcal{C}^J = \{\mathbf{c}|_{\bar{J}} : \mathbf{c} \in \mathcal{C} \textit{ and } \mathbf{c}|_{J=0}\}.$$

The generator matrix G^J of the shortened code $\mathcal{C}^J[n-|J|, k-|J|]$ is obtained from the generator matrix G of the code $\mathcal{C}[n,k]$ as follows:

- For all $j \in J$, find a unique row in which the j^{th} column is non zero.
- Delete those rows and all columns belong to J in the matrix G. Then such matrix obtained from G generates the shortened code.

It can be easily verify that shortened code and punctured code of MDS code is MDS.

Claim 1. *For any $m_1 < m$ (say $l = m - m_1$), if $E \in \mathcal{E}_{m_1,n}(a,b)$ then $E \in \mathcal{E}_{m,n}(a,b)$ and $E \cup E^{(l)} \in \mathcal{E}_{m,n}(a+l,b)$, where $E^{(l)}$ is an erasure pattern of size $m \times n$ with ln number of erasures which are located in last l rows.*

Proof. For any $E \in \mathcal{E}_{m_1,n}(a,b)$, there exist a code say $\mathcal{C} = \mathcal{C}_{\text{col}} \otimes \mathcal{C}_{\text{row}}$ which corrects E. Let $G = G_{\text{col}} \otimes G_{\text{row}}$ is the corresponding generator matrix, which implies $G|_E$ is full rank (i.e., $(m_1-a)(n-b)$).

Now, we can construct the code $\mathcal{C}^{'} = \mathcal{C}^{'}_{\text{col}} \otimes \mathcal{C}_{\text{row}}$ for the topology $T_{m,n}(a,b,0)$ with the generator matrix $G^{'} = G^{'}_{\text{col}} \otimes G_{\text{row}}$. Where

$$G^{'}_{\text{col}} = \left[\begin{array}{c|c} \underbrace{G_{\text{col}}}_{(m_1-a)\times(m_1)} & \underbrace{0}_{(m_1-a)\times(l)} \\ \hline \underbrace{0}_{(l)\times(m_1)} & \underbrace{I}_{(l)\times(l)} \end{array}\right].$$

Therefore, $G^{'} = \text{diag}(G_{\text{col}} \otimes G_{\text{row}}, I_l \otimes G_{\text{row}})$. Clearly, $\text{rank}(G^{'}|_{\overline{E}}) = (m_1-a)(n-b) + l(n-b) = (m-a)(n-b)$. Hence, $E \in \mathcal{E}_{m,n}(a,b)$.

Similarly, we can prove $E \cup E^{(l)} \in \mathcal{E}_{m,n}(a+l,b)$ by constructing the code $\mathcal{C}^{'} = \mathcal{C}^{'}_{\text{col}} \otimes \mathcal{C}_{\text{row}}$ for the topology $T_{m,n}(a+l,b,0)$ with the generator matrix $G^{'} = G^{'}_{\text{col}} \otimes G_{\text{row}}$. Where $G^{'}_{\text{col}} = [G_{\text{col}}|P]$, here P is the the matrix of size $(m_1-a) \times l$ in which columns are belong to the column space of G_{col}. Therefore, $G^{'} = [G_{\text{col}} \otimes G_{\text{row}} | P \otimes G_{\text{row}}]$. Hence, $\text{rank}(G^{'}|_{\overline{E \cup E^{(l)}}}) = \text{rank}(G|_{\overline{E}}) = (m_1-a)(n-b) = (m-a-l)(n-b)$.

□

Lemma 4.2.1. *If $\mathcal{C} = \mathcal{C}_{col} \otimes \mathcal{C}_{row}$ is an MR code for the topology $T_{m,n}(a, b, 0)$, then for any $M_1 \subset [m]$ and $|M_1| = m_1 < m$, by removing the coordinates $\bar{M}_1 = [m] \setminus M_1$ of column code $\mathcal{C}_{col}$ (say $\mathcal{C}_{col}|_{M_1}$) and with the same row code $\mathcal{C}_{row}$ we get the MR code for $T_{m_1,n}(a - m + m_1, b, 0)$.*

Proof. Suppose $G = G_{\text{col}} \otimes G_{\text{row}}$ is the generator matrix for the code $\mathcal{C}$ of particular form, where G_{col} is systematic. WLOG, assume that $M_1 = [m_1]$. The punctured code $\mathcal{C}'$ is obtained from the code $\mathcal{C}$ by puncturing all the coordinates belong to last $l = m - m_1$ rows of every codeword. The corresponding generator matrix $G' = G_{\text{col}}|_{M_1} \otimes G_{\text{row}}$, notice that $G' = G|_{[m_1 n]}$ (i.e. last ln columns are punctured).

From claim 1, for every erasure pattern $E \in \mathcal{E}_{m_1,n}(a - l, b)$ there exist an erasure pattern $E \cup E^{(l)} \in \mathcal{E}_{m,n}(a, b)$. Since C is an MR code, it can correct the $E \cup E^{(l)}$. Therefore, rank$(G|_{\overline{E \cup E^{(l)}}}) = (m - a)(n - b)$ and rank$(G|_{\overline{E \cup E^{(l)}}}) =$ rank$(G'|_{\overline{E}})$. Hence, the punctured code $\mathcal{C}'$ is an MR code for $T_{m_1,n}(a - m + m_1, b, 0)$.

□

Lemma 4.2.2. *If $\mathcal{C} = \mathcal{C}_{col} \otimes \mathcal{C}_{row}$ is an MR code for the topology $T_{m,n}(a, b, 0)$, then for any $M_1 \subset [m]$ and $|M_1| = m_1 < m$, by shortening the column code at $\bar{M}_1 = [m] \setminus M_1$ locations of column code $\mathcal{C}_{col}$ (say $\mathcal{C}_{col}^{\bar{M}_1}$) and with the same row code $\mathcal{C}_{row}$ we get the MR code for the topology $T_{m_1,n}(a, b, 0)$.*

Proof. Suppose $G = G_{\text{col}} \otimes G_{\text{row}}$ is the generator matrix for the code $\mathcal{C}$ of particular form, where G_{col} is systematic. WLOG, assume that $M_1 = [1 : m_1 - a, (m - a + 1) : m]$. Set of all codewords belong to $\mathcal{C}$ which are having zeros in the rows $\bar{M}_1 = [m_1 - a + 1 : m - a]$ is the subcode of $\mathcal{C}$ say $\mathcal{C}_{\text{sub}}^{\bar{M}_1}$. Let $\mathcal{C}^{\bar{M}_1}$ is the shortened code obtained by removing the coordinates belong to zeros in $\mathcal{C}_{\text{sub}}^{\bar{M}_1}$. I.e., $\mathcal{C}^{\bar{M}_1} = \mathcal{C}_{\text{col}}^{\bar{M}_1} \otimes \mathcal{C}_{\text{row}}$ and corresponding generator matrix $G^{\bar{M}_1} = G_{\text{col}}^{\bar{M}_1} \otimes G_{\text{row}}$. Now, we need to prove that $\mathcal{C}^{\bar{M}_1}$ is the MR code for the topology $T_{m_1,n}(a, b, 0)$. From claim 1, for every erasure pattern belong to $\mathcal{E}_{m_1,n}(a, b)$ there exists exactly same kind of erasure pattern in $\mathcal{E}_{m,n}(a, b)|_{M_1}$.

Any correctable erasure pattern $E \in \mathcal{E}_{m_1,n}(a, b)$ is correctable for all the codewords belong to $\mathcal{C}_{\text{sub}}^{\bar{M}_1}$, Since $\mathcal{C}$ is MR code. In fact, all the coordinates belong to $\bar{M}_1$ in $\mathcal{C}_{\text{sub}}^{\bar{M}_1}$ are zeros and are not useful to correct the erasures in E. Therefore, the shortened code $\mathcal{C}^{\bar{M}_1}$ corrects E.

□

Lemma 4.2.3. *For the topology $T_{m,n}(a,b,0)$ the code $\mathcal{C} = \mathcal{C}_{col} \otimes \mathcal{C}_{row}$ is an MR code if and only if $\mathcal{C}^* = \mathcal{C}_{row} \otimes \mathcal{C}_{col}$ is an MR code for $T_{n,m}(b,a,0)$.*

Proof. Firstly, we prove that for every code $\mathcal{C} = \mathcal{C}_{\text{col}} \otimes \mathcal{C}_{\text{row}}$ for the topology $T_{m,n}(a,b,0)$ there is an equivalent code $\mathcal{C}^* = \mathcal{C}_{\text{row}} \otimes \mathcal{C}_{\text{col}}$ with $T_{n,m}(b,a,0)$.

Suppose $G = G_{\text{col}} \otimes G_{\text{row}}$ is the generator matrix for the code $\mathcal{C}$. Consequently, $G^* = G_{\text{row}} \otimes G_{\text{col}}$ is the generator matrix for the code $\mathcal{C}^*$. From the tensor product of two matrices, We can see that G^* can be obtained from G by few row and column permutations. Therefore, the codes $\mathcal{C}$ and $\mathcal{C}^*$ are equivalent. But in the view of topologies, codewords of two codes are differ in its array structure and row code becomes column code and column code becomes row code. Thus, If the code $\mathcal{C}$ corrects an erasure pattern say E then $\mathcal{C}^*$ corrects E^T. Hence, there exist a bijective map $\phi : \mathcal{E}_{m,n}(a,b) \to \mathcal{E}_{n,m}(b,a)$, such that $\phi(E) = E^T$.

Now we prove one direction of the lemma and the other direction can also be done similarly. suppose the code $\mathcal{C}$ is MR code then it corrects every $E \in \mathcal{E}_{m,n}(a,b)$. So, it's equivalent code $\mathcal{C}^*$ can correct corresponding erasure pattern E^T which is belong to $\mathcal{E}_{n,m}(b,a)$. Since ϕ is bijective, $\mathcal{C}^*$ correct all the erasure patterns belong to $\mathcal{E}_{n,m}(b,a)$. Hence $\mathcal{C}^*$ is MR Code for the topology $T_{n,m}(b,a,0)$. □

Corollary 4.2.4. *If an erasure pattern is correctable in square product topology $T_{n,n}(a,a,0)$, then its transpose is also correctable in $T_{n,n}(a,a,0)$.*

4.3 Characterization of Recoverable Erasure Patterns

In this section, we discuss the characterization of correctable and non-correctable erasure patterns, the construction of bipartite graphs for Regular Erasure Patterns (REPs), the recoverability of REPs for the case when $a = 1$, and a subset of recoverable erasure patterns specifically when $a = 2$.

4.3.1 Correctable and Incorrectable Erasure Patterns

In this section, we first prove an independent claim which shows that for the case of $a = 1$, if a code can correct an erasure pattern, it can also correct a row permutation of the erasure pattern. Then, we go on to show that unions of regular erasure patterns are regular and correctable in a higher topology. Finally, we identify a class of erasure patterns based on unions of tensor products which are not correctable. This class is a generalization of the counter example of regular, incorrectable erasure pattern provided in [52].

Claim 2. *If the code $\mathcal{C}$ for $T_{m,n}(a = 1, b, 0)$ is constructed over the finite field of characteristic two and can recover from an erasure pattern E, then it can also recover from another erasure pattern E' which is obtained by permuting the rows of E.*

Proof. Suppose $G = G_{\text{col}} \otimes G_{\text{row}}$ is the generator matrix for the code $\mathcal{C}$, where $G_{\text{col}} = [I_{m-1} \,|\, \underline{1}]$. Let E be an erasure pattern in $[m] \times [n]$. Let $E_i \subseteq [n], i \in [m]$ denote the i^{th} row of erasures of E and let $L_i = [n] \backslash E_i$. Note that $E_i = \phi$, if i^{th} row has no erasures. An erasure pattern is recoverable if and only if the residual generator matrix after removing the columns corresponding to the erasures has full row rank.

$$G|_L = \begin{bmatrix} G_{\text{row}}|_{L_1} & & & G_{\text{row}}|_{L_m} \\ & \ddots & & \vdots \\ & & G_{\text{row}}|_{L_{m-1}} & G_{\text{row}}|_{L_m} \end{bmatrix}. \tag{4.1}$$

Wolog, we will consider the permutation which involves interchanging of just two coordinates. If it is interchanging within the rows $u_1, u_2 \in [m-1]$, the full row rank property of the residual generator matrix is clear. Now, consider the case when $u_1 = 1, u_2 = m$ rows are interchanged. The residual generator matrix $G|_{L'}$ is obtained by interchanging $G_{\text{row}}|_{L_1}$ and $G_{\text{row}}|_{L_m}$ in (4.1).

By premultiplying $G|_{L'}$ with $A_{(m-1)(n-b)\times(m-1)(n-b)} = \begin{bmatrix} I_{n-b} & & \\ \vdots & \ddots & \\ I_{n-b} & & I_{n-b} \end{bmatrix}$, we have a matrix equivalent (upto row and column permutations) to $G|_L$ (finite field is of characteristic 2). Hence rank$(G|_{L'})$ = rank$(G|_L)$ and the claim follows. □

Though the above claim is an independent statement, we believe that it sheds some light on the reason behind $a = 1$ case being easier to characterize than the general a case.

Definition 9 (l-partitioned REP (l-REP) for $T_{m,n}(a, b, 0)$)**.** *An Irreducible Regular Erasure Pattern (IREP) say E, and supports of E is enclosed by $U \times V \subseteq [m] \times [n]$ is l-REP if there exists $E_1, E_2, \ldots, E_l \in \mathcal{E}_{m,n}(a, b)$ and supp(E_i) is enclosed by $U_i \times V_i \subseteq [m] \times [n]$, $\forall i \in [l]$ such that $U_i \cap U_j = \emptyset$, $\forall i \neq j$ and $\bigcup_{i=1}^{l} U_i = U$.*

Lemma 4.3.1. *Any l-REP for topology $T_{m,n}(a, b, 0)$ is a REP with respect to the topology $T_{m,n}(al, b, 0)$.*

Proof. Let E is any l-REP for $T_{m,n}(a, b)$. Suppose $supp(E)$ is enclosed by $U \times V \subseteq [m] \times [n]$, where $|U| = u$, $|V| = v$.

$$\begin{aligned}
|E| = |E \cap (U \times V)| &= \sum_{i=1}^{l} |E_i \cap (U_i \times V_i)| \ (\because E \text{ is } l\text{-REP}) \\
&\leq \sum_{i=1}^{l} (v_i a + u_i b - ab) \ (\because E_i \text{ is IREP}) \\
&= v(al) + ub - (al)b.
\end{aligned}$$

□

In Lemma 4.2.2, we have shown that the shortening of an MR code is also an MR code. Utilizing this property, we now prove that an l-REP for $T_{m,n}(a, b, 0)$ is correctable in a different topology with the same array size ($m \times n$), but with the number of parities set to al (i.e., $T_{m,n}(al, b, 0)$).

Theorem 4.3.2. *Any l-partitioned REP for the topology $T_{m,n}(a, b, 0)$ is correctable with the topology $T_{m,n}(al, b, 0)$.*

Proof. Suppose E is any l-partitioned REP in which $supp(E)$ are enclosed by $U \times V \subseteq [m] \times [n]$ with $|U| = u$, $|V| = v$. Therefore, there exists $E_1, E_2, \ldots, E_l \in \mathcal{E}_{m,n}(a,b)$ such that $U_i \cap U_j = \emptyset$, $\forall i \neq j$ and $\bigcup_{i=1}^{l} U_i = U$. Where $supp(E_i)$ is enclosed by $U_i \times V_i \subseteq [m] \times [n]$ $\forall i \in [l]$ with $|U_i| = u_i$ and $|V_i| = v_i$. Let $\mathcal{C} = \mathcal{C}_{\text{col}} \otimes \mathcal{C}_{\text{row}}$ is the MR code with the generator matrix $G = G_{\text{col}} \otimes G_{\text{row}}$. From Lemma 4.2.2, each E_i correctable by the corresponding shortened code $\mathcal{C}^{\bar{U}_i} = \mathcal{C}_{\text{col}}^{\bar{U}_i} \otimes \mathcal{C}_{\text{row}}$ for the topology $T_{u_i,n}(a,b,0)$ and corresponding generator matrices are $G^{\bar{U}_i} = G_{\text{col}}^{\bar{U}_i} \otimes G_{\text{row}}$ $\forall i \in [l]$. Then E is correctable by the following generator matrix (G') for the topology $T_{m,n}(al,b,0)$:

$$G' = \left[\begin{array}{c|c|c|c|c} G^{\bar{U}_1} & 0 & \cdots & 0 & 0 \\ \hline 0 & G^{\bar{U}_2} & \cdots & 0 & 0 \\ \hline \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline 0 & 0 & \cdots & G^{\bar{U}_l} & 0 \\ \hline 0 & 0 & \cdots & 0 & G^* \end{array}\right], \tag{4.2}$$

where $G^* = I_{m-u} \otimes G_{\text{row}}$.

□

The sufficient condition of the Conjecture 1 given in [16] (i.e., every regular erasure pattern is correctable) is disproved by providing the following counter example in [52].

Example 3 ([52])**.** *An example of the REP, which is not correctable. Let E be the maximal erasure pattern for $T_{5,5}(2,2,0)$ and let the complement of erasure pattern be $\overline{E} = \{(1,1), (2,2), (2,3), (3,2), (3,3), (4,4), (4,5), (5,4), (5,5)\}$ (see Fig. 4.1).*

Alternatively, the given example of REP (Example 3) is demonstrated to be not correctable through the following lemma in [55].

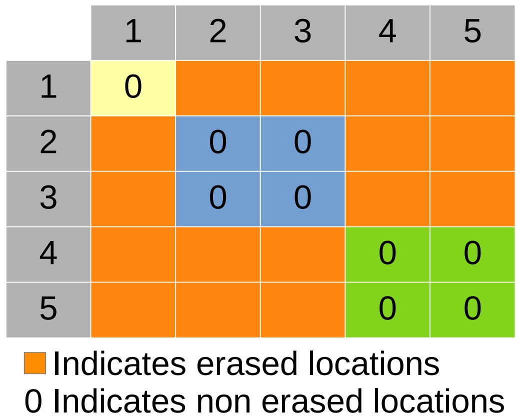

Figure 4.1: Example for a maximal erasure pattern in $T_{5,5}(2,2,0)$

Lemma 4.3.3 (Lemma 5.3 in [55])**.** *If U_1, U_2 are subspaces of U and V_1, V_2 are subspaces of V, then*

$$dim(U_1 \otimes V_1 + U_2 \otimes V_2) = dim(U_1).dim(V_1) + dim(U_2).dim(V_2) - dim(U_1 \cap U_2).dim(V_1 \cap V_2). \tag{4.3}$$

In the Example 3, we can find $U_1 = \{2,3\}, V_1 = \{2,3\}, U_2 = \{4,5\}, V_2 = \{4,5\}$. We get, $dim(U_1 \otimes V_1 + U_2 \otimes V_2) = 7$ which is less than the $|(U_1 \otimes V_1) \cup (U_2 \otimes V_2)| = 8$. Therefore, from Eq. (4.3), the erasure pattern E is not correctable.

From Eq. (4.3), we can observe that an erasure pattern E is not correctable whenever there exists U_1, V_1, U_2 and V_2 in E such that $dim(U_1 \cap U_2).dim(V_1 \cap V_2) > 0$. The collection of all such erasure patterns are not correctable REPs. Building upon this condition, we have characterized a class of erasure patterns that are not correctable, as clarified in the following theorem.

Theorem 4.3.4. *Suppose E is the maximal erasure pattern for $T_{m,n}(a,b,0)$, i.e., $\overline{E} = (m-a)(n-b)$. For any such E, If there exist $(U_1 \times V_1)$ & $(U_2 \times V_2) \subset [m] \times [n]$ in $\overline{E}$ such that*

i) $|U_i| < (m-a), |V_i| < (n-b)\, \forall i = \{1,2\}$ and

ii) U_1, U_2, V_1 and V_2 satisfies any of the following conditions:

a) $|U_1 \cup U_2| > (m-a)$ *and* $|V_1 \cup V_2| > (n-b)$.

b) $|U_1 \cup U_2| > (m-a)$, $|V_1 \cup V_2| \leq (n-b)$ *and* $|V_1 \cap V_2| \neq 0$.

c) $|U_1 \cup U_2| \leq (m-a)$, $|V_1 \cup V_2| > (n-b)$ *and* $|U_1 \cap U_2| \neq 0$.

then E is not correctable.

Proof. We use the same notation to represent subspaces of column code U_1, U_2 with respect to the column indices U_1, U_2 of it's generator matrix interchangeably and also for V_1, V_2. Now using Lemma 4.3.3, we get

$$dim(U_1 \otimes V_1 + U_2 \otimes V_2) = |U_1||V_1| + |U_2||V_2| - dim(U_1 \cap U_2).dim(V_1 \cap V_2)$$

In above, equality holds from the condition i) of the theorem. Now, for every case of condition ii), we can verify $dim(U_1 \cap U_2).dim(V_1 \cap V_2) > |U_1 \cap U_2|.|V_1 \cap V_2|$. Therefore, $dim(U_1 \otimes V_1 + U_2 \otimes V_2) < |(U_1 \otimes V_1) \cup (U_2 \otimes V_2)|$. Hence, E is not correctable. □

Note that the erasure pattern given as Example 3 falls under the case of i)a) of the theorem 4.3.4. Another example is illustrated in the following.

Example 4. *Let E is the maximal erasure pattern for $T_{7,7}(3,3,0)$ and* $\overline{E} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3),(3,4),(4,2),(4,3),(4,4),(5,2),(5,3),(5,4)\}$.

In this $\overline{E}$ we can find $U_1 = \{1,2,3\}, V_1 = \{1,2,3\}, U_2 = \{3,4,5\}, V_2 = \{2,3,4\}$ and are satisfies the conditions i) and ii)b) of the theorem 4.3.4. We get $dim(U_1 \otimes V_1 + U_2 \otimes V_2) = 15$ which is less than the $|(U_1 \otimes V_1) \cup (U_2 \otimes V_2)| = 16$. Therefore, E is not correctable. Without checking the regularity condition we can find it is not correctable. We would like to note here that these patterns are incorrectable irrespective of if these patterns are regular and not. Hence, there is no need to check for the regularity condition.

4.3.2 Bipartite Graphs for Regular, Irreducible Erasure Patterns

In this section, we construct two bipartite graphs based on an erasure pattern and derive some properties of these graphs.

Construction 4.3.5 (Bipartite Graph between erasures and non-erasures for general $a \geq 1$)**.** *Consider a row-wise irreducible erasure pattern E with enclosing grid $U \times V \subseteq [m] \times [n], |U| = u, |V| = v$, where enclosing grid is used to refer to the smallest grid containing the erasure pattern E. Assuming that the elements of U are sorted, let the erasure pattern be such that each row has $b+r_i, i \in U$ erasures. Let $U_L \subseteq U$ be arbitrary subset of $u-a$ elements and $U_R = U \setminus U_L$. We construct a bipartite graph as follows:*

- *For each $i \in U_L$, we create r_i vertices on the left. The r_i left vertices corresponding to $i \in U_L$ are denoted by $e(i,1), e(i,2), \ldots, e(i,r_i)$. Hence, the total number of vertices on the left are $\sum_{i \in U_L} r_i$.*
- *Each vertex on the right corresponds to one non-erasure in the rows U_R. Let there be w non-erasures in the rows U_R. The vertices on the right are denoted by $d_1, d_2, \ldots, d_w$.*
- *We place an edge between a left vertex $e(i,j)$ and a right vertex d_ℓ if there exists an erasure in the position $(s,t) \in [m] \times [n]$ where s is the row number of the erasure $e(i,j)$ and t is the column number of the non-erasure d_ℓ.*

Lemma 4.3.6. *If an erasure pattern is regular and row-wise irreducible for topology $T_{m,n}(a,b,0)$, then there exists a complete matching[1] in the bipartite graph (for the erasure pattern) resulting from Construction 4.3.5.*

Proof of Lemma 4.3.6. We will prove that there exists a matching by verifying the Hall's condition. To do so, we consider all the left vertices corresponding to $U_S \subseteq U_L$, where $|U_S| = s$. The number of such vertices on the left are given by $\sum_{i \in U_S} r_i$. Let $U_S \times V_T$ denote the enclosing grid of all the erasures in the rows U_S. Denote $|V_T| = t$. Consider the erasures in the

[1]By complete matching in a bipartite graph, we refer to a matching in which all the left vertices are included. In this paper, whenever we refer to matching in a bipartite graph, we mean complete matching.

	1	2	3	4	5	6	7	8	9	10
1							×	×	×	×
2						×	×	×		
3			×						×	×
4				×	×	×				
5			×	×	×					
6										

Figure 4.2: Example of a regular erasure pattern, $(m, n) = (6, 10), (a, b) = (1, 2)$. Enclosing grid of the erasure pattern is $[1 : 5] \times [3 : 10]$.

	1	2	3	4	5	6	7	8	9	10
1							×	⊗	⊗	×
2						×	⊗	×		
3			×						×	⊗
4				×	×	⊗				
5			×	×	×	◯	◯	◯	◯	◯
6										

Figure 4.3: The matching in construction II.1.

grid $(U_S \cup U_R) \times V_T$ of $s + a$ rows and t columns. Let x denote the number of erasures in the subgrid $U_R \times V_T$. Since the erasure pattern is regular and irreducible, we apply the condition in (2.11) to the grid $(U_S \cup U_R) \times V_T$. Then, we have

$$sb + \sum_{i \in U_S} r_i + x \leq ta + (s + a)b - ab. \tag{4.4}$$

Thus, we have an upper bound on x as $x \leq at - \sum_{i \in U_S} r_i$. Thus, the number of non-erasures in these t columns is lower bounded by $p = at - x \geq \sum_{i \in U_S} r_i$. This proves that the neighbourhood of a set of size $\sum_{i \in U_S} r_i$ is at least $\sum_{i \in U_S} r_i$. Hence, for any set A where we consider all the vertices corresponding to any s rows in the bipartite graph, we have that $|N(A)| \geq |A|$.

Now, consider the case when we take sets A such that A partially intersects s rows. Since the neighbourhood $N(A)$ in this case is the same as that we would have obtained when we consider all the vertices corresponding to these s rows, it is true that $|N(A)| \geq |A|$ even in this case.

□

Construction 4.3.7 (Bipartite Graph between rows and columns for $a = 1$)**.** *Consider a row-wise irreducible erasure pattern E with enclosing grid $U \times V \subseteq [m] \times [n], |U| = u, |V| = v$. Let ℓ denote an arbitrary element of U and the support of $b + r_\ell$ erasures in the row given by the set V_ℓ. Consider the erasures in the grid $(U \setminus \ell) \times (V \setminus V_\ell)$. We construct a bipartite graph as follows:*

- *The vertices on the left correspond to the elements of the set $(U \setminus \ell)$.*
- *The vertices on the right correspond to the elements of the set $(V \setminus V_\ell)$*
- *We place an edge between two vertices i and j if the array element (i, j) is erased in E.*

Lemma 4.3.8. *Consider an erasure pattern which is regular and row-wise irreducible for topology $T_{m,n}(a = 1, b, 0)$. Consider the bipartite graph (for the erasure pattern) resulting from Construction 4.3.7. The following property holds for this bipartite graph: If $A \subseteq U \setminus \ell$ (left vertices), then the neighbourhood of A, $N(A)$ satisfies $|N(A)| \geq \sum_{i \in A} r_i$.*

Proof of Lemma 4.3.8. Consider the left vertices corresponding to $U_S \subseteq (U \setminus \ell)$, where $|U_S| = s$. Let $(U_S \cup \ell) \times V_T$ denote the enclosing grid of all the erasures in the rows $U_S \cup \ell$. We note that $|V_\ell| = b + r_\ell$. Also we denote $|V_T \setminus V_\ell| = t$. Since the erasure pattern is regular and irreducible, we apply the condition in (2.11) to the grid $(U_S \cup \ell) \times V_T$. Then, we have

$$(s+1)b + \sum_{i \in (U_S \cup \ell)} r_i \leq (t + b + r_\ell) + (s+1)b - b. \tag{4.5}$$

The above equation implies that $t \geq \sum_{i \in U_S} r_i \geq s$. □

We would like to note that for the case of $a = 1$, both the above constructions result in the same bipartite graph.

	3	4	5	9	10
1				×	×
3	×			×	×
4		×	×		
5	×	×	×		

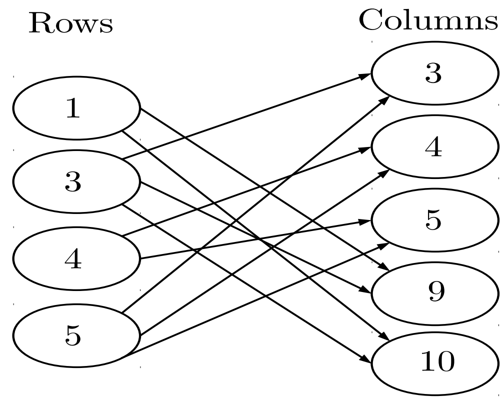

Figure 4.4: Continuing from previous example in Fig. 4.2, we have $\ell = 2, V_\ell = \{6, 7, 8\}$, $r_1 = 2, r_3 = 1, r_4 = 1, r_5 = 1$. Note that the neighbourhoods of subsets of left vertices satisfy the condition in Lemma 4.3.8.

4.3.3 Recoverability of Regular Erasure Patterns for $a = 1$

In this section, we give an alternate proof for the Theorem 4.3.11 which is the Conjecture 1 for $a = 1$ [16]. The following two lemmas would be useful in the proving the theorem.

Lemma 4.3.9. *Consider a square matrix B of size $n \times n$. The matrix consists of zeros at some positions and distinct variables (indeterminates) in the rest of the positions. Consider a bipartite graph constructed based on this matrix as follows:*

- *The left vertices correspond to rows.*
- *The right vertices correspond to columns.*
- *We place an edge between two vertices i, j, whenever there is a variable in the position (i, j).*

If there is a matching in the bipartite graph thus constructed, then det(B) is a non-zero (multivariate) polynomial and the variables can be assigned values from a large enough finite field $\mathbb{F}_q$ such that the matrix is full rank.

Proof. If a variable is present in position (i, j), then we denote the variable by $x_{i,j}$. Let $x_{i_1,j_1}, x_{i_2,j_2}, \ldots x_{i_n,j_n}$ be the variables involved in the matching. The determinant of the matrix is a multi-variate polynomial and due to the matching, $\prod_{\ell=1}^{n} x_{i_\ell j_\ell}$ is one of the monomials adding to the determinant polynomial. $\prod_{\ell=1}^{n} x_{i_\ell, j_\ell}$ has a nonzero coefficient as no other term in the determinant would give the same monomial. This is due to the fact that all the variables in the matrix are distinct. Hence, the determinant polynomial is a non-zero polynomial. It follows by Schwartz-Zippel Lemma that the indeterminates can be assigned values from a large enough finite field such that the determinant of the matrix is nonzero and hence the matrix is full rank. □

Lemma 4.3.10 ([16]). *Consider an erasure pattern $E \subseteq [m] \times [n]$. Let $E' \subseteq E$ be a row-wise irreducible erasure pattern obtained as follows: If i^{th} row ($1 \leq i \leq m$) of E has $\geq b+1$ erasures, then i^{th} row of E' is identical to i^{th} row of E. All the rest of the rows are non erasures in E'. Then E is recoverable if and only if E' is recoverable.*

Theorem 4.3.11 ([16]). *For the topology $T_{m,n}(a = 1, b, 0)$, if an erasure pattern is regular, then it is recoverable.*

Proof. Based on the above lemma, in order to prove Theorem 4.3.11, it is enough to consider row-wise irreducible, regular erasure patterns. In [16], the proof of Theorem 4.3.11 considered the following two cases:

- Case 1: E have exactly $b+1$ erasures in each row (which has nonzero erasures). This can be considered as the base case.

- Case 2: E have $b + r_i, r_i \geq 1, i \in U$ erasures in each row (where $U \times V$ is the enclosing grid of E).

We will give an alternate proof which unifies both the cases. This proof will be generalized later to the case of $a = 2$ for some erasure patterns.

Consider a row-wise irreducible, regular erasure pattern E which has an enclosing grid of $U \times V$ and has $b + r_i, r_i \geq 1, i \in U$ erasures in each row. If $|U| = 1$, a simple parity check code as the column code will suffice to correct the erasure pattern. So, we assume that $|U| \geq 2$. To prove that E is recoverable, we need to construct a code $\mathcal{C}$ which is an instantiation of topology $T_{m,n}(a = 1, b, 0)$ such that $\dim(\mathcal{C}|_{D \setminus E}) = \dim(\mathcal{C})$, where $D = [m] \times [n]$. Since $\mathcal{C}$ is an instantiation of topology $T_{m,n}(a = 1, b, 0)$ and Definition 1 for $h = 0$ case is precisely the definition of product of codes [61], we have $\mathcal{C} = \mathcal{C}_{\text{col}} \otimes \mathcal{C}_{\text{row}}$. To construct $\mathcal{C}$, we construct the generator matrices of $\mathcal{C}_{\text{col}}$ and $\mathcal{C}_{\text{row}}$ [16], denoted by G_{col} and G_{row} respectively.

For correcting any row-wise irreducible, regular erasure pattern E, the column code $\mathcal{C}_{\text{col}}$ is a simple parity check code, the generator matrix of which is given by

$$G_{\text{col}} = [\underline{1} \quad I_{m-1}], \tag{4.6}$$

where G_{col} is a $(m-1) \times m$ matrix.

The row code $\mathcal{C}_{\text{row}}$ is constructed based on the erasure pattern E. The generator matrix of the row code G_{row} is of the size $(n-b) \times n$ and the entries of the generator matrix are either variables(indeterminates) or zeros. A variable present at position (i, j) is denoted by $x_{i,j}$.

- For $j \in [n] \setminus V$, which has no erasures, a row is added in the generator matrix G_{row} which has a variable in the j^{th} position and zeros in all the other positions.

- Consider a row of the erasure pattern E which has $b + r_i, i \in U$ erasures and let $i \times V_i$ denote the enclosing grid of the row of erasures. Let V_T denote a b element subset of V_i. r_i rows are added in the generator matrix corresponding to this row of the erasure pattern. Each of the r_i rows of the generator matrix is formed by placing variables in columns V_T and at one additional column in $V_i \setminus V_T$. All the rest of the entries are zeros.

- Until now, the number of rows of generator matrix which have already been filled are $n - v + \sum_{i \in U} r_i$. Since the erasure pattern in regular, we have that

  $$ub + \sum_{i \in U} r_i \leq v + ub - b.$$

  Hence, to complete the $n - b$ rows of the generator matrix, we have to add $n - b - (n - v + \sum_{i \in U} r_i) = v - b - \sum_{i \in U} r_i = t$ rows. Each of these rows is formed by placing variables in the V columns and zeros in the other $[n] \setminus V$ columns.

Combining all the above, G_{row} (upto permutation of columns) can be written as

$$G_{\text{row}} = \begin{bmatrix} \underbrace{G_I}_{(n-v)\times(n-v)} & 0 \\ 0 & \underbrace{G_S}_{(\sum_{i\in U} r_i)\times v} \\ 0 & \underbrace{G_T}_{t\times v} \end{bmatrix}.$$

$$\begin{bmatrix} x_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & x_{2,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_{3,7} & x_{3,8} & x_{3,9} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_{4,7} & x_{4,8} & 0 & x_{4,10} \\ 0 & 0 & 0 & 0 & 0 & x_{5,6} & x_{5,7} & x_{5,8} & 0 & 0 \\ 0 & 0 & x_{6,3} & 0 & 0 & 0 & 0 & 0 & x_{6,9} & x_{6,10} \\ 0 & 0 & 0 & x_{7,4} & x_{7,5} & x_{7,6} & 0 & 0 & 0 & 0 \\ 0 & 0 & x_{8,3} & x_{8,4} & x_{8,5} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 4.5: G_{row} for the erasure pattern in the earlier example. Rows 1 and 2 in the above matrix correspond to the first two non-erasure columns. Rows 3 and 4 correspond to the first row of the erasure pattern. Note that $V_1 = \{7, 8, 9, 10\}$ and $V_T = \{7, 8\}$. Rows $5, 6, 7$ and 8 correspond to the next four rows of the erasure pattern. In this matrix, there is no G_T component.

The generator matrix G of the product code [61] in terms of the generator matrices of the row and column codes is given by

$$\begin{aligned} G &= G_{\text{col}} \otimes G_{\text{row}} \\ &= \begin{bmatrix} G_{\text{row}} & G_{\text{row}} & & \\ G_{\text{row}} & & \ddots & \\ G_{\text{row}} & & & G_{\text{row}} \end{bmatrix}. \end{aligned} \tag{4.7}$$

Now, we have to prove that the erasure pattern E is recoverable by the code $\mathcal{C}$. It is enough to show that there exists an assignment of the variables in G_{row} such that $\text{rank}(G|_{D\setminus E}) =$

$(n-b)(m-1)$. Without loss of generality, we assume that the parity block column (the one which has m copies of G_{row}) is always included in E. Otherwise, the columns of G_{col} can be permuted so that it is included.

To examine the structure of $G|_{D\setminus E}$, we will first consider the systematic part (last $m-1$ block columns in (4.7)). G_{row} corresponding to $i \in U$ has erasures and the submatrix which remains after deleting the columns corresponding to the erasures has the structure[2]

$$G_{\text{row}}|_{[n]\setminus V_i} = \begin{bmatrix} \underbrace{G_I}_{(n-v)\times(n-v)} & 0 \\ 0 & \underbrace{G_{S_i}}_{(\sum_{i\in U} r_i \times v-b-r_i)} \\ 0 & \underbrace{G_{T_i}}_{(t\times v-b-r_i)} \end{bmatrix}.$$

It can be observed based on the construction of G_{row} that G_{S_i} has r_i zero rows. Let G_{Z_i} denote the matrix which remains after removing the r_i zero rows from G_{S_i}. G_{row} corresponding to $i \in [m]\setminus U$ remains unchanged, since there are no erasures in these rows. For consistency of notation, we have $V_i = \phi$, $G_{S_i} = G_{Z_i} = G_S$, $G_{T_i} = G_T$ for $i \in [m]\setminus U$. For ease of notation, we denote $\begin{bmatrix} G_{Z_i} \\ G_{T_i} \end{bmatrix}, i \in [m]$ by G_{Y_i}.

By rearranging the rows of $G|_{D\setminus E}$ so that all the zero rows in $G_{S_i}, \forall i \in U$ are shuffled to the top, the resulting matrix G_π has the following structure:

[2]The matrices G_{S_i}, G_{T_i}, G_{Z_i} and G_{Y_i} are used to denote particular sub matrices of G_{row}. Note that S_i, T_i, Z_i, Y_i by themselves do not refer to anything.

$$
G_{\pi} = \left[\begin{array}{c|ccccc}
G_P & & & & & \\
\hline
 & G_I & & & & \\
 & & G_{Y_1} & & & \\
G_L & & & \ddots & & \\
 & & & & G_I & \\
 & & & & & G_{Y_{m-1}}
\end{array}\right],
$$

where G_P is of size $(\sum_{j\in U} r_j) \times (n-b-r_1)$.

Claim 3. *Consider the matrix $G_{Y_i}, i \in [m]$. There exists a complete matching in the bipartite graph constructed based on this matrix as in Lemma 4.3.9.*

Proof. First, we will consider the case when $i \in U$. We will show that there is a matching in G_{Z_i} and since G_{T_i} contains rows completely filled with variables, the matching in G_{Z_i} can be easily extended to a matching in G_{Y_i}. In order to show that there is a matching in G_{Z_i}, we will verify the Hall's condition. Consider a subset A formed by including all the $\sum_{j\in U_S} r_j$ vertices associated with rows $U_S \subseteq U$. The mapping between rows U_S and left vertices of the bipartite graph can be done since the rows of G_{row} (and hence G_{Z_i}) are constructed based on the rows U. Applying Lemma 4.3.8 (since G_{Z_i} is obtained by removing columns V_i from G_{row}), we have that $|N(A)| \geq \sum_{j\in U_S} r_j$. Now, we consider the case when the subset A is formed by t_j of r_j vertices corresponding to rows U_S in U, where $t_j < r_j, j \in U_S$. Note that $|A| = \sum_{j\in U_S} t_j$. Based on the construction of matrix G_S, we have that by removing $r_j - t_j$ vertices corresponding to j^{th} row, the neighbourhood can reduce almost by $r_j - t_j$. Hence, it follows that $|N(A)| \geq \sum_{j\in U_S} r_j - \sum_{j\in U_S}(r_j - t_j) = \sum_{j\in U_S} t_j$.

Now, consider the case when $i \in [m] \setminus U$. Since $|U| \geq 2$, there is at least some i such that bipartite graph of G_{Y_i} has a matching (say M_1). The r_i rows and the $b+r_i$ columns indexed by V_i, which have been erased to obtain G_{Y_i}, have a matching within themselves (say M_2), since the neighbourhood of any one of the r_i rows has exactly one column unique to itself. Then, $M_1 \cup M_2$ is a matching in $\begin{bmatrix} G_S \\ G_T \end{bmatrix}$. □

Let $G_{Y_i'}, i \in [m]$ denote the square submatrix of G_{Y_i} which is associated with the matching in Claim 3. Applying Lemma 4.3.9, we have that $\det(G_{Y_i'})$ is a non-zero polynomial.

Now consider the matching which results by applying Lemma 4.3.6 to the erasure pattern E with $U_R = \{1\}$. Let $V_M \subseteq [n]$ denote the columns (right vertices) in the matching. In the example in Fig. 4.3, $V_M = \{6, 7, 8, 9, 10\}$. Let $G_{P'}$ be square submatrix of G_P by restricting to V_M columns. It can seen that the all the variables in $G_{P'}$ are all distinct, and by Lemma 4.3.6, there exists a matching between the $\sum_{j \in U} r_j$ rows and the columns that are retained in $G_{P'}$. Hence, applying Lemma 4.3.9, we have that $\det(G_{P'})$ is also a non-zero polynomial.

Consider the following square submatrix of G_π:

$$G_{\pi'} = \left[\begin{array}{c|ccccc} G_{P'} & & & & & \\ \hline & G_I & & & & \\ & & G_{Y_1'} & & & \\ G_{L'} & & & \ddots & & \\ & & & & G_I & \\ & & & & & G_{Y_{m-1}'} \end{array}\right],$$

$$\det(G_{\pi'}) = \det(G_{P'})\det(G_I)^{m-1} \prod_{i=1}^{m-1} \det(G_{Y_i'}).$$

It follows that $\det(G_{\pi'})$ is a non-zero multivariate polynomial, since each of the factors in the product are non-zero. Hence, the variables can be assigned values from a sufficiently large finite field $\mathbb{F}_q$ such that $G_{\pi'}$ is a full rank matrix. Hence, $\text{rank}(G|_{D \setminus E}) = \text{rank}(G_{\pi'}) = (n-b)(m-1)$. Thus, we have proved that the erasure pattern E is recoverable.

□

4.3.4 Partial Characterization of Recoverable Erasure Patterns for $a = 2$

In this section, we define an extended erasure pattern $E^{'}$ of E where E is an erasure pattern for topology $T_{m,n}(a = 1, b, 0)$, E' is for $T_{m+m',n}(a = 2, b, 0)$ and E' is obtained from E by replicating some rows of erasures in E. If E is row-wise irreducible and regular, we prove that $E^{'}$ is also regular and recoverable.

Definition 10 (Extended Erasure Pattern). *Consider an erasure pattern $E \subseteq [m] \times [n]$ which is row-wise irreducible and regular for the topology $T_{m,n}(a = 1, b, 0)$. Let $U \times V$ denote the enclosing grid of E in $[m] \times [n]$. Let $i \times V_i$ denote the enclosing grid for the erasures in i^{th} row $i \in U$. Consider an erasure pattern E' for the topology $T_{m+m',n}(a = 2, b, 0)$, $m' \leq m$ formed by extending E as follows:*

- *Rows of the erasure pattern are replicated i.e., $V_{m+\ell} = V_j$, $1 \leq \ell \leq m', 1 \leq j \leq m$.*
- *The replication factor of any row of the erasure pattern is atmost two, i.e., $V_{m+\ell} \neq V_{m+\ell'}$ when $\ell \neq \ell'$.*

The erasure pattern E' will be referred to as extended erasure pattern.

Lemma 4.3.12. *Any extended erasure pattern resulting from Definition 10 is row-wise irreducible and regular for the topology $T_{m+m',n}(a = 2, b, 0)$.*

Proof. Let E' be an extended erasure pattern of E. It is clear that E' is row-wise irreducible. Consider a sub grid $U \times V \subseteq [m + m'] \times [n]$. It is enough to consider $|U| \geq a + 1 = 3$ and $|V| \geq b + 1$ to verify the regularity condition.

Let $U_1 = U \cap [m]$ and $U_2 = U \cap \{m + 1, \ldots, m + m'\}$. By the definition of extended erasure pattern, corresponding to U_2, there is a set $U_2' \in [m]$ such that the structure of erasures in $U_2 \times V$ is the same as that in $U_2' \times V$.

$$
\begin{aligned}
|E' \cap (U \times V)| &= |E' \cap ((U_1 \cup U_2) \times V)| \\
&= |E' \cap (U_1 \times V)| + |E' \cap (U_2 \times V)| \\
&= |E \cap (U_1 \times V)| + |E \cap (U_2' \times V)| \\
&\overset{(a)}{\leq} (v + u_1 b - b) + (v + u_2 b - b) \\
&= 2v + ub - 2b,
\end{aligned}
$$

where (a) follows since E is regular for topology $T_{m,n}(a = 1, b, 0)$.

□

Theorem 4.3.13. *Any extended erasure pattern resulting from Definition 10 is recoverable for the topology* $T_{m+m',n}(a = 2, b, 0)$.

Proof. Let E' be the extended erasure pattern of E, where E is row-wise irreducible and regular for the topology $T_{m,n}(a = 1, b, 0)$. Let $U \times V$ denote the enclosing grid of E' in $[m + m'] \times [n]$. To recover E', we employ the same row code as the one used for recovering E in $T_{m,n}(a = 1, b, 0)$, the construction of which is described in the proof of Theorem 4.3.11. The generator matrix of the column code G_{col} is given by

$$
G_{\text{col}} = \begin{bmatrix} \Sigma_{(m+m'-2)\times 2} & \Lambda_{(m+m'-2)\times(m+m'-2)} \end{bmatrix}, \tag{4.8}
$$

where $\Sigma = [\sigma_{i,j}]$, $1 \leq i \leq m + m' - 2$, $1 \leq j \leq 2$ and all the entries in Σ are indeterminates, Λ is a diagonal matrix with entries $\lambda_{i,i}$ as indeterminates. The product code has the following generator matrix

$$
\begin{aligned}
G &= G_{\text{col}} \otimes G_{\text{row}} = [\Sigma \otimes G_{\text{row}} \quad \Lambda \otimes G_{\text{row}}] \\
&= \begin{bmatrix}
\sigma_{1,1} G_{\text{row}} & \sigma_{1,2} G_{\text{row}} & \lambda_{1,1} G_{\text{row}} & & \\
\vdots & \vdots & & \ddots & \\
\sigma_{\ell,1} G_{\text{row}} & \sigma_{\ell,2} G_{\text{row}} & & & \lambda_{\ell,\ell} G_{\text{row}}
\end{bmatrix},
\end{aligned}
$$

where $\ell = m + m' - 2$.

Similar to the $a = 1$ case, after rearranging the zero rows of $G|_{D\setminus E'}$, the resulting matrix G_π has the following structure.

$$G_\pi = \left[\begin{array}{c|cccccc} G_P & & & & & & \\ \hline & \lambda_{1,1}G_I & & & & & \\ & & \lambda_{11}G_{Y_1} & & & & \\ G_L & & & \ddots & & & \\ & & & & \lambda_{\ell,1}G_I & & \\ & & & & & \lambda_{\ell,1}G_{Y_\ell} \end{array}\right],$$

where G_P is of size $(\sum_{j\in U} r_j) \times (2n - 2b - r_1 - r_2)$. Note that G_P and G_L are obtained by combining the first two block columns in $G|_{D\setminus E'}$. The matching in $G_{Y_i}, i \in \ell$ follows from the $a = 1$ case since the row code is the same.

Now consider the matching which results by applying Lemma 4.3.6 to the erasure pattern E' with $U_R = \{1, 2\}$. Let V_M denote the right vertices in the matching. Let $G_{P'}$ be square submatrix of G_P by restricting to V_M columns. By Lemma 4.3.6, there exists a matching between the $\sum_{j\in U} r_j$ rows and the columns that are retained in $G_{P'}$. However, note that unlike the $a = 1$ case, each non-zero entry in this case is a product of variables $\sigma_{\alpha,\beta}$ and $x_{j,k}$. Also, note that the product of variables given by the matching is a monomial which cannot be cancelled by any other term in det$(G_{P'})$. To show this, assume that one of the entry in the matching is $\sigma_{1,\beta}x_{j,k}$. We would like to note that there can be atmost one more variable in $G_{P'}$ containing $x_{j,k}$ and if it is present, then necessarily it must be multiplied by $\sigma_{2,\beta}$. Hence, the monomial formed by the matching is unique, following which det$(G_{P'})$ is a non-zero polynomial. Rest of the proof is exactly same as the $a = 1$ case.

□

4.4 Generic Matrices and Higher Order MDS Codes

In this section, we will introduce generic matrices and higher order MDS codes which have been shown in [55], that these are constituent row codes for the case of $a = 1$ case. We prove a certain inclusion-exclusion type principle relating to the dimension of intersection spaces of generic matrices. We will use this result to derive a correspondence between $(n, 3)$-MDS(3) code and its associated projective space.

Definition 11 (Generic Matrix). *A matrix $W \in \mathbb{R}^{k\times n}$ (i.e., taking values from real numbers) is said to be a generic matrix, if the set of columns of W avoid a fixed low-dimensional algebraic variety with probability one.*

We can interpret the genericity of a matrix as follows:

- If we pick all the elements from a large finite field, the probability that the vectors are from low dimensional algebraic variety is close to zero.
- This also means that the sum space of column spaces in general have maximum rank possible and the intersection of column spaces have minimum rank possible with high probability.

We will now prove an inclusion-exclusion principle type result related to the dimension of generic matrices.

Theorem 4.4.1. *Consider a generic matrix W of size $k \times n$ over $\mathbb{F}$. Consider pair-wise disjoint subsets $A_1 \ldots, A_l$ of $[n]$, then we have that*

$$\begin{aligned}
&dim(W_{A_1} \cap \ldots \cap W_{A_l}) = \sum_{i}^{l} \min(|A_i|, k) \\
&- \sum_{1\le i<j\le l} \min(|A_i| + |A_j|, k) \\
&+ \sum_{1\le i<j<t\le l} \min(|A_i| + |A_j| + |A_t|, k) \\
&\ldots + (-1)^{l+1} \min(|A_1| + |A_2| + \ldots + |A_l|, k).
\end{aligned}$$

Proof. We will prove this result by induction. For the case of $l = 2$, we have that

$$\begin{aligned} & dim(W_{A_1} \cap W_{A_2}) \\ = & \; dim(W_{A_1}) + dim(W_{A_2}) - dim(W_{A_1} + W_{A_2}) \\ = & \; \min(|A_1|, k) + \min(|A_2|, k) - \min(|A_1| + |A_2|, k) \end{aligned}$$

Assuming that the induction hypothesis is true $l - 1$, we will prove it for the case of l.

$$\begin{aligned} & dim(W_{A_1} \cap \ldots \cap W_{A_l}) = \\ & dim(W_{A_1} \cap \ldots W_{A_{l-1}}) + dim(W_{A_l}) \\ & -dim((W_{A_1} \cap \ldots \cap W_{A_{l-1}}) + W_{A_l}) \end{aligned}$$

Since $A_1, A_2, \ldots, A_l$ are disjoint, we have that $dim((W_{A_1} \cap \ldots W_{A_{l-1}}) + W_{A_l}) = \min(dim(W_{A_1} \cap \ldots W_{A_{l-1}}) + dim(W_{A_l}), k)$. Applying this and the induction hypothesis, the statement of the theorem follows. □

Note that such inclusion-exclusion type principle is in general not true for the dimensions of vector spaces. However, it is interesting that for the case of generic matrices with column subspaces picked based on disjoint matrices, such a relation holds.

Definition 12 (Higher order MDS code (MDS(l)). *Let C be an (n, k) code with generator matrix $V_{k \times n}$. For $l \geq 2$, we say that C is an MDS(l) code if for all $A_1, \ldots, A_l \subset [n]$,*

$$dim(V_{A_1} \cap \ldots \cap V_{A_l}) = dim(W_{A_1} \cap \ldots \cap W_{A_l}),$$

where W is a $k \times n$ generic matrix.

Theorem 4.4.2 (Equivalent definition of MDS(l)). *Let V is the generator matrix for the (n, k) MDS code. Then the code generated by V is MDS(l) if and only if for all $A_1, \ldots, A_l \subseteq [n]$ such that $|A_i| \leq k$, $\sum_{i=1}^{l} |A_i| = (l - 1)k$ and $A_1 \cap \ldots \cap A_k = \emptyset$, we have that*

$$dim(V_{A_1} \cap \ldots \cap V_{A_l}) = 0 \Longleftrightarrow dim(W_{A_1} \cap \ldots \cap W_{A_l}) = 0.$$

Theorem 4.4.3 ([55]). *Let $\mathcal{C} = \mathcal{C}_{col} \otimes \mathcal{C}_{row}$ be product code for the topology $T_{m,n}(1, b, 0)$. Then $\mathcal{C}$ is MR code if and only if $\mathcal{C}_{row}$ is an MDS(m) code. Here $\mathcal{C}_{col}$ can be any single parity check MDS code.*

Proof. In [55] the above theorem is been proved by assuming $\mathcal{C}_{\text{col}}$ as simple parity check code. We can easily verify that it can be extended to any general single parity check MDS code. □

Theorem 4.4.4. *For the topology $T_{m,n}(a, b, 0)$ if the code $\mathcal{C} = \mathcal{C}_{col} \otimes \mathcal{C}_{row}$ is MR then $\mathcal{C}_{row}$ is an MDS($m - a + 1$) code and $\mathcal{C}_{col}$ is MDS($n - b + 1$) code.*

Proof. Suppose $\mathcal{C} = \mathcal{C}_{\text{col}} \otimes \mathcal{C}_{\text{row}}$ is MR code for $T_{m,n}(a, b, 0)$. Then from lemma 4.2.1 the code $\mathcal{C}' = \mathcal{C}_{\text{col}}|_{M_1} \otimes \mathcal{C}_{\text{row}}$ is MR code for $T_{m-a+1,n}(1, b, 0)$, here $M_1 \subset [m], |M_1| = m - a + 1$ and $\mathcal{C}_{\text{col}}|_{M_1}$ is the punctured code of $\mathcal{C}_{\text{col}}$. So, from the corollary 4.4.3 it is clear that $\mathcal{C}_{\text{row}}$ should be MDS($m - a + 1$) code.

Now by using lemma 4.2.3 we can say that $\mathcal{C}^* = \mathcal{C}_{\text{row}} \otimes \mathcal{C}_{\text{col}}$ is MR code for $T_{n,m}(b, a, 0)$. Similarly, we can prove that $\mathcal{C}_{\text{col}}$ is MDS($n - b + 1$) code.

□

Definition 13 (Projective Space ($PG(k - 1, q)$) [62]). *The $(k - 1)$-dimensional projective space over F_q is a set of points in which a point corresponds to a 1-dimensional subspace of k-dimensional vector space over $\mathbb{F}_q$ (say $VS(k, q)$). It is denoted by $PG(k - 1, q)$.*

In canonical form $PG(2, q) = \{(1, x_1, x_2)| \, x_1, x_2 \in F_q\} \cup \{(0, 1, x_2)| \, x_2 \in F_q\} \cup \{(0, 0, 1)\}$, every point is the representative of a 1-dimensional subspace in $VS(3, q)$, i.e. set of all scalar multiples of the point. And line joining of any two points in $PG(2, q)$, corresponds to 2-dimensional subspace in $VS(3, q)$. So, line contains the representatives of each 1-dimensional subspace of 2-dimensional subspace. The projective space $PG(2, q)$ has $q^2 + q + 1$ points and $q^2 + q + 1$ lines with each line containing $q + 1$ points and each point lying on $q + 1$ lines. In this view point we can state:

- Any three points are collinear in $PG(2, q)$ implies linear dependency among three vectors in $VS(2, q)$.

- Any three lines in $PG(2,q)$ are concurrent implies three 2-dimensional subspaces meets in $VS(2,q)$, which gives a 1-dimensional subspace. The concurrent point is the representative of that 1-dimensional subspace.

Theorem 4.4.5. *A code C with a generator matrix V is an $(n,3)$-MDS(3) code if and only if the columns of V satisfy the following conditions in $PG(2,q)$:*

i) No three points are collinear.

ii) Take any three lines which are formed by distinct pair of points. Such three lines are non concurrent.

Proof. Suppose the columns of V are the points in $PG(2,q)$ and holds the properties $i)$ and $ii)$. Now to prove that the code generated by V is $(n,3)$-MDS(3) code, we show that V fulfills the conditions given in the Theorem 4.4.2. Particularly for $l=3$, we get three cases which we prove one by one.

Case 1: $|A_1|=3, |A_2|=3$ and $|A_3|=0$.
Since $V_{A_3}=\emptyset$, it is obvious that $dim(V_{A_1}\cap V_{A_2}\cap V_{A_3})=0$.

Case 2: $|A_1|=3, |A_2|=2$ and $|A_3|=1$.
From property $i)$, it is clear that $V_{A_1}=F_q^3$ and $V_{A_2}\cup V_{A_3}=F_q^3$. Therefore, $dim(V_{A_2}\cap V_{A_3})=0$ and $dim(V_{A_1}+(V_{A_2}\cap V_{A_3}))=3$. Hence, $dim(V_{A_1}\cap V_{A_2}\cap V_{A_3})=dim(V_{A_1})+dim(V_{A_2}\cap V_{A_3})-dim(V_{A_1}+(V_{A_2}\cap V_{A_3}))=0$.

Case 3: $|A_1|=2, |A_2|=2$ and $|A_3|=2$.
From property $ii)$, it is clear that $V_{A_1}\cap V_{A_2}\cap V_{A_3}=\underline{0}$ for any pair-wise disjoint subsets A_i's. Therefore, $dim(V_{A_1}\cap V_{A_2}\cap V_{A_3})=0$.

Now we prove only if part. Assume V is the generator matrix for the $(n,3)$-MDS(3) code. For all the above three cases, we can see $dim(W_{A_1}\cap W_{A_2}\cap W_{A_3})=0$. For the case 1 and case 2 it is obvious. Now for the case 3, we use the formula given in the Theorem 4.4.1 to find the $dim(W_{A_1}\cap W_{A_2}\cap W_{A_3})$, which equals to

$$\sum_{i}^{3} \min(|A_i|, 3) - \sum_{1 \le i < j \le 3} \min(|A_i| + |A_j|, 3)$$
$$+ \min(|A_1| + |A_2| + |A_3|, 3) = 3.2 - 3.3 + 3 = 0.$$

So from the theorem 4.4.2, for every case $dim(V_{A_1} \cap V_{A_2} \cap V_{A_3}) = 0$.

We prove the non collinear property $i)$ by using the specific case 2. Since V is full rank, We can pick A_1 with $|A_1| = 3$ such that $dim(V_{A_1}) = 3$. Therefore, for any A_2, A_3 we get $dim(V_{A_1} + (V_{A_2} \cap V_{A_3})) = 3$. Since $dim(V_{A_1} \cap V_{A_2} \cap V_{A_3}) = 0$, we get $dim(V_{A_2} \cap V_{A_3}) = 0$. Which implies $dim(V_{A_2} \cup V_{A_3}) = 3$. Therefore, Any three points in $PG(2, q)$ are non collinear.

Now we use case 3 to prove non concurrent property $ii)$. In this case all A_i's are pair-wise disjoint with cardinality 2. Since $dim(V_{A_1} \cap V_{A_2} \cap V_{A_3}) = 0$, We can say $V_{A_1} \cap V_{A_2} \cap V_{A_3} = \underline{0}$. Hence, there exist no point in $PG(2, q)$ which lies on all these three lines formed by points with the corresponding columns of V.

□

Note 5. *For the construction of* $(5, 3)$*-MDS(*3*) code, the generator matrix which satisfies the non-collinear property is sufficient. Since we have only 5 columns, the non-concurrent condition doesn't arise. Therefore, for* $n = 5$ *any MDS code is MDS(*3*).*

4.5 Conclusion

In this chapter, we have discussed shortening and puncturing properties of MRPCs. We have expanded the class of correctable and incorrectable erasure patterns for general product topologies. We also constructed a bipartite graph between a subset of rows of erasures and non-erasures in a disjoint subset of rows. We proved that for a row-wise irreducible, regular erasure pattern, there exists a complete matching in this graph. For the case of $a = 1$, we constructed another bipartite graph between rows and columns of erasure sub-patterns and proved a certain matching condition property of this graph. We gave an alternate proof of the sufficiency of regularity for $a = 1$ case. We considered the generator matrix G of the product code and expand it as tensor product $G_{\text{col}} \otimes G_{\text{row}}$ of generator matrices of column and row codes. We proved that a certain square submatrix of this tensor product is full rank, by applying the properties of bipartite graphs which we derived. We consider a subset of regular erasure patterns for the case of $a = 2$, which are obtained by extending regular erasure patterns for $a = 1$. We prove that these regular erasure patterns are also recoverable. Also, we have characterized $(n, 3)$-MDS(3) codes in terms of the points/lines of the associated projective space. Its part of ongoing work to use this characterization to construct $(n, 3)$-MDS(3) codes.

Chapter 5

Sparse and Balanced Generator Matrix for the Local MRCs

In this chapter, we study sparse and balanced generator matrices of codes. A generator matrix of a code is said to be sparse if each row of the generator matrix has minimal weight. The advantage of sparse generator matrix is that in the event of a message symbol update, the number of code symbols that need to be updated is minimum. A generator matrix is said to be balanced if the weights of the columns of the generator matrix are all equal or differ at most by one. A balanced generator matrix offers the benefit of approximately the same computation time for all the code symbols. We present sparse generator matrices for MRC with locality for single erasure and also sparse and balanced generator matrices for MRC with locality parameter 2 for a large set of parameters.

Section 5.1 briefly reviews the Sparse and Balanced (SB) generator matrices of MDS codes and LRCs. In Section 5.2, focuses on the construction of sparse matrix and balance matrix of Local MRCs for single erasure by using the generator matrix of well-known PMDS codes. Lastly, Section 5.3 presents the conclusion of the chapter.

5.1 Review of SB Generator Matrix for MDS Codes and LRCs

In this section, we review the results of [63]. Firstly, we will define a sparse and balanced MDS code and state a result about the existence of these codes over a sufficiently large field ($q > \binom{n-1}{k-1}$).

Definition 17 (Sparse and Balanced (SB) MDS code:). *An $[n, k, d]_q$ MDS code is called sparse and balanced MDS code. If the generator matrix $G_{k\times n}$ satisfies following two conditions:*

(1) Sparse: each row of G has minimum Hamming weight $d = n - k + 1$.

(2) Balanced: Hamming weights of any two columns of G are equal or differ by one.

Codes with this structure minimize the maximal computation time of computing any code symbol. Because, sparsity minimizes the no. of nodes required to update with respect to a modified data symbol in the data. Balanced property assures almost uniform load distribution on each node.

The following Theorem 5.1.1 states the necessary conditions regarding the support constraints on any set of rows of a generator matrix in order to generate MDS codes. Initially, it was conjectured in [64] by Dau et al. and subsequently verified for all $k \leq 7$. Later, the conjecture was proven in [65].

Theorem 5.1.1. *If the specified supports of the generator matrices (M) satisfy the so-called MDS condition:*

$$| \cup_{i\in I} \mathit{supp}(M_i)| \geq n - k + |I|. \tag{5.1}$$

for all non empty subsets $I \subseteq \{1, 2, \ldots, k\}$, where $\mathit{supp}(M_i) = \{j|1 \leq j \leq n, m_{ij} \neq 0\}$ is the support of ith row of M. Then for every prime power $q \geq n + k - 1$, there exists an $[n, k]_q$ MDS code whose generator matrix G fits with M.

A way of Explicit construction of *SB Reed-Solomon* codes from the cyclic RS code is shown in [66] with the condition $\frac{k}{n}(n - k + 1)$ is an integer and later provided the same by relaxing the condition.

Definition 18. *$w-$ Balanced Matrix: A matrix $B_{k\times n}$ is called $w-$balanced if the following conditions hold:*

- *Every row of B has the same weight w.*
- *Every column is of weight $\left\lceil \frac{kw}{n} \right\rceil$ or $\left\lfloor \frac{kw}{n} \right\rfloor$*

In [67], the authors presented a method that produces a w-balanced generator matrix for a given full-length cyclic RS code. In particular, each row is a codeword of weight w, such that $d \leq w \leq n-1$. Also provided a way of construction for a given full-length cyclic LRC with locality r (given in [5], generally known as Tamo-Barg codes), where $(\frac{n}{r+1} - \frac{k}{r} - 1)(r+1) \leq w \leq (\frac{n}{r+1} - 1)(r+1)$.

5.2 Sparse Matrix and Balanced Matrix of Local MRCs

In this section, we discuss the properties of a balanced generator matrix and also provide the construction of a sparse and balanced generator matrix for local MRC with $r = 2$ and $\delta = 2$. Before that, for readability, we recall the construction of Local MRCs, which is given in Subsection 2.2.3 (see before Table 2.3). The MRD property of Gabidulin code is used to construct the generator matrix of Local MRCs, which has the following structure:

$$G = G_{GC} \left[\begin{array}{c|c|c|c} G_{MDS} & 0 & \cdots & 0 \\ \hline 0 & G_{MDS} & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & G_{MDS} \end{array}\right] \tag{5.2}$$

Here, G_{GC} is the generator matrix of Gabidulin code $\mathscr{C}_{GC}[n_1 = lr, k, d_1 = n_1 - k + 1]$ over the field F_{q^M} and G_{MDS} is the generator matrix of MDS code $\mathscr{C}_{MDS}[n_2 = r+m, r, d_2 = m+1]$ over the field F_q.

5.2.1 Sparse and Balanced Local MRCs

To understand the construction of SB matrices, it needs to rewrite the necessary condition on support constraints of the generator matrix to be MDS code given in (5.1) with respect to zeros for $G \in F_{q^M}^{k\times n}$ as follows:

Support Constraints of G [68]: Support constraints of the generator matrix $G \in F_{q^M}^{k\times n}$ described through the subsets $Z_1, Z_2, \ldots, Z_k \subset [n]$ as follows:

$$\forall i \in [k], \forall j \in Z_i, G_{ij} = 0 \tag{5.3}$$

A necessary condition for a code to be MDS is

$$\mid \cap_{i\in\Omega} Z_i \mid + \mid \Omega \mid \leq k, \; \forall\, \Omega \neq \emptyset \subset k \tag{5.4}$$

Theorem 5.2.1 ([68])**.** *For any $M \geq \max\{n,\, k-1+\log_q k\}$, if (5.4) is satisfied, then there exists a Gabidulin code in F_{q^M} of length n and dimension k such that its generator matrix satisfies the support constrains in (5.3).*

Theorem 5.2.2. *Suppose $\mathscr{C}_{GC}[n_1 = k+h, k, d_1 = h+1]$ Gabidulin code over F_{q^M}, where $M \geq \{k+h, k-1+\log_q k\}$ and $\mathscr{C}_{MDS}[n_2 = r+1, r, d_2 = 2]$ is systematic MDS code and the corresponding generator matrices are G_{GC} and G_{MDS}. For any $r \geq 2$ which divides k, h and $(r+1)$ divides n, there exists $d = h + \frac{h}{r} + 2$-sparse generator matrix for the Local MRC of length $k + h + (\frac{k+h}{r})$ and dimension k.*

Proof. From the Theorem 5.2.1 provided there exist a Gabidulin code over F_{q^M}, if it satisfies the MDS condition (5.4). So, it is possible to have d_1-sparse generator matrix for Gabidulin code say G_{SGC} which can have non zero entries in every j^{th} row with the column indices from $j-1$ to $j-1+h$. Now replace G_{GC} with G_{SGC} in the equation (5.2), then it is clear that weight of each row of G_{GSC} is distributed up to $(\frac{h+1}{r})$ number of G_{MDS}'s in G. Therefore, weight of every row in G is equal to $(\frac{h}{r}+1)+h+1 = d$.

□

Now we discuss the construction of $w-$balanced mask matrix given in [69], which is used to identify the locations of zeros to be placed in order to construct the $w-$balanced generator matrix.

Construction 5.2.3 (w-Balanced Mask Matrix of size $k \times n$)**.** *Let both k and w be strictly less than n. Define the quantities $g := \gcd(w, n)$, $\eta := \frac{n}{g}, \phi = \lfloor \frac{k}{n} \rfloor$ and $\rho = k - \eta\phi$. Define the* *index sets*

$$\mathcal{I}_1 = \{jw + i : 0 \le j \le \eta - 1, 0 \le i \le \phi - 1\},$$
$$\mathcal{I}_2 = \{jw + \phi : 0 \le j \le \rho - 1\},$$

and $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2$. *The matrix* A *whose rows are given by* $\{a_l : l \in \mathcal{I}\}$. *Here* a *is the vector of the length* n *which has consecutive ones at first* w *locations and rest of them are zeros, i.e.* $a = (1, \dots, 1, 0, \dots 0)$ *and* a_l *denote the right cyclic shift of* a *by* l *positions.*

Lemma 5.2.4. *Let* A *be any* w*- balanced mask matrix constructed from the procedure mentioned in construction 5.2.3 with the index set* $\mathcal{I}$*. For any two consecutive columns of* A *say* $A|_i$ *and* $A|_{i+1}, 0 \le i \le n_1 - 2$,

$$|\text{supp}(A|_i) \cap \text{supp}(A|_{i+1})| \ge \xi - 1 \tag{5.5}$$

where $\xi = \max\{|\text{supp}(A|_i)|, |\text{supp}(A|_{i+1})|\}$

Proof. WLOG, assume $\xi = |\text{supp}(A|_i)|$. Suppose $|\text{supp}(A|_i) \cap \text{supp}(A|_{i+1})| = |\text{supp}(A|_i)| - \gamma$, where $\gamma \ge 2$. It indicates that there exists γ number of rows say $j_1, j_2, \dots, j_\gamma$ in A, which has γ number of zeroes in $A|_i$ and γ number of ones in $A|_{i+1}$. Therefore, from construction of A we can say there exists $l_1, l_2, \dots, l_\gamma \in \mathcal{I}$ such that $l_1 \mod n_1 = l_1 \mod n_1 = \dots = l_\gamma \mod n_1 = i$. So $j_1, j_2, \dots, j_\gamma$ rows of A are identical. This contradicts the property that the rows of A are pairwise distinct. Similarly, we can prove for $\xi = \max\{|\text{supp}(A|_{i+1})|$. □

Theorem 5.2.5. *Suppose* G_{BGC} *is the* w_1*-balanced generator matrix of* $G_{GC} = T^{-1}G_{BGC}$*, where* T *is the invertible square matrix of size* $k \times k$*. Let* $A_{BGC} \in \{0,1\}^{k \times n_1}$ *is the corresponding mask matrix. For any* $w_1 = h + \delta$ *where* $\delta \ge 1$ *is an integer* ≥ 1*,* $g_1 = \gcd(w_1, n_1 = k + h), \eta_1 = \frac{n_1}{g_1}$ *divides* k *and* n_1, k *are even integers. If* G_{MDS} *is the systematic generator matrix of size* 2×3 *then*

1. *for any odd integer* δ*, the generator matrix of Local MRC* G *given in* (5.2) *can form* $w = \frac{3(h+\delta)+1}{2}$ *- balanced generator matrix.*

2. *for any even integer δ and $\phi = \frac{k}{\eta_1} = 1$ the Local MRC generator matrix G given in* (5.2) *is $w_1$balanced matrix.*

Proof. Suppose $G_{BGC} = \begin{bmatrix} \underline{g}_0 & \underline{g}_1 & \cdots & \underline{g}_{n_1-1} \end{bmatrix}$ and with the given $G_{MDS} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, the generator matrix G of Local MRC obtained from equation (5.2) can be viewed as $[G_1|G_2|\dots|G_{l-1}]$. Hence, $G_i = \begin{bmatrix} \underline{g}_{2i} & \underline{g}_{2i+1} & \underline{p}_i = \underline{g}_{2i} + \underline{g}_{2i+1} \end{bmatrix}$ for all $i \in \{0, 1, \dots, l-1\}$.

Proof of 1): Since, G_{GBC} is the w_1-balanced matrix with the corresponding mask matrix, weight of any row in G is $w = w_1 + \frac{w_{1-1}}{2} + 1 = \frac{3(h+\delta)+1}{2}$.

To show G is w-balanced matrix, we need to show that the weight of the any column of G is either $\lfloor \frac{wk}{n} \rfloor$ or $\lceil \frac{wk}{n} \rceil$. Since, $\eta_1 = \gcd(w_1, n_1) \mid n_1$, $\rho = k - \eta_1 \frac{k}{\eta_1} = 0$. i.e. weight of any column in $G_{GBC} = \frac{w_1 k}{n_1}$.

$$\begin{aligned} |\text{supp}(\underline{p}_i)| &= |\text{supp}(\underline{g}_{2i}) \cup \text{supp}(\underline{g}_{2i+1})| \\ &= 2|\text{supp}(\underline{g}_{2i})| - |\text{supp}(\underline{g}_{2i}) \cap \text{supp}(\underline{g}_{2i+1})| \\ &\leq |\text{supp}(\underline{g}_{2i})| + 1 \end{aligned}$$

Inequality in the above follows from the Lemma 5.2.4. Therefore, weight of any column in G is either $\frac{w_1 k}{n_1}$ or $\frac{w_1 k}{n_1} + 1$.

Note 6. *The difference $D = \frac{wk}{n} - \frac{w_1 k}{n_1} \in (0, 1)$. It can be easily verified as follows:*

$$D = \frac{\left[\frac{3(h+\delta)+1}{2}\right] k}{\frac{(k+h)3}{2}} - \frac{(h+\delta)k}{(k+h)} = \frac{k}{3(k+h)} < 1$$

From the above note, it is clear that $\frac{w_1 k}{n_1} = \lfloor \frac{wk}{n} \rfloor$ and $\frac{w_1 k}{n_1} + 1 = \lceil \frac{wk}{n} \rceil$. Hence, the balanced property of the G has been satisfied.

proof for 2: Since w_1 is even integer the difference $D = 0$. Therefore G is balanced only if every column of G should have the weight $\frac{w_1 k}{n_1}$. Note that every column of $G_{BGC} = \frac{w_1 k}{n_1}$, because ϕ_1 is integer. For any $0 \leq i \leq n_1 - 1$

$$\begin{aligned} |\text{supp}(\underline{p}_i)| &= |\text{supp}(\underline{g}_{2i}) \cup \text{supp}(\underline{g}_{2i+1})| \\ &= 2|\text{supp}(\underline{g}_{2i})| - |\text{supp}(\underline{g}_{2i}) \cap \text{supp}(\underline{g}_{2i+1})| \\ &= |\text{supp}(\underline{g}_{2i})| \end{aligned}$$

In the above last equality holds because $|\text{supp}(\underline{g}_{2i})| = |\text{supp}(\underline{g}_{2i+1})|$. Since any even number modulo with even number is even.

Note 7. *We observed that for any even integer δ and $\phi = \frac{k}{\eta_1} > 1$ the Local MRC generator matrix G given in* (5.2) *can't be the balanced matrix through lot of examples and proof is left as future work.*

□

Now, we illustrate our construction in the proof of Theorem 5.2.5 through an example.

Example 5. *Let Gabidulin code $\mathscr{C}_{GC}[n_1 = 24, k = 16, d_1 = 9]$. Suppose G_{BGC} is the $w_1 = 9$-balanced generator matrix of $G_{GC} = T^{-1}G_{BGC}$, where T is the invertible square matrix of size $k \times k$. Let $A_{BGC} \in \{0,1\}^{k \times n_1}$ is the corresponding mask matrix. $w_1 = h + 1 = 8 + 1$, here $\delta = 1$ is an integer, $g_1 = \gcd(w_1, n_1 = k + h) = (9, 24) = 3, \eta_1 = \frac{n_1}{g_1} = \frac{24}{3} = 8$ divides $k = 16$ and n_1, k are even integers. If G_{MDS} is the systematic generator matrix of size 2×3 then for odd integer $\delta = 1$, the generator matrix of Local MRC G given in* (5.2) *can form $w = \frac{3(h+\delta)+1}{2} = \frac{28}{2} = 14$ - balanced generator matrix.*

5.3 Conclusion

This chapter provided a study of sparse and balanced matrices and their properties, along with a brief review of sparse and balanced matrices for MDS codes and Gabidulin codes. By considering generator matrices of PMDS code, we proved that the sparse $(d = h + \frac{h}{r} + 2)$ generator matrices always exist for any $r \geq 2$, which divides k, h and $(r + 1)$ divides n. We also proved for $r = 2$, if $\eta_1 = \frac{n_1}{g_1}$ divides k and n_1, k are even integers, there exist $w_1 = (h+\delta)$-balanced generator matrix for Local MRC specific to the cases for any odd integer δ and for any even integer δ whenever $k = \eta_1$.

Chapter 6

H-LRC Codes with Availability

In this chapter, we explore the concept of locally recoverable codes with availability. The concept of availability in LRCs refers to the ability to have multiple repair sets to repair a single node. Having multiple repair sets in LRCs is advantageous because it allows for the distribution of the repair load among different nodes. In other words, when a node in the storage system fails or loses data, instead of burdening a subset of nodes with the repair process, the load can be divided among several nodes. This helps in balancing the workload and avoiding excessive download of repair data from the specific nodes that are accessed more frequently.

We begin the chapter with a brief literature review on the upper bound on the minimum distance for both single and multiple erasures. We extend the concept of availability to LRCs with hierarchical locality. We will refer to these codes as Hierarchical Locally Recoverable Codes (H-LRCs) with availability. Our study investigates the upper bound on the minimum distance for certain range of parameters of H-LRCs with availability.

Section 6.1 offers a concise review of LRCs with availability. Section 6.2 provides a definition of H-LRC code with availability and the Singleton-like bound for the code. Lastly, Section 6.3 presents the Conclusion of the chapter.

6.1 Review of LRCs with Availability

The study by Pamies et al. [70] explores utilizing multiple repair sets to recover from a single erasure. In this work, alternative repair options are possible by slightly increasing the number of contacted nodes during the repair.

Additionally, the studies discussed in [71] and [72] explore the trade-offs between code rate and minimum Hamming distance in relation to the desirable properties of locality and availability in codes. The paper by Wang et al. [71] proposed the $(r, \delta)_c$ -locality providing $\delta - 1$ non overlapping local repair groups of size no more than r for a coordinate (by modifying the structure of (r, δ) LRCs) and derived the upper bound on the minimum distance. Later on, The paper [6] presented the upper bound on the minimum distance d for locally recoverable codes (LRCs) with (r, τ)-Availability. This bound is given as follows:

$$d \leq (n - k + 1) - \left(\left\lceil \frac{\tau(k-1)+1}{\tau(r-1)+1} \right\rceil - 1 \right). \tag{6.1}$$

Note that the upper bound on the minimum distance obtained in [71] is similar to the (6.1) with $\tau = \delta - 1$.

LRCs with availability are also studied for non-uniform localities in [73] and the upper bound on d is as follows:

$$d \leq (n - k + 1) - \left(\left\lceil \frac{\tau(k-1)+1}{\sum_{i=1}^{\tau}(r_i - 1) + 1} \right\rceil - 1 \right). \tag{6.2}$$

6.2 Singleton-like Bound for H-LRC Code with Availability

In the paper [35], the concept of availability has been extended to codes with (r, δ) locality and codes with hierarchical locality. However, the authors have introduced the aforementioned definitions for non-uniform localities and constructed codes from curves strictly when the localities are non-uniform (their constructions are not applicable for uniform locality case). Now, we provide the formal definition for codes with (r, δ) locality and availability and the upper bound on the minimum distance given in [74].

Definition 14 (LRC code with (r, δ) locality and availability [74])**.** *A linear code $\mathcal{C}$ is an LRC with (r, δ) locality and availability τ, if for any $i \in [n]$ there are τ punctured codes $\mathcal{C}_1^{(i)}, \mathcal{C}_2^{(i)}, \ldots, \mathcal{C}_\tau^{(i)}$ such that for all $p \in [\tau]$ the following conditions hold:*

i) $i \in supp(\mathcal{C}_p^{(i)})$,

ii) $dim(\mathcal{C}_p^{(i)}) \leq r$,

iii) $d_{min} \geq \delta$ *and*

iv) $R_p^{(i)} \cap R_q^{(i)} = \{i\}$ *for all* $p \neq q$. *Here* $R_p^{(i)} = supp(\mathcal{C}_p^{(i)})$ *is called the repair set for the* i^{th} *coordinate.*

The upper bound on the minimum distance is provided as follows in equation (6.3) [74]:

$$d \leq (n-k+1) - \left(\left\lceil \frac{\tau(k-1)+1}{\tau(r-1)+1} \right\rceil - 1 \right) (\delta - 1). \tag{6.3}$$

Let $N_p^{(i)} = R_1^{(i)} \cup \ldots \cup R_p^{(i)}$, for any $p \in [\tau]$. The union of all repair sets for the index i is denoted by $R^{(i)}$ then $R^{(i)} = N_\tau^i$ and let $\mathcal{C}^{(i)}$ be the code $\mathcal{C}$ restricted to $R^{(i)}$, i.e., $\mathcal{C}^{(i)} = \mathcal{C}|_{R^{(i)}}$.

Let I be an information set if $i \in I$, then

$$rank(N_p^{(i)}) \leq 1 + p(r_2 - 1). \tag{6.4}$$

Now, we extend the concept to codes with hierarchical locality.

Definition 15 (H-LRC code with availability)**.** *Let $\tau_1, \tau_2 \geq 1$, $\delta_2 < \delta_1$ and $r_2 < r_1$. A linear code with hierarchical locality $\mathcal{C}[n, k, d, r_1, r_2, \delta_1, \delta_2]$ is H-LRC with availability τ_1, τ_2 if*

- *The code has locality (r_1, δ_1) and availability τ_1. Let $\mathcal{C}_1^{(j)}, \mathcal{C}_2^{(j)}, \ldots, \mathcal{C}_{\tau_1}^{(j)}$ be the punctured codes for any $j \in [n]$ and corresponding repairs sets are denoted by $R_1^{(j)}, R_2^{(j)} \ldots R_{\tau_1}^{(j)}$. From Definition 14, $R_p^{(j)} \cap R_{p'}^{(j)} = \{j\}$ for all $p \neq p'$, $R^{(j)} = \bigcup_{p=1}^{\tau_1} R_p^{(j)}$, and $\mathcal{C}^{(j)} = \mathcal{C}|_{R^{(j)}}$.*

- *For any $p \in [\tau_1]$ the code $\mathcal{C}_p^{(j)}$ is LRC code with (r_2, δ_2) locality and availability τ_2. For any $i \in R_p^{(j)}$, let $\mathcal{C}_{p,1}^{(j,i)}, \mathcal{C}_{p,2}^{(j,i)}, \ldots, \mathcal{C}_{p,\tau_2}^{(j,i)}$ be the punctured codes, and let $R_{p,1}^{(j,i)}, R_{p,2}^{(j,i)}, \ldots R_{p,\tau_2}^{(j,i)}$ be their corresponding repair sets. From the definition 14, notice that $R_{p,q}^{(j,i)} \cap R_{p,q'}^{(j,i)} = \{i\}$ for all $q \neq q'$. Let $R_p^{(j,i)}$ is the union of all the repair sets of i^{th} coordinate, i.e., $R_p^{(j,i)} = \bigcup_{q=1}^{\tau_2} R_{p,q}^{(j,i)}$ and let $\mathcal{C}_p^{(j,i)}$ be the code restricted to $R_p^{(j,i)}$, i.e., $\mathcal{C}_p^{(j,i)} = \mathcal{C}^{(j)}|_{R_p^{(j,i)}}$. We can also denote for any $i \in R^{(j)}$, $\mathcal{C}^{(j,i)} = \mathcal{C}^{(j)}|_{R^{(j,i)}}$, where $R^{(j,i)} = \bigcup_{p=1}^{\tau_1} R_p^{(j,i)}$, which we use in the Algorithm 1 to prove the following theorem.*

Theorem 6.2.1. *Let $\mathcal{C}[n, k, d, r_1, r_2, \delta_1, \delta_2]$ be an H-LRC code with availability and if $\tau_1 = \tau_2 = \tau$ then*

$$d \leq (n-k+1) - \left(\left\lceil \frac{\tau(k-1)+1}{\tau(r_2-1)+1} \right\rceil - 1 \right) (\delta_2 - 1) - \left(\left\lceil \frac{\tau(k-1)+1}{\tau(r_1-1)+1} \right\rceil - 1 \right) (\delta_1 - \delta_2). \quad (6.5)$$

Proof. In order to prove the theorem, we identify a punctured code $\mathcal{C}_s$ of $\mathcal{C}$, which has dimension $k-1$ with the largest support, by combining and developing the techniques used in [codes with hierarchical locality] and [repair locality with multiple erasure tolerance]. Then we will use the fact given in Lemma 1, i.e., $d = n - |supp(\mathcal{C}_s)|$.

The algorithm referred to as Algorithm 1 is utilized for the purpose of finding $\mathcal{C}_s$. During each iteration j, the algorithm detects a Middle Code with Availability (MCA), say $M_j = \mathcal{C}^{(j')}$ for some $j' \in [n]$ in $\mathcal{C}$ with repair set $R^{(j')}$, which gains additional rank to the rank of the previous iteration $j-1$. In step 3, Local Code with Availability (LCA) $L_i = \mathcal{C}^{(j',i')}$ is picked up at each iteration indexed by i, within the specified MCA (identified in step 2) with repair set $R^{(j',i')}$, which gains additional rank to the rank of the previous iteration $i-1$. It is evident that the algorithm's termination becomes apparent when the overall rank is bounded by k. The variables i_{end} and j_{end} correspond to the final values of i and j, respectively. The support of L_i is denoted by S_i and V_i represents the column space of the matrix $G|_{Si}$. If no further LCA is added from the MCA M_j, the support of the last added LCA is removed, and an additional support T_j of M_j is included in ψ (step 8).

The incremental rank is denoted by a_i and the incremental support by s_i while adding an LCA L_i. For all $i \in [i_{end}]$, it holds that $s_i \geq a_i + \tau_2(\delta_2 - 1)$. Since,

Algorithm 1 For the proof of Theorem 6.2.1

1: $j = 0, i = 0, W = \emptyset, \psi = \emptyset$
2: **while** $\exists$ a middle code with availability M_j (with repair set $R^{(j')}$) in $\mathcal{C}$ such that $rank(G|_{\psi \cup R^{(j')}}) > rank(G|_{\psi})$ **do**
3:     **while** $\exists$ a local code with availability L_i in M_j such that $V_i \subsetneq W$ **do**
4:         $W = W + V_i$
5:         $\psi = \psi \cup S_i$
6:         $i = i + 1$
7:     **end while**
8:     $\psi = (\psi \setminus S_{i-1}) \cup T_j$
9:     $j = j + 1$
10: **end while**

- a_i is always greater than zero in each iteration (as per the condition given in step 3),
- Among all τ_2 local codes in L_i, any two repair sets have only one element in common, which is the coordinate of the specified LCA L_i, say $i' \in [n]$ and
- Each local code has minimum distance δ_2.

Let $i(j)$ denote the index of the last LCA added from M_j. There are τ_1 number of middle codes in MCA, and any two repair sets have only one common element for any M_j say $j' \in [n]$. Every rank accumulating local code brings at least one new information symbol. Thus, for all $j \in [j_{end}]$, $t_j = |T_j| \geq a_{i(j)} + \tau_1(\delta_1 - 1) = a_{i(j)} + \tau_1[(\delta_2 - 1) + (\delta_1 - \delta_2)]$.

By observing the algorithm and from (6.4), it is evident that $(k-1) \leq \sum_{h=1}^{i_{end}-1} rank(G|_{s_h}) = (i_{end} - 1)(1 + \tau_2(r_2 - 1))$. Consequently, the inequality

$$(i_{end} - 1) \geq \left\lceil \frac{k-1}{\tau_2(r_2 - 1) + 1} \right\rceil \tag{6.6}$$

is established. Similarly, $(k-1) \leq \sum_{h=1}^{j_{end}-1} rank(G|_{R^{(h)}}) = (j_{end} - 1)(1 + \tau_1(r_1 - 1))$. Therefore,

$$(j_{end} - 1) \geq \left\lceil \frac{k-1}{\tau_1(r_1 - 1) + 1} \right\rceil \tag{6.7}$$

Since, after adding the LCA $L_{i_{end}-1}$, we would have accumulated rank that is less than or equal to $(k-1)$. Hence, we can always pick $s_e = (k-1) - \sum_{i=1}^{i_{end}-1} a_i$ and note that $s_e \geq 0$. The resultant punctured code is identified as $\mathcal{C}_s$. Let $E = \{i(j) \,|\, 1 \leq j \leq j_{end}\}$. Then

$$|Supp(\mathcal{C}_s)| \geq \sum_{i \notin E, i=1}^{i_{end}-1} s_i + s_e + \sum_{j=1}^{j_{end}-1} t_j \tag{6.8}$$

In (6.8) the last term $\sum_{j=1}^{j_{end}-1} t_j$ includes a sum of only $(j_{end}-1)$ terms because we could have possibly accumulated a rank of $(k-1)$ after adding $L_{i_{end}} - 1$, i.e., $s_e = 0$. Thus, we have,

$$\begin{aligned}
|Supp(\mathcal{C}_s)| &\geq \sum_{i \notin E, i=1}^{i_{end}-1} s_i + (k-1) - \sum_{i=1}^{i_{end}-1} a_i + \sum_{j=1}^{j_{end}-1} t_j \\
&\geq \sum_{i \notin E, i=1}^{i_{end}-1} (a_i + \tau_2(\delta_2 - 1)) + (k-1) - \sum_{i=1}^{i_{end}-1} a_i + \sum_{j=1}^{j_{end}-1} (a_{i(j)} + \tau_1(\delta_2 - 1) + \tau_1(\delta_2 - \delta_1)) \\
&= \sum_{i=1}^{i_{end}-1} \tau(\delta_2 - 1) + (k-1) + \sum_{j=1}^{j_{end}-1} \tau(\delta_1 - \delta_2) \qquad (\because \tau_1 = \tau_2 = \tau) \\
&= (k-1) + (i_{end}-1)\tau(\delta_2 - 1) + (j_{end}-1)\tau(\delta_1 - \delta_2) \\
&\geq (k-1) + \left(\left\lceil \frac{k-1}{\tau(r_2-1)+1} \right\rceil \right) \tau(\delta_2 - 1) + \left(\left\lceil \frac{k-1}{\tau(r_1-1)+1} \right\rceil \right) \tau(\delta_1 - \delta_2)
\end{aligned}$$

The last inequality is obtained by substituting the values of $i_{end} - 1$ from (6.6) and $j_{end} - 1$ from (6.7). We know that from Lemma 1,

$$\begin{aligned}
d &= n - |Supp(\mathcal{C}_s)| \\
&\leq (n-k+1) - \left(\left\lceil \frac{k-1}{\tau(r_2-1)+1} \right\rceil \right) \tau(\delta_2 - 1) - \left(\left\lceil \frac{k-1}{\tau(r_1-1)+1} \right\rceil \right) \tau(\delta_1 - \delta_2) \\
&\leq (n-k+1) - \left(\left\lceil \frac{\tau(k-1)+1}{\tau(r_2-1)+1} \right\rceil - 1 \right) (\delta_2 - 1) - \left(\left\lceil \frac{\tau(k-1)+1}{\tau(r_1-1)+1} \right\rceil - 1 \right) (\delta_1 - \delta_2),
\end{aligned}$$

where the last inequality is obtained by using the following inequalities:

$$\left(\left\lceil\frac{k-1}{\tau(r_2-1)+1}\right\rceil\right)\tau \geq \left(\frac{\tau(k-1)}{\tau(r_2-1)+1}\right)$$
$$\geq \left(\left\lfloor\frac{\tau(k-1)}{\tau(r_2-1)+1}\right\rfloor\right) = \left(\left\lfloor\frac{\tau(k-1)+1}{\tau(r_2-1)+1}\right\rfloor\right) - 1$$

and (we can replace r_2 with r_1)

$$\left(\left\lceil\frac{k-1}{\tau(r_1-1)+1}\right\rceil\right)\tau \geq \left(\left\lfloor\frac{\tau(k-1)+1}{\tau(r_1-1)+1}\right\rfloor\right) - 1.$$

□

The following figure (Fig. 6.1) depicts the comparison between upper bounds on all minimum distances of all the topologies discussed till now in this chapter. i.e., codes with locality and availability for single erasure, multiple erasures, and hierarchical locality. In this, we considered fixed values for $(n = 60, r_1 = 9, r_2 = 5, \delta_1 = 4, \delta_2 = 2, \tau = 4)$. We took 10 samples by decreasing the values of k from 30 to 21. Plotted the graph code rate (k/n) versus relative distance (d/n).

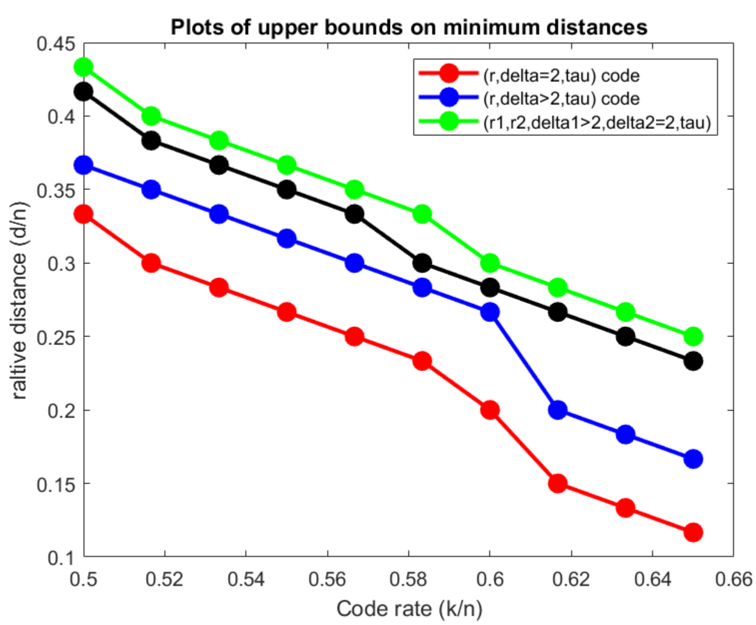

Figure 6.1: Comparison of upper bounds on d

6.3 Conclusion

This chapter has focused on the definition of the H-LRC code with availability, incorporating uniform localities for both local codes and middle codes within the code structure. The derived upper bound on the minimum distance of the H-LRC code with availability has been determined for the case when τ_1 is equal to τ_2. However, obtaining a closed form expression for the case when τ_1 is not equal to τ_2 appears to be a challenging task and is left for future work.

Chapter 7

Conclusions and Future Directions

7.1 Summary and Conclusions

In this thesis, we have studied Distributed Storage Systems (DSS) and their importance. Specifically focusing on the challenge of data recovery, we have looked into different types of topologies that have evolved to address this challenge by providing low repair degrees for different types of erasures, including single, multiple, and correlated failures. In the context of these topologies, we have examined the properties, constructions and bounds of Maximally Recoverable Codes (MRCs).

Our research has specifically focused on the challenges and problems associated with Locally Recoverable Codes (LRCs) for multiple erasures. Within the domain of LRCs, we have addressed four key aspects.

- The first aspect of our work involves the construction of Hierarchical Local MRCs (HL-MRCs) with improved field sizes compared to previously known constructions. We have also determined the lower bound on the required field size for HL-MRCs.
- The second aspect of our research is related to Maximally Recoverable Product Codes (MRPCs). We identifying the open problem of identifying the set of all erasure patterns which can be recovered in MRPCs. To address this challenge, we have characterized certain classes of recoverable and non-recoverable erasure patterns by deriving properties of MRPCs. Furthermore, we have delved into higher-order Maximum Distance Separa-

ble (MDS) codes and their interpretation in projective spaces, specifically focusing on MRPC construction for the case where $a = 1$.

- The third aspect of our work was dedicated to finding sparse and balanced Local MRCs for single erasure scenarios. The objective here is to reduce the encoding complexity associated with these codes, thereby facilitating more efficient data storage and recovery processes.

- Finally, we extended the concept of availability in LRCs to codes with hierarchical locality. By investigating Hierarchical Locally Recoverable Codes (H-LRC) with availability, we provided an upper bound on the minimum distance for the case $\tau_1 = \tau_2$.

7.2 Future Scope

- While this thesis has made progress in constructing Hierarchical Locally Repairable MRCs (HL-MRCs) with improved field sizes, there is still room for further advancements. Future research can focus on exploring alternative construction techniques and algorithms to achieve even better field sizes, ultimately enhancing the efficiency and performance of HL-MRCs.

- This thesis evidently states that there are no clear conditions that guarantee the recoverability of erasures in product topologies. Future research can focus on investigating this area further to determine the conditions expanding the set of recoverable erasure patterns.

- While sparse and balanced generator matrices for Local MRCs for single erasure have been explored in this thesis, future investigations can focus on sparse and balanced generator matrices for Local MRCs for multiple erasures.

- The extension of availability concept to LRCs with hierarchical locality is presented in this thesis. The upper bound on the minimum distance of the H-LRC code with availability has been determined for the case when τ_1 is equal to τ_2. Future studies can further explore to find the upper bound on the minimum distance for the general case and also investigate optimal code designs.

Related publications

Journals :

1. **D. Shivakrishna**, Aaditya M Nair, V. Lalitha, ”Maximally Recoverable Codes with Hierarchical Locality: Constructions and Field-Size Bounds,” in *IEEE Transactions on Information Theory*, vol. 69, no. 2, pp. 842-859, Feb. 2023, doi: 10.1109/TIT.2022.3212076.

Conferences:

1. **D. Shivakrishna** and V. Lalitha, ”Some Results on Maximally Recoverable Codes with Locality and Hierarchical Locality,” *2022 IEEE International Symposium on Information Theory (ISIT)*, Espoo, Finland, 2022, pp. 1181-1186, doi: 10.1109/ISIT50566.2022.9834863.

2. **D. Shivakrishna** and V. Lalitha, ”Properties of Maximally Recoverable Product Codes and Higher Order MDS Codes,” *2022 National Conference on Communications (NCC)*, Mumbai, India, 2022, pp. 233-238, doi: 10.1109/NCC55593.2022.9806469.

3. **D. Shivakrishna**, A. M. Nair and V. Lalitha, ”A Field Size Bound and Constructions of Maximally Recoverable Codes with Hierarchical Locality,” *2021 IEEE International Symposium on Information Theory (ISIT)*, Melbourne, Australia, 2021, pp. 2048-2053, doi: 10.1109/ISIT45174.2021.9517844.

4. **D. Shivakrishna**, V. A. Rameshwar, V. Lalitha and B. Sasidharan, ”On Maximally Recoverable Codes for Product Topologies,” *2018 Twenty Fourth National Conference on Communications (NCC), Hyderabad, India*, 2018, pp. 1-6, doi: 10.1109/NCC.2018.8599965.

Appendix A

Proofs for Chapter 3

A.1 Proof of Theorem 3.3.5

The following results related to the determinants of matrices will be useful in proving Theorem 3.3.5.

Lemma A.1.1 ([56]). *Let C_1 be an $a \times (a+1)$ matrix, C_2 be an $a \times (a+2)$ matrix, D_1 be a $3 \times (a+1)$ matrix and D_2 be a $3 \times (a+2)$ matrix and let $D_i^{(j)}$ be the j^{th} row of D_i. Then,*

$$\det \left[\begin{array}{c|c} C_1 & 0 \\ \hline 0 & C_2 \\ \hline D_1 & D_2 \end{array}\right] = (-1)^a \cdot \Bigg(\det \begin{pmatrix} C_1 \\ D_1^{(1)} \end{pmatrix} \cdot \det \begin{pmatrix} C_2 \\ D_2^{(2)} \\ D_2^{(3)} \end{pmatrix} - \det \begin{pmatrix} C_1 \\ D_1^{(2)} \end{pmatrix} \cdot \det \begin{pmatrix} C_2 \\ D_2^{(1)} \\ D_2^{(3)} \end{pmatrix} + \det \begin{pmatrix} C_1 \\ D_1^{(3)} \end{pmatrix} \cdot \det \begin{pmatrix} C_2 \\ D_2^{(1)} \\ D_2^{(2)} \end{pmatrix} \Bigg)$$

Lemma A.1.2 ([56]). *Given C_1 and C_2 to be $a \times (a+1)$ matrices and C_3 to be an $a \times (a+2)$ matrix. Also, D_1 and D_2 are $4 \times (a+1)$ matrices while D_3 is a $4 \times (a+2)$ matrix. It is also* *given that* $D_3^{(1)}, D_1^{(2)}, D_2^{(2)} = [0]$. *Then,*

$$\det \left[\begin{array}{c|c|c} C_1 & 0 & 0 \\ \hline 0 & C_2 & 0 \\ \hline 0 & 0 & C_3 \\ \hline D_1 & D_2 & D_3 \end{array}\right] = (-1)^a \cdot \left(\det \begin{pmatrix} C_1 \\ D_1^{(1)} \end{pmatrix} \cdot \det \begin{pmatrix} C_2 \\ D_2^{(3)} \end{pmatrix} \cdot \det \begin{pmatrix} C_3 \\ D_3^{(2)} \\ D_3^{(4)} \end{pmatrix} \right.$$

$$+ \det \begin{pmatrix} C_1 \\ D_1^{(1)} \end{pmatrix} \cdot \det \begin{pmatrix} C_2 \\ D_2^{(4)} \end{pmatrix} \cdot \det \begin{pmatrix} C_3 \\ D_3^{(2)} \\ D_3^{(3)} \end{pmatrix}$$

$$+ \det \begin{pmatrix} C_1 \\ D_1^{(3)} \end{pmatrix} \cdot \det \begin{pmatrix} C_2 \\ D_2^{(1)} \end{pmatrix} \cdot \det \begin{pmatrix} C_3 \\ D_3^{(2)} \\ D_3^{(4)} \end{pmatrix}$$

$$\left. - \det \begin{pmatrix} C_1 \\ D_1^{(4)} \end{pmatrix} \cdot \det \begin{pmatrix} C_2 \\ D_2^{(1)} \end{pmatrix} \cdot \det \begin{pmatrix} C_3 \\ D_3^{(2)} \\ D_3^{(3)} \end{pmatrix} \right)$$

Proof: Follows as a result of Lemmas B.2 in [56]. □

Lemma A.1.3 (Cauchy Matrix [75]). *Let* $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n \in \mathbb{F}_q$ *be all distinct. Then,*

$$\det \begin{bmatrix} \frac{1}{a_1-b_1} & \frac{1}{a_2-b_1} & \cdots & \frac{1}{a_n-b_1} \\ \frac{1}{a_1-b_2} & \frac{1}{a_2-b_2} & \cdots & \frac{1}{a_n-b_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{a_1-b_n} & \frac{1}{a_2-b_n} & \cdots & \frac{1}{a_n-b_n} \end{bmatrix} = \frac{\prod_{i>j}(a_i - a_j)(b_i - b_j)}{\prod_{i,j}(a_i - b_j)}.$$

Such a matrix is called an Cauchy Matrix. Every minor of a Cauchy matrix is also an Cauchy matrix.

Again as in previous proof, we consider the case when there are δ erasures per local code, one erasure per mid-level code and two more global erasures anywhere in the code. We again look at the erasure patterns within each mid-level codes. There are three distinct patterns possible,

1. No global erasures occur in that mid-level code.

2. Either one or both of the global erasures occur in the mid-level code.

We show that all the above erasure patterns are correctable.

Let $\gamma_{i,j} = \frac{1}{\alpha_j - \beta_i}$.

1. When no global erasures occur in the mid-level code, there are δ erasures per local code and one more erasure per mid-level code.

   In this scenario, we involve the mid-level parities. Let l be the affected mid-level code and l' be the local code within the mid-level code where the erasure occurs. The matrix, B_l

$$B_l = \begin{bmatrix} \gamma_{1,l'_1} & \gamma_{1,l'_2} & \cdots & \gamma_{1,l'_{\delta+1}} \\ \gamma_{2,l'_1} & \gamma_{2,l'_2} & \cdots & \gamma_{2,l'_{\delta+1}} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{\delta,l'_1} & \gamma_{\delta,l'_2} & \cdots & \gamma_{\delta,l'_{\delta+1}} \\ \gamma_{\delta+1,l'_1} & \gamma_{\delta+1,l'_2} & \cdots & \gamma_{\delta+1,l'_{\delta+1}} \end{bmatrix}.$$

   Where $\{l'_1, l'_2, \ldots, l'_{\delta+1}\}$ are the erased coordinates in local code l'. This is a Cauchy matrix and hence $\det(B_l) \neq 0$.

2. When there are global erasures, there are δ erasures per local code, one erasure per mid-level code, and two more erasures anywhere in the code. Here we have a lot more sub-cases.

   (a) Two global erasures are in the same local code as the mid-level erasure. Let l be the affected mid-level code and l' be the local code in the mid-level code where the erasure happens. The matrix B_l in that case,

$$B_l = \begin{bmatrix} \gamma_{1,l'_1} & \cdots & \gamma_{1,l'_{\delta+3}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta+1,l'_1} & \cdots & \gamma_{\delta+1,l'_{\delta+3}} \\ \lambda_{l'} \cdot \gamma_{\delta+2,l'_1} & \cdots & \lambda_{l'} \cdot \gamma_{\delta+2,l'_{\delta+3}} \\ \mu_{l'} \cdot \gamma_{\delta+3,l'_1} & \cdots & \mu_{l'} \cdot \gamma_{\delta+3,l'_{\delta+3}} \end{bmatrix}.$$

This is also a Cauchy matrix with the last two rows scaled to $\lambda_{l'}$ and $\mu_{l'}$ respectively. Hence $\det(B_l) \neq 0$ and this erasure pattern is correctible.

(b) Two global erasures are in one local code and the mid-level erasure is in a different local code for that mid-level code.

Assume that the l^{th} mid-level code is affected. Let l'' be the local code with two erasures while l' be the other one within this mid-level code.

$$B_l = \begin{bmatrix} \gamma_{1,l'_1} & \cdots & \gamma_{1,l'_{\delta+1}} & & & \\ \vdots & \ddots & \vdots & & & \\ \gamma_{\delta,l'_1} & \cdots & \gamma_{\delta,l'_{\delta+1}} & & & \\ & & & \gamma_{1,l''_1} & \cdots & \gamma_{1,l''_{\delta+2}} \\ & & & \vdots & \ddots & \vdots \\ & & & \gamma_{\delta,l''_1} & \cdots & \gamma_{\delta,l''_{\delta+2}} \\ \gamma_{\delta+1,l'_1} & \cdots & \gamma_{\delta+1,l'_{\delta+1}} & \gamma_{\delta+1,l''_1} & \cdots & \gamma_{\delta+1,l''_{\delta+2}} \\ \lambda_{l'} \cdot \gamma_{\delta+2,l'_1} & \cdots & \lambda_{l'} \cdot \gamma_{\delta+2,l'_{\delta+1}} & \lambda_{l''} \cdot \gamma_{\delta+2,l''_1} & \cdots & \lambda_{l''} \cdot \gamma_{\delta+2,l''_{\delta+2}} \\ \mu_{l'} \cdot \gamma_{\delta+3,l'_1} & \cdots & \mu_{l'} \cdot \gamma_{\delta+3,l'_{\delta+1}} & \mu_{l''} \cdot \gamma_{\delta+3,l''_1} & \cdots & \mu_{l''} \cdot \gamma_{\delta+3,l''_{\delta+2}} \end{bmatrix}.$$

Expanding this via the Lemma A.1.1,

$$\det(B_l) = \det\begin{pmatrix} \gamma_{1,l'_1} & \cdots & \gamma_{1,l'_{\delta+1}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l'_1} & \cdots & \gamma_{\delta,l'_{\delta+1}} \\ \gamma_{\delta+1,l'_1} & \cdots & \gamma_{\delta+1,l'_{\delta+1}} \end{pmatrix} \cdot \det\begin{pmatrix} \gamma_{1,l''_1} & \cdots & \gamma_{1,l''_{\delta+2}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l''_1} & \cdots & \gamma_{\delta,l''_{\delta+2}} \\ \lambda_{l''}\cdot\gamma_{\delta+2,l''_1} & \cdots & \lambda_{l''}\cdot\gamma_{\delta+2,l''_{\delta+2}} \\ \mu_{l''}\cdot\gamma_{\delta+3,l''_1} & \cdots & \mu_{l''}\cdot\gamma_{\delta+3,l''_{\delta+2}} \end{pmatrix}$$

$$- \det\begin{pmatrix} \gamma_{1,l'_1} & \cdots & \gamma_{1,l'_{\delta+1}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l'_1} & \cdots & \gamma_{\delta,l'_{\delta+1}} \\ \lambda_{l'}\cdot\gamma_{\delta+2,l'_1} & \cdots & \lambda_{l'}\cdot\gamma_{\delta+2,l'_{\delta+1}} \end{pmatrix} \cdot \det\begin{pmatrix} \gamma_{1,l''_1} & \cdots & \gamma_{1,l''_{\delta+2}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l''_1} & \cdots & \gamma_{\delta,l''_{\delta+2}} \\ \gamma_{\delta+1,l''_1} & \cdots & \gamma_{\delta+1,l''_{\delta+2}} \\ \mu_{l''}\cdot\gamma_{\delta+3,l''_1} & \cdots & \mu_{l''}\cdot\gamma_{\delta+3,l''_{\delta+2}} \end{pmatrix}$$

$$+ \det\begin{pmatrix} \gamma_{1,l'_1} & \cdots & \gamma_{1,l'_{\delta+1}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l'_1} & \cdots & \gamma_{\delta,l'_{\delta+1}} \\ \mu_{l'}\cdot\gamma_{\delta+3,l'_1} & \cdots & \mu_{l'}\cdot\gamma_{\delta+3,l'_{\delta+1}} \end{pmatrix} \cdot \det\begin{pmatrix} \gamma_{1,l''_1} & \cdots & \gamma_{1,l''_{\delta+2}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l''_1} & \cdots & \gamma_{\delta,l''_{\delta+2}} \\ \gamma_{\delta+1,l''_1} & \cdots & \gamma_{\delta+1,l''_{\delta+2}} \\ \lambda_{l''}\cdot\gamma_{\delta+2,l''_1} & \cdots & \lambda_{l''}\cdot\gamma_{\delta+2,l''_{\delta+2}} \end{pmatrix}$$

$$\det(B_l) = \lambda_{l''}\mu_{l''} \cdot \det\begin{pmatrix} \gamma_{1,l'_1} & \cdots & \gamma_{1,l'_{\delta+1}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l'_1} & \cdots & \gamma_{\delta,l'_{\delta+1}} \\ \gamma_{\delta+1,l'_1} & \cdots & \gamma_{\delta+1,l'_{\delta+1}} \end{pmatrix} \cdot \det\begin{pmatrix} \gamma_{1,l''_1} & \cdots & \gamma_{1,l''_{\delta+2}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l''_1} & \cdots & \gamma_{\delta,l''_{\delta+2}} \\ \gamma_{\delta+2,l''_1} & \cdots & \gamma_{\delta+2,l''_{\delta+2}} \\ \gamma_{\delta+3,l''_1} & \cdots & \gamma_{\delta+3,l''_{\delta+2}} \end{pmatrix}$$

$$- \lambda_{l'}\mu_{l''} \cdot \det\begin{pmatrix} \gamma_{1,l'_1} & \cdots & \gamma_{1,l'_{\delta+1}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l'_1} & \cdots & \gamma_{\delta,l'_{\delta+1}} \\ \gamma_{\delta+2,l'_1} & \cdots & \gamma_{\delta+2,l'_{\delta+1}} \end{pmatrix} \cdot \det\begin{pmatrix} \gamma_{1,l''_1} & \cdots & \gamma_{1,l''_{\delta+2}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l''_1} & \cdots & \gamma_{\delta,l''_{\delta+2}} \\ \gamma_{\delta+1,l''_1} & \cdots & \gamma_{\delta+1,l''_{\delta+2}} \\ \gamma_{\delta+3,l''_1} & \cdots & \gamma_{\delta+3,l''_{\delta+2}} \end{pmatrix}$$

$$+ \lambda_{l''}\mu_{l'} \cdot \det\begin{pmatrix} \gamma_{1,l'_1} & \cdots & \gamma_{1,l'_{\delta+1}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l'_1} & \cdots & \gamma_{\delta,l'_{\delta+1}} \\ \gamma_{\delta+3,l'_1} & \cdots & \gamma_{\delta+3,l'_{\delta+1}} \end{pmatrix} \cdot \det\begin{pmatrix} \gamma_{1,l''_1} & \cdots & \gamma_{1,l''_{\delta+2}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l''_1} & \cdots & \gamma_{\delta,l''_{\delta+2}} \\ \gamma_{\delta+1,l''_1} & \cdots & \gamma_{\delta+1,l''_{\delta+2}} \\ \gamma_{\delta+2,l''_1} & \cdots & \gamma_{\delta+2,l''_{\delta+2}} \end{pmatrix}$$

Each term in this determinant is $\lambda_i\mu_j$ multiplied by a Cauchy matrix $\in \mathbb{F}_{q_0}$. The determinant is again a linear combination of $\lambda_{l'}$ and $\lambda_{l''}$. Again, this determinant cannot be zero because λ's are 4-wise independent.

(c) Two global and the one mid-level erasure are all in three different local codes within the same mid-level code. Let the affected mid-level code be l and the local codes within, where the erasure occurs, be $l^{(1)}$, $l^{(2)}$ and $l^{(3)}$. The matrix B_l,

$$B_l = \begin{bmatrix} \gamma_{1,l_1^{(1)}} & \cdots & \gamma_{1,l_{\delta+1}^{(1)}} & & & & & & \\ \vdots & \ddots & \vdots & & & & & & \\ \gamma_{\delta,l_1^{(1)}} & \cdots & \gamma_{\delta,l_{\delta+1}^{(1)}} & & & & & & \\ & & & \gamma_{1,l_1^{(2)}} & \cdots & \gamma_{1,l_{\delta+1}^{(2)}} & & & \\ & & & \vdots & \ddots & \vdots & & & \\ & & & \gamma_{\delta,l_1^{(2)}} & \cdots & \gamma_{\delta,l_{\delta+1}^{(2)}} & & & \\ & & & & & & \gamma_{1,l_1^{(3)}} & \cdots & \gamma_{1,l_{\delta+1}^{(3)}} \\ & & & & & & \vdots & \ddots & \vdots \\ & & & & & & \gamma_{\delta,l_1^{(3)}} & \cdots & \gamma_{\delta,l_{\delta+1}^{(3)}} \\ \gamma_{\delta+1,l_1^{(1)}} & \cdots & \gamma_{\delta+1,l_{\delta+1}^{(1)}} & \gamma_{\delta+1,l_1^{(2)}} & \cdots & \gamma_{\delta+1,l_{\delta+1}^{(2)}} & \gamma_{\delta+1,l_1^{(3)}} & \cdots & \gamma_{\delta+1,l_{\delta+1}^{(3)}} \\ \lambda_{l^{(1)}} \cdot \gamma_{\delta+2,l_1^{(1)}} & \cdots & \lambda_{l^{(1)}} \cdot \gamma_{\delta+2,l_{\delta+1}^{(1)}} & \lambda_{l^{(2)}} \cdot \gamma_{\delta+2,l_1^{(2)}} & \cdots & \lambda_{l^{(2)}} \cdot \gamma_{\delta+2,l_{\delta+1}^{(2)}} & \lambda_{l^{(3)}} \cdot \gamma_{\delta+2,l_1^{(3)}} & \cdots & \lambda_{l^{(3)}} \cdot \gamma_{\delta+2,l_{\delta+1}^{(3)}} \\ \mu_{l^{(1)}} \cdot \gamma_{\delta+3,l_1^{(1)}} & \cdots & \mu_{l^{(1)}} \cdot \gamma_{\delta+3,l_{\delta+1}^{(1)}} & \mu_{l^{(2)}} \cdot \gamma_{\delta+3,l_1^{(2)}} & \cdots & \mu_{l^{(2)}} \cdot \gamma_{\delta+3,l_{\delta+1}^{(2)}} & \mu_{l^{(3)}} \cdot \gamma_{\delta+3,l_1^{(3)}} & \cdots & \mu_{l^{(3)}} \cdot \gamma_{\delta+3,l_{\delta+1}^{(3)}} \end{bmatrix}.$$

$\det(B_l)$ can be expanded via Lemma 3.3.2. After doing that and setting the determinant to zero,

$$\det(B_l) = 0,$$

we get,

$$\det\begin{bmatrix} \frac{c_{l^{(1)}}d\prod_{i\in[\delta]}(\beta_i-\beta_{\delta+1})}{e_{l^{(1)}}\prod_{i\in l_S^{(1)}}(\alpha_i-\beta_{\delta+1})} & \frac{c_{l^{(2)}}d\prod_{i\in[\delta]}(\beta_i-\beta_{\delta+1})}{e_{l^{(2)}}\prod_{i\in l_S^{(2)}}(\alpha_i-\beta_{\delta+1})} & \frac{c_{l^{(3)}}d\prod_{i\in[\delta]}(\beta_i-\beta_{\delta+1})}{e_{l^{(3)}}\prod_{i\in l_S^{(3)}}(\alpha_i-\beta_{\delta+1})} \\ \lambda_{l^{(1)}}\cdot\frac{c_{l^{(1)}}d\prod_{i\in[\delta]}(\beta_i-\beta_{\delta+2})}{e_{l^{(1)}}\prod_{i\in l_S^{(1)}}(\alpha_i-\beta_{\delta+2})} & \lambda_{l^{(2)}}\cdot\frac{c_{l^{(2)}}d\prod_{i\in[\delta]}(\beta_i-\beta_{\delta+2})}{e_{l^{(2)}}\prod_{i\in l_S^{(2)}}(\alpha_i-\beta_{\delta+2})} & \lambda_{l^{(3)}}\cdot\frac{c_{l^{(3)}}d\prod_{i\in[\delta]}(\beta_i-\beta_{\delta+2})}{e_{l^{(3)}}\prod_{i\in l_S^{(3)}}(\alpha_i-\beta_{\delta+2})} \\ \mu_{l^{(1)}}\cdot\frac{c_{l^{(1)}}d\prod_{i\in[\delta]}(\beta_i-\beta_{\delta+3})}{e_{l^{(1)}}\prod_{i\in l_S^{(1)}}(\alpha_i-\beta_{\delta+3})} & \mu_{l^{(2)}}\cdot\frac{c_{l^{(2)}}d\prod_{i\in[\delta]}(\beta_i-\beta_{\delta+3})}{e_{l^{(2)}}\prod_{i\in l_S^{(2)}}(\alpha_i-\beta_{\delta+3})} & \mu_{l^{(3)}}\cdot\frac{c_{l^{(3)}}d\prod_{i\in[\delta]}(\beta_i-\beta_{\delta+3})}{e_{l^{(3)}}\prod_{i\in l_S^{(3)}}(\alpha_i-\beta_{\delta+3})} \end{bmatrix} = 0$$

$$\det\begin{bmatrix} 1 & 1 & 1 \\ \lambda_{l^{(1)}}\prod_{i\in l_S^{(1)}}\frac{\alpha_i-\beta_{\delta+1}}{\alpha_i-\beta_{\delta+2}} & \lambda_{l^{(2)}}\prod_{i\in l_S^{(2)}}\frac{\alpha_i-\beta_{\delta+1}}{\alpha_i-\beta_{\delta+2}} & \lambda_{l^{(3)}}\prod_{i\in l_S^{(3)}}\frac{\alpha_i-\beta_{\delta+1}}{\alpha_i-\beta_{\delta+3}} \\ \mu_{l^{(1)}}\prod_{i\in l_S^{(1)}}\frac{\alpha_i-\beta_{\delta+1}}{\alpha_i-\beta_{\delta+3}} & \mu_{l^{(2)}}\prod_{i\in l_S^{(2)}}\frac{\alpha_i-\beta_{\delta+1}}{\alpha_i-\beta_{\delta+2}} & \mu_{l^{(3)}}\prod_{i\in l_S^{(3)}}\frac{\alpha_i-\beta_{\delta+1}}{\alpha_i-\beta_{\delta+3}} \end{bmatrix} = 0,$$

where,

- $l_S^{(i)} = \{l_1^{(i)}, \ldots, l_{\delta+1}^{(i)}\}$.
- $c_{l^{(i)}} = \prod_{f>g, f,g\in l_S^{(i)}}(\alpha_f - \alpha_g)$.
- $d = \prod_{f>g, f,g\in[\delta]}(\beta_f - \beta_g)$.
- $e_{l^{(i)}} = \prod_{f\in l_S^{(i)}, g\in[\delta]}(\alpha_f - \beta_g)$.

Now, by the choice of α's, $\prod_{i\in l_S^{(k)}}\frac{\alpha_i-\beta_{\delta+1}}{\alpha_i-\beta_{\delta+3}} \in G$. And because μ_i belong to different cosets in G, the last row in the above matrix consists of distinct elements. This determinant is a linear combination in the three λ's. Hence the determinant is non-zero because the λ's are 4-wise independent.

(d) Two global erasures are in different mid-level codes. In both the mid-level codes, the mid-level erasure is in the same local code as the global erasure.

Assume k^{th} and l^{th} mid-level codes are affected. The local codes within them, where the erasure occurs, are k' and l'. The matrix $B_{k,l}$,

$$B_{k,l} = \begin{bmatrix} & \gamma_{1,k'_1} & \cdots & \gamma_{1,k'_{\delta+2}} & & & & \\ & \vdots & \ddots & \vdots & & & & \\ & \gamma_{\delta+1,k'_1} & \cdots & \gamma_{\delta+1,k'_{\delta+2}} & & & & \\ & & & & \gamma_{1,l'_1} & \cdots & \gamma_{1,l'_{\delta+2}} & \\ & & & & \vdots & \ddots & \vdots & \\ & & & & \gamma_{\delta+1,l'_1} & \cdots & \gamma_{\delta+1,l'_{\delta+2}} & \\ \lambda_{k'} \cdot \gamma_{\delta+2,k'_1} & & \cdots & \lambda_{k'} \cdot \gamma_{\delta+2,k'_{\delta+2}} & \lambda_{l'} \cdot \gamma_{\delta+2,l'_1} & \cdots & \lambda_{l'} \cdot \gamma_{\delta+2,l'_{\delta+2}} & \\ \mu_{k'} \cdot \gamma_{\delta+3,k'_1} & & \cdots & \mu_{k'} \cdot \gamma_{\delta+3,k'_{\delta+2}} & \mu_{l'} \cdot \gamma_{\delta+3,l'_1} & \cdots & \mu_{l'} \cdot \gamma_{\delta+3,l'_{\delta+2}} & \end{bmatrix}.$$

Therefore, for $\det(B_{k,l}) = 0$,

$$\det(B_{k,l}) = \det \begin{bmatrix} & \gamma_{1,k'_1} & \cdots & \gamma_{1,k'_{\delta+2}} & & & & \\ & \vdots & \ddots & \vdots & & & & \\ & \gamma_{\delta+1,k'_1} & \cdots & \gamma_{\delta+1,k'_{\delta+2}} & & & & \\ & & & & \gamma_{1,l'_1} & \cdots & \gamma_{1,l'_{\delta+2}} & \\ & & & & \vdots & \ddots & \vdots & \\ & & & & \gamma_{\delta+1,l'_1} & \cdots & \gamma_{\delta+1,l'_{\delta+2}} & \\ \lambda_{k'} \cdot \gamma_{\delta+2,k'_1} & & \cdots & \lambda_{k'} \cdot \gamma_{\delta+2,k'_{\delta+2}} & \lambda_{l'} \cdot \gamma_{\delta+2,l'_1} & \cdots & \lambda_{l'} \cdot \gamma_{\delta+2,l'_{\delta+2}} & \\ \mu_{k'} \cdot \gamma_{\delta+3,k'_1} & & \cdots & \mu_{k'} \cdot \gamma_{\delta+3,k'_{\delta+2}} & \mu_{l'} \cdot \gamma_{\delta+3,l'_1} & \cdots & \mu_{l'} \cdot \gamma_{\delta+3,l'_{\delta+2}} & \end{bmatrix} = 0$$

$$\Rightarrow \det \begin{bmatrix} \det \begin{pmatrix} \gamma_{1,k'_1} & \cdots & \gamma_{1,k'_{\delta+2}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta+1,k'_1} & \cdots & \gamma_{\delta+1,k'_{\delta+2}} \\ \lambda_{k'} \cdot \gamma_{\delta+2,k'_1} & \cdots & \lambda_{k'} \cdot \gamma_{\delta+2,k'_{\delta+2}} \end{pmatrix} & \det \begin{pmatrix} \gamma_{1,l'_1} & \cdots & \gamma_{1,l'_{\delta+2}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta+1,l'_1} & \cdots & \gamma_{\delta+1,l'_{\delta+2}} \\ \lambda_{l'} \cdot \gamma_{\delta+2,l'_1} & \cdots & \lambda_{l'} \cdot \gamma_{\delta+2,l'_{\delta+2}} \end{pmatrix} \\ \det \begin{pmatrix} \gamma_{1,k'_1} & \cdots & \gamma_{1,k'_{\delta+2}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta+1,k'_1} & \cdots & \gamma_{\delta+1,k'_{\delta+2}} \\ \mu_{k'} \cdot \gamma_{\delta+3,k'_1} & \cdots & \mu_{k'} \cdot \gamma_{\delta+3,k'_{\delta+2}} \end{pmatrix} & \det \begin{pmatrix} \gamma_{1,l'_1} & \cdots & \gamma_{1,l'_{\delta+2}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta+1,l'_1} & \cdots & \gamma_{\delta+1,l'_{\delta+2}} \\ \mu_{l'} \cdot \gamma_{\delta+3,l'_1} & \cdots & \mu_{l'} \cdot \gamma_{\delta+3,l'_{\delta+2}} \end{pmatrix} \end{bmatrix} = 0$$

$$\Rightarrow \det \begin{bmatrix} \lambda_{k'} \cdot \frac{\prod_{i\in[\delta+1]}(\beta_i-\beta_{\delta+2})}{\prod_{i\in k'_S}(\alpha_i-\beta_{\delta+2})} & \lambda_{l'} \cdot \frac{\prod_{i\in[\delta+1]}(\beta_i-\beta_{\delta+2})}{\prod_{i\in l'_S}(\alpha_i-\beta_{\delta+2})} \\ \mu_{k'} \cdot \frac{\prod_{i\in[\delta+1]}(\beta_i-\beta_{\delta+3})}{\prod_{i\in k'_S}(\alpha_i-\beta_{\delta+3})} & \mu_{l'} \cdot \frac{\prod_{i\in[\delta+1]}(\beta_i-\beta_{\delta+3})}{\prod_{i\in l'_S}(\alpha_i-\beta_{\delta+3})} \end{bmatrix} = 0$$

$$\Rightarrow \det \begin{bmatrix} \lambda_{k'} & \lambda_{l'} \\ \mu_{k'} \prod_{i\in k'_S} \frac{(\alpha_i-\beta_{\delta+2})}{(\alpha_i-\beta_{\delta+3})} & \mu_{l'} \prod_{i\in l'_S} \frac{(\alpha_i-\beta_{\delta+2})}{(\alpha_i-\beta_{\delta+3})} \end{bmatrix} = 0,$$

where $k'_S = \{k'_1, \ldots, k'_{\delta+2}\}$ and $l'_S = \{l'_1, \ldots, l'_{\delta+2}\}$. The terms $c_{l^{(i)}}$, d and $e_{l^{(i)}}$ were factored out from the above determinant where,

- $c_{l^{(i)}} = \prod_{f>g, f,g\in l^{(i)}_S}(\alpha_f - \alpha_g)$.
- $d = \prod_{f>g, f,g\in[\delta+1]}(\beta_f - \beta_g)$.
- $e_{l^{(i)}} = \prod_{f\in l^{(i)}_S, g\in[\delta+1]}(\alpha_f - \beta_g)$.

By the choice of α_i's, $\prod_{i\in x} \frac{(\alpha_i-\beta_{\delta+2})}{(\alpha_i-\beta_{\delta+3})} \in G$ for $x = k'_S, l'_S$. This yet again is a linear combination of two λ's. Hence this determinant is non-zero and the erasure pattern correctable.

(e) Two global erasures are in two different mid-level codes and in each mid-level code, they are in a different local code as compared to the mid-level erasure for that mid-level code.

There are four local groups where the erasures occur, two in each mid-level code. Let the affected mid-level codes be k and l while the local codes within, where the erasure occurs, be $k^{(1)}$ and $k^{(2)}$ and $l^{(1)}$ and $l^{(2)}$ respectively. The matrix $B_{k,l}$ is similar to that in Lemma A.1.1.

$$B_{k,l} = \begin{bmatrix} A & \\ & B \\ C & D \end{bmatrix} \Rightarrow \det(B_{k,l}) = \det \begin{bmatrix} \det \begin{pmatrix} A \\ C^{(1)} \end{pmatrix} & \det \begin{pmatrix} B \\ D^{(1)} \end{pmatrix} \\ \det \begin{pmatrix} A \\ C^{(2)} \end{pmatrix} & \det \begin{pmatrix} B \\ D^{(2)} \end{pmatrix} \end{bmatrix} = 0$$

$$A = \begin{bmatrix} \gamma_{1,k_1^{(1)}} & \cdots & \gamma_{1,k_{\delta+1}^{(1)}} & & & \\ \vdots & \ddots & \vdots & & & \\ \gamma_{\delta,k_1^{(1)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(1)}} & & & \\ & & & \gamma_{1,k_1^{(2)}} & \cdots & \gamma_{1,k_{\delta+1}^{(2)}} \\ & & & \vdots & \ddots & \vdots \\ & & & \gamma_{\delta,k_1^{(2)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(2)}} \\ \gamma_{\delta+1,k_1^{(1)}} & \cdots & \gamma_{\delta+1,k_{\delta+1}^{(1)}} & \gamma_{\delta+1,k_1^{(2)}} & \cdots & \gamma_{\delta+1,k_{\delta+1}^{(2)}} \end{bmatrix}$$

$$B = \begin{bmatrix} \gamma_{1,l_1^{(1)}} & \cdots & \gamma_{1,l_{\delta+1}^{(1)}} & & & \\ \vdots & \ddots & \vdots & & & \\ \gamma_{\delta,l_1^{(1)}} & \cdots & \gamma_{\delta,l_{\delta+1}^{(1)}} & & & \\ & & & \gamma_{1,l_1^{(2)}} & \cdots & \gamma_{1,l_{\delta+1}^{(2)}} \\ & & & \vdots & \ddots & \vdots \\ & & & \gamma_{\delta,l_1^{(2)}} & \cdots & \gamma_{\delta,l_{\delta+1}^{(2)}} \\ \gamma_{\delta+1,l_1^{(1)}} & \cdots & \gamma_{\delta+1,l_{\delta+1}^{(1)}} & \gamma_{\delta+1,l_1^{(2)}} & \cdots & \gamma_{\delta+1,l_{\delta+1}^{(2)}} \end{bmatrix}$$

$$C = \begin{bmatrix} \lambda_{k^{(1)}} \cdot \gamma_{\delta+2,k_1^{(1)}} & \cdots & \lambda_{k^{(1)}} \cdot \gamma_{\delta+2,k_{\delta+1}^{(1)}} & \lambda_{k^{(2)}} \cdot \gamma_{\delta+2,k_1^{(2)}} & \cdots & \lambda_{k^{(2)}} \cdot \gamma_{\delta+2,k_{\delta+1}^{(2)}} \\ \mu_{k^{(1)}} \cdot \gamma_{\delta+3,k_1^{(1)}} & \cdots & \mu_{k^{(1)}} \cdot \gamma_{\delta+3,k_{\delta+1}^{(1)}} & \mu_{k^{(2)}} \cdot \gamma_{\delta+3,k_1^{(2)}} & \cdots & \mu_{k^{(2)}} \cdot \gamma_{\delta+3,k_{\delta+1}^{(2)}} \end{bmatrix}$$

$$D = \begin{bmatrix} \lambda_{l^{(1)}} \cdot \gamma_{\delta+2,l_1^{(1)}} & \cdots & \lambda_{l^{(1)}} \cdot \gamma_{\delta+2,l_{\delta+1}^{(1)}} & \lambda_{l^{(2)}} \cdot \gamma_{\delta+2,l_1^{(2)}} & \cdots & \lambda_{l^{(2)}} \cdot \gamma_{\delta+2,l_{\delta+1}^{(2)}} \\ \mu_{l^{(1)}} \cdot \gamma_{\delta+3,l_1^{(1)}} & \cdots & \mu_{l^{(1)}} \cdot \gamma_{\delta+3,l_{\delta+1}^{(1)}} & \mu_{l^{(2)}} \cdot \gamma_{\delta+3,l_1^{(2)}} & \cdots & \mu_{l^{(2)}} \cdot \gamma_{\delta+3,l_{\delta+1}^{(2)}} \end{bmatrix}$$

To calculate the whole determinant, We consider the first element,

$$\det\begin{pmatrix} A \\ C^{(1)} \end{pmatrix} = \det\begin{bmatrix} & \gamma_{1,k_1^{(1)}} & \cdots & \gamma_{1,k_{\delta+1}^{(1)}} & & & & \\ & \vdots & \ddots & \vdots & & & & \\ & \gamma_{\delta,k_1^{(1)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(1)}} & & & & \\ & & & & \gamma_{1,k_1^{(2)}} & \cdots & \gamma_{1,k_{\delta+1}^{(2)}} & \\ & & & & \vdots & \ddots & \vdots & \\ & & & & \gamma_{\delta,k_1^{(2)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(2)}} & \\ & \gamma_{\delta+1,k_1^{(1)}} & \cdots & \gamma_{\delta+1,k_{\delta+1}^{(1)}} & \gamma_{\delta+1,k_1^{(2)}} & \cdots & \gamma_{\delta+1,k_{\delta+1}^{(2)}} & \\ \lambda_{k^{(1)}}\cdot\gamma_{\delta+2,k_1^{(1)}} & & \cdots & \lambda_{k^{(1)}}\cdot\gamma_{\delta+2,k_{\delta+1}^{(1)}} & \lambda_{k^{(2)}}\cdot\gamma_{\delta+2,k_1^{(2)}} & \cdots & \lambda_{k^{(2)}}\cdot\gamma_{\delta+2,k_{\delta+1}^{(2)}} & \end{bmatrix}$$

$$= \det\begin{bmatrix} \det\begin{pmatrix} \gamma_{1,k_1^{(1)}} & \cdots & \gamma_{1,k_{\delta+1}^{(1)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,k_1^{(1)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(1)}} \\ \gamma_{\delta+1,k_1^{(1)}} & \cdots & \gamma_{\delta+1,k_{\delta+1}^{(1)}} \end{pmatrix} & \det\begin{pmatrix} \gamma_{1,k_1^{(2)}} & \cdots & \gamma_{1,k_{\delta+1}^{(2)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,k_1^{(2)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(2)}} \\ \gamma_{\delta+1,k_1^{(2)}} & \cdots & \gamma_{\delta+1,k_{\delta+1}^{(2)}} \end{pmatrix} \\ \det\begin{pmatrix} \gamma_{1,k_1^{(1)}} & \cdots & \gamma_{1,k_{\delta+1}^{(1)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,k_1^{(1)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(1)}} \\ \lambda_{k^{(1)}}\cdot\gamma_{\delta+2,k_1^{(1)}} & \cdots & \lambda_{k^{(1)}}\cdot\gamma_{\delta+2,k_{\delta+1}^{(1)}} \end{pmatrix} & \det\begin{pmatrix} \gamma_{1,k_1^{(2)}} & \cdots & \gamma_{1,k_{\delta+1}^{(2)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,k_1^{(2)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(2)}} \\ \lambda_{k^{(2)}}\cdot\gamma_{\delta+2,k_1^{(2)}} & \cdots & \lambda_{k^{(2)}}\cdot\gamma_{\delta+2,k_{\delta+1}^{(2)}} \end{pmatrix} \end{bmatrix}$$

$$= \det\begin{bmatrix} \frac{c_{k^{(1)}} d \prod_{i\in[\delta]}(\beta_i-\beta_{\delta+1})}{e_{k^{(1)}}\prod_{i\in k_S^{(1)}}(\alpha_i-\beta_{\delta+1})} & \frac{c_{k^{(2)}} d \prod_{i\in[\delta]}(\beta_i-\beta_{\delta+1})}{e_{k^{(2)}}\prod_{i\in k_S^{(2)}}(\alpha_i-\beta_{\delta+1})} \\ \lambda_{k^{(1)}}\cdot\frac{c_{k^{(1)}} d \prod_{i\in[\delta]}(\beta_i-\beta_{\delta+2})}{e_{k^{(1)}}\prod_{i\in k_S^{(1)}}(\alpha_i-\beta_{\delta+2})} & \lambda_{k^{(2)}}\cdot\frac{c_{k^{(2)}} d \prod_{i\in[\delta]}(\beta_i-\beta_{\delta+2})}{e_{k^{(2)}}\prod_{i\in k_S^{(2)}}(\alpha_i-\beta_{\delta+2})} \end{bmatrix}$$

$$= \frac{c_{k^{(1)}}c_{k^{(2)}}d^2}{e_{k^{(1)}}e_{k^{(2)}}}\prod_{i\in[\delta]}(\beta_i-\beta_{\delta+1})(\beta_i-\beta_{\delta+2})\cdot\det\begin{bmatrix} \prod_{i\in k_S^{(1)}}\frac{1}{\alpha_i-\beta_{\delta+1}} & \prod_{i\in k_S^{(2)}}\frac{1}{\alpha_i-\beta_{\delta+1}} \\ \lambda_{k^{(1)}}\cdot\prod_{i\in k_S^{(1)}}\frac{1}{\alpha_i-\beta_{\delta+2}} & \lambda_{k^{(2)}}\cdot\prod_{i\in k_S^{(2)}}\frac{1}{\alpha_i-\beta_{\delta+2}} \end{bmatrix}.$$

Where,

- $k_S^{(i)} = \{k_1^{(i)}, \ldots, k_{\delta+2}^{(i)}\}$.
- $c_{k^{(i)}} = \prod_{f>g, f,g \in k_S^{(i)}} (\alpha_f - \alpha_g)$.
- $d = \prod_{f>g, f,g \in [\delta]} (\beta_f - \beta_g)$.
- $e_{k^{(i)}} = \prod_{f \in k_S^{(i)}, g \in [\delta]} (\alpha_f - \beta_g)$.

Applying all this in the main determinant and setting,

$$\det(B_{k,l}) = 0$$

and factoring out the common multiples, we get

$$\det \begin{bmatrix} \det \begin{pmatrix} \prod_{i \in k_S^{(1)}} \frac{1}{\alpha_i - \beta_{\delta+1}} & \prod_{i \in k_S^{(2)}} \frac{1}{\alpha_i - \beta_{\delta+1}} \\ \lambda_{k^{(1)}} \cdot \prod_{i \in k_S^{(1)}} \frac{1}{\alpha_i - \beta_{\delta+2}} & \lambda_{k^{(2)}} \cdot \prod_{i \in k_S^{(2)}} \frac{1}{\alpha_i - \beta_{\delta+2}} \end{pmatrix} & \det \begin{pmatrix} \prod_{i \in l_S^{(1)}} \frac{1}{\alpha_i - \beta_{\delta+1}} & \prod_{i \in l_S^{(2)}} \frac{1}{\alpha_i - \beta_{\delta+1}} \\ \lambda_{l^{(1)}} \cdot \prod_{i \in l_S^{(1)}} \frac{1}{\alpha_i - \beta_{\delta+2}} & \lambda_{l^{(2)}} \cdot \prod_{i \in l_S^{(2)}} \frac{1}{\alpha_i - \beta_{\delta+2}} \end{pmatrix} \\ \det \begin{pmatrix} \prod_{i \in k_S^{(1)}} \frac{1}{\alpha_i - \beta_{\delta+1}} & \prod_{i \in k_S^{(2)}} \frac{1}{\alpha_i - \beta_{\delta+1}} \\ \mu_{k^{(1)}} \cdot \prod_{i \in k_S^{(1)}} \frac{1}{\alpha_i - \beta_{\delta+3}} & \mu_{k^{(2)}} \cdot \prod_{i \in k_S^{(2)}} \frac{1}{\alpha_i - \beta_{\delta+3}} \end{pmatrix} & \det \begin{pmatrix} \prod_{i \in l_S^{(1)}} \frac{1}{\alpha_i - \beta_{\delta+1}} & \prod_{i \in l_S^{(2)}} \frac{1}{\alpha_i - \beta_{\delta+1}} \\ \mu_{l^{(1)}} \cdot \prod_{i \in l_S^{(1)}} \frac{1}{\alpha_i - \beta_{\delta+3}} & \mu_{l^{(2)}} \cdot \prod_{i \in l_S^{(2)}} \frac{1}{\alpha_i - \beta_{\delta+3}} \end{pmatrix} \end{bmatrix} = 0$$

$$\det \begin{bmatrix} \lambda_{k^{(2)}} \cdot \prod_{i \in k_S^{(2)}} \frac{\alpha_i - \beta_{\delta+1}}{\alpha_i - \beta_{\delta+2}} - \lambda_{k^{(1)}} \cdot \prod_{i \in k_S^{(1)}} \frac{\alpha_i - \beta_{\delta+1}}{\alpha_i - \beta_{\delta+2}} & \lambda_{l^{(2)}} \cdot \prod_{i \in l_S^{(2)}} \frac{\alpha_i - \beta_{\delta+1}}{\alpha_i - \beta_{\delta+2}} - \lambda_{l^{(1)}} \cdot \prod_{i \in l_S^{(1)}} \frac{\alpha_i - \beta_{\delta+1}}{\alpha_i - \beta_{\delta+2}} \\ \mu_{k^{(2)}} \cdot \prod_{i \in k_S^{(2)}} \frac{\alpha_i - \beta_{\delta+1}}{\alpha_i - \beta_{\delta+3}} - \mu_{k^{(1)}} \cdot \prod_{i \in k_S^{(1)}} \frac{\alpha_i - \beta_{\delta+1}}{\alpha_i - \beta_{\delta+3}} & \mu_{l^{(2)}} \cdot \prod_{i \in l_S^{(2)}} \frac{\alpha_i - \beta_{\delta+1}}{\alpha_i - \beta_{\delta+3}} - \mu_{l^{(1)}} \cdot \prod_{i \in l_S^{(1)}} \frac{\alpha_i - \beta_{\delta+1}}{\alpha_i - \beta_{\delta+3}} \end{bmatrix} = 0$$

Where similarly, $l_S^{(i)} = \{l_1^{(i)}, \ldots, l_{\delta+2}^{(i)}\}$.

Now, since the λ_i's are 4-wise independent over $\mathbb{F}_{q_0}$, the first row is never zero. Similarly, all the μ_j's are in different cosets of G and by choice of α's $\prod_{i \in l_S^{(j)}, k_S^{(j)}} \frac{\alpha_i - \beta_{\delta+1}}{\alpha_i - \beta_{\delta+3}} \in G$. Hence the last row isn't zero either. Then this determinant resolves into a linear combination for 4 different values of λ_is. Hence, by linear independence rules of λ, this determinant is also non-zero.

(f) Two global erasures are in two different mid-level codes. In one of the mid-level code, the mid-level erasure is in the same local code as the global erasure, while in the other they are in different local codes. Assume that the k^{th} and l^{th} mid-level codes are affected. Let the local codes within, where the erasure occurs, be $k^{(1)}, k^{(2)}$ and $l^{(1)}$. The matrix $B_{k,l}$,

$$
B_{k,l} = \begin{bmatrix}
\gamma_{1,k_1^{(1)}} & \cdots & \gamma_{1,k_{\delta+1}^{(1)}} & & & & & & \\
\vdots & \ddots & \vdots & & & & & & \\
\gamma_{\delta,k_1^{(1)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(1)}} & & & & & & \\
& & & \gamma_{1,k_1^{(2)}} & \cdots & \gamma_{1,k_{\delta+1}^{(2)}} & & & \\
& & & \vdots & \ddots & \vdots & & & \\
& & & \gamma_{\delta,k_1^{(2)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(2)}} & & & \\
\gamma_{\delta+1,k_1^{(1)}} & \cdots & \gamma_{\delta+1,k_{\delta+1}^{(1)}} & \gamma_{\delta+1,k_1^{(2)}} & \cdots & \gamma_{\delta+1,k_{\delta+1}^{(2)}} & & & \\
& & & & & & \gamma_{1,l_1^{(1)}} & \cdots & \gamma_{1,l_{\delta+2}^{(1)}} \\
& & & & & & \vdots & \ddots & \vdots \\
& & & & & & \gamma_{\delta+1,l_1^{(1)}} & \cdots & \gamma_{\delta+1,l_{\delta+2}^{(1)}} \\
\lambda_{k^{(1)}} \cdot \gamma_{\delta+2,k_1^{(1)}} & \cdots & \lambda_{k^{(1)}} \cdot \gamma_{\delta+2,k_{\delta+1}^{(1)}} & \lambda_{k^{(2)}} \cdot \gamma_{\delta+2,k_1^{(2)}} & \cdots & \lambda_{k^{(2)}} \cdot \gamma_{\delta+2,k_{\delta+1}^{(2)}} & \lambda_{l^{(1)}} \cdot \gamma_{\delta+2,l_1^{(1)}} & \cdots & \lambda_{l^{(1)}} \cdot \gamma_{\delta+2,l_{\delta+2}^{(1)}} \\
\mu_{k^{(1)}} \cdot \gamma_{\delta+3,k_1^{(1)}} & \cdots & \mu_{k^{(1)}} \cdot \gamma_{\delta+3,k_{\delta+1}^{(1)}} & \mu_{k^{(2)}} \cdot \gamma_{\delta+3,k_1^{(2)}} & \cdots & \mu_{k^{(2)}} \cdot \gamma_{\delta+3,k_{\delta+1}^{(2)}} & \mu_{l^{(1)}} \cdot \gamma_{\delta+3,l_1^{(1)}} & \cdots & \mu_{l^{(1)}} \cdot \gamma_{\delta+3,l_{\delta+2}^{(1)}}
\end{bmatrix}
$$

Now, after permuting one row, we can apply A.1.2 to expand the matrix for the determinant,

$$
\det(B_{k,l}) =
$$

$$
\det\begin{pmatrix} \gamma_{1,k_1^{(1)}} & \cdots & \gamma_{1,k_{\delta+1}^{(1)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,k_1^{(1)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(1)}} \\ \gamma_{\delta+1,k_1^{(1)}} & \cdots & \gamma_{\delta+1,k_{\delta+1}^{(1)}} \end{pmatrix}
\det\begin{pmatrix} \gamma_{1,k_1^{(2)}} & \cdots & \gamma_{1,k_{\delta+1}^{(2)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,k_1^{(2)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(2)}} \\ \lambda_{k^{(2)}}\cdot\gamma_{\delta+2,k_1^{(2)}} & \cdots & \lambda_{k^{(2)}}\cdot\gamma_{\delta+2,k_{\delta+1}^{(2)}} \end{pmatrix}
\det\begin{pmatrix} \gamma_{1,l_1^{(1)}} & \cdots & \gamma_{1,l_{\delta+2}^{(1)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l_1^{(1)}} & \cdots & \gamma_{\delta,l_{\delta+2}^{(1)}} \\ \gamma_{\delta+1,l_1^{(1)}} & \cdots & \gamma_{\delta+1,l_{\delta+2}^{(1)}} \\ \mu_{l^{(1)}}\cdot\gamma_{\delta+3,l_1^{(1)}} & \cdots & \mu_{l^{(1)}}\cdot\gamma_{\delta+3,l_{\delta+2}^{(1)}} \end{pmatrix} +
$$

$$
\det\begin{pmatrix} \gamma_{1,k_1^{(1)}} & \cdots & \gamma_{1,k_{\delta+1}^{(1)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,k_1^{(1)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(1)}} \\ \gamma_{\delta+1,k_1^{(1)}} & \cdots & \gamma_{\delta+1,k_{\delta+1}^{(1)}} \end{pmatrix}
\det\begin{pmatrix} \gamma_{1,k_1^{(2)}} & \cdots & \gamma_{1,k_{\delta+1}^{(2)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,k_1^{(2)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(2)}} \\ \mu_{k^{(2)}}\cdot\gamma_{\delta+3,k_1^{(2)}} & \cdots & \mu_{k^{(2)}}\cdot\gamma_{\delta+3,k_{\delta+1}^{(2)}} \end{pmatrix}
\det\begin{pmatrix} \gamma_{1,l_1^{(1)}} & \cdots & \gamma_{1,l_{\delta+2}^{(1)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l_1^{(1)}} & \cdots & \gamma_{\delta,l_{\delta+2}^{(1)}} \\ \gamma_{\delta+1,l_1^{(1)}} & \cdots & \gamma_{\delta+1,l_{\delta+2}^{(1)}} \\ \lambda_{l^{(1)}}\cdot\gamma_{\delta+2,l_1^{(1)}} & \cdots & \lambda_{l^{(1)}}\cdot\gamma_{\delta+2,l_{\delta+2}^{(1)}} \end{pmatrix} +
$$

$$
\det\begin{pmatrix} \gamma_{1,k_1^{(1)}} & \cdots & \gamma_{1,k_{\delta+1}^{(1)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,k_1^{(1)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(1)}} \\ \lambda_{k^{(1)}}\cdot\gamma_{\delta+2,k_1^{(1)}} & \cdots & \lambda_{k^{(1)}}\cdot\gamma_{\delta+2,k_{\delta+1}^{(1)}} \end{pmatrix}
\det\begin{pmatrix} \gamma_{1,k_1^{(1)}} & \cdots & \gamma_{1,k_{\delta+1}^{(1)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,k_1^{(1)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(1)}} \\ \gamma_{\delta+1,k_1^{(2)}} & \cdots & \gamma_{\delta+1,k_{\delta+1}^{(2)}} \end{pmatrix}
\det\begin{pmatrix} \gamma_{1,l_1^{(1)}} & \cdots & \gamma_{1,l_{\delta+2}^{(1)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l_1^{(1)}} & \cdots & \gamma_{\delta,l_{\delta+2}^{(1)}} \\ \gamma_{\delta+1,l_1^{(1)}} & \cdots & \gamma_{\delta+1,l_{\delta+2}^{(1)}} \\ \mu_{l^{(1)}}\cdot\gamma_{\delta+3,l_1^{(1)}} & \cdots & \mu_{l^{(1)}}\cdot\gamma_{\delta+3,l_{\delta+2}^{(1)}} \end{pmatrix} -
$$

$$
\det\begin{pmatrix} \gamma_{1,k_1^{(1)}} & \cdots & \gamma_{1,k_{\delta+1}^{(1)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,k_1^{(1)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(1)}} \\ \mu_{k^{(1)}}\cdot\gamma_{\delta+3,k_1^{(1)}} & \cdots & \mu_{k^{(1)}}\cdot\gamma_{\delta+3,k_{\delta+1}^{(1)}} \end{pmatrix}
\det\begin{pmatrix} \gamma_{1,k_1^{(2)}} & \cdots & \gamma_{1,k_{\delta+1}^{(2)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,k_1^{(2)}} & \cdots & \gamma_{\delta,k_{\delta+1}^{(2)}} \\ \gamma_{\delta+1,k_1^{(2)}} & \cdots & \gamma_{\delta+1,k_{\delta+1}^{(2)}} \end{pmatrix}
\det\begin{pmatrix} \gamma_{1,l_1^{(1)}} & \cdots & \gamma_{1,l_{\delta+2}^{(1)}} \\ \vdots & \ddots & \vdots \\ \gamma_{\delta,l_1^{(1)}} & \cdots & \gamma_{\delta,l_{\delta+2}^{(1)}} \\ \gamma_{\delta+1,l_1^{(1)}} & \cdots & \gamma_{\delta+1,l_{\delta+2}^{(1)}} \\ \lambda_{l^{(1)}}\cdot\gamma_{\delta+2,l_1^{(1)}} & \cdots & \lambda_{l^{(1)}}\cdot\gamma_{\delta+2,l_{\delta+2}^{(1)}} \end{pmatrix}
$$

Now, in this massive expansion, we can take λ_i and μ_j out of the determinants. What we will find is that each term is $\lambda_i\mu_j$ multiplied by the product of the determinant of three Cauchy matrices. Each of those determinant $\in \mathbb{F}_{q_0}$.

Hence the final determinant is actually the linear combination of three λ_i in $\mathbb{F}_{q_0}$. Hence $\det(B_{k,l}) \neq 0$.

Bibliography

[1] M. Sathiamoorthy, M. Asteris, D. Papailiopoulos, A. G. Dimakis, R. Vadali, S. Chen, and D. Borthakur, "Xoring elephants: Novel erasure codes for big data," *Proc. VLDB Endow.*, vol. 6, no. 5, pp. 325–336, Mar. 2013. [Online]. Available: http://dx.doi.org/10.14778/2535573.2488339

[2] C. Huang, H. Simitci, Y. Xu, A. Ogus, B. Calder, P. Gopalan, J. Li, and S. Yekhanin, "Erasure coding in windows azure storage," in *Presented as part of the 2012 USENIX Annual Technical Conference (USENIX ATC 12)*. Boston, MA: USENIX, 2012, pp. 15–26. [Online]. Available: https://www.usenix.org/conference/atc12/technical-sessions/presentation/huang

[3] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols," *IEEE Transactions on Information Theory*, vol. 58, no. 11, pp. 6925–6934, Nov 2012.

[4] C. Huang, M. Chen, and J. Li, "Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems," in *Sixth IEEE International Symposium on Network Computing and Applications (NCA 2007)*, July 2007, pp. 79–86.

[5] I. Tamo and A. Barg, "A family of optimal locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4661–4676, Aug. 2014.

[6] A. S. Rawat, D. S. Papailiopoulos, A. G. Dimakis, and S. Vishwanath, "Locality and availability in distributed storage," *IEEE Transactions on Information Theory*, vol. 62, no. 8, pp. 4481–4493, Aug 2016.

[7] T. Ernvall, T. Westerbäck, and C. Hollanti, "Constructions of optimal and almost optimal locally repairable codes," in *2014 4th International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems (VITAE)*. IEEE, 2014, pp. 1–5.

[8] S. Goparaju and R. Calderbank, "Binary cyclic codes that are locally repairable," in *2014 IEEE International Symposium on Information Theory*, 2014, pp. 676–680.

[9] J. Hao, S.-T. Xia, and B. Chen, "On optimal ternary locally repairable codes," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 171–175.

[10] B. Chen, W. Fang, S.-T. Xia, J. Hao, and F.-W. Fu, "Improved bounds and singleton-optimal constructions of locally repairable codes with minimum distance 5 and 6," *IEEE Transactions on Information Theory*, vol. 67, no. 1, pp. 217–231, 2020.

[11] P. Gopalan, C. Huang, B. Jenkins, and S. Yekhanin, "Explicit maximally recoverable codes with locality," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5245–5256, Sep. 2014.

[12] V. Cadambe and A. Mazumdar, "An upper bound on the size of locally recoverable codes," in *2013 International Symposium on Network Coding (NetCod)*, June 2013, pp. 1–5.

[13] J. Hao, S. Xia, B. Chen, and F. Fu, "On the weight hierarchy of locally repairable codes," in *2017 IEEE Information Theory Workshop (ITW)*, Nov 2017, pp. 31–35.

[14] J. Hao, K. Shum, S. Xia, and Y. Yang, "On the maximal code length of optimal linear locally repairable codes," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 1326–1330.

[15] V. Guruswami, C. Xing, and C. Yuan, "How long can optimal locally repairable codes be?" *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3662–3670, June 2019.

[16] P. Gopalan, G. Hu, S. Kopparty, S. Saraf, C. Wang, and S. Yekhanin, "Maximally recoverable codes for grid-like topologies," in *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA '17. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2017, pp. 2092–2108. [Online]. Available: http://dl.acm.org/citation.cfm?id=3039686.3039822

[17] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar, "Optimal linear codes with a local-error-correction property," in *2012 IEEE International Symposium on Information Theory Proceedings*, July 2012, pp. 2776–2780.

[18] G. M. Kamath, N. Prakash, V. Lalitha, and P. V. Kumar, "Codes with local regeneration and erasure correction," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4637–4660, Aug 2014.

[19] M. Grezet, R. Freij-Hollanti, T. Westerbäck, and C. Hollanti, "Alphabet-dependent bounds for linear locally repairable codes based on residual codes," *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6089–6100, Oct 2019.

[20] H. Cai, Y. Miao, M. Schwartz, and X. Tang, "On optimal locally repairable codes with super-linear length," *CoRR*, vol. abs/1812.11942, 2018. [Online]. Available: http://arxiv.org/abs/1812.11942

[21] I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis, "Optimal locally repairable codes and connections to matroid theory," *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 6661–6671, 2016.

[22] W. Song, S. H. Dau, C. Yuen, and T. J. Li, "Optimal locally repairable linear codes," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 5, pp. 1019–1036, 2014.

[23] S. Kadhe, S. El Rouayheb, I. Duursma, and A. Sprintson, "Codes with locality in the rank and subspace metrics," *IEEE Transactions on Information Theory*, vol. 65, no. 9, pp. 5454–5468, 2019.

[24] G. Luo, M. F. Ezerman, and S. Ling, "Three new constructions of optimal locally repairable codes from matrix-product codes," *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 75–85, 2023.

[25] I. Tamo, A. Barg, S. Goparaju, and R. Calderbank, "Cyclic LRC codes and their subfield subcodes," in *2015 IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 1262–1266.

[26] B. Chen, S.-T. Xia, J. Hao, and F.-W. Fu, "Constructions of optimal cyclic (r, δ) locally repairable codes," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2499–2511, 2017.

[27] A. Barg, I. Tamo, and S. Vlăduţ, "Locally recoverable codes on algebraic curves," *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 4928–4939, 2017.

[28] X. Li, L. Ma, and C. Xing, "Optimal locally repairable codes via elliptic curves," *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 108–117, 2019.

[29] B. Kim, "Locally recoverable codes in hermitian function fields with certain types of divisors," *AIMS Mathematics*, vol. 7, no. 6, pp. 9656–9667, 2022.

[30] B. Sasidharan, G. K. Agarwal, and P. V. Kumar, "Codes with hierarchical locality," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1257–1261.

[31] G. Zhang and H. Liu, "Constructions of optimal codes with hierarchical locality," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7333–7340, Dec. 2020.

[32] B. Chen, G. Zhang, and W. Li, "New optimal linear codes with hierarchical locality," *IEEE Transactions on Information Theory*, vol. 69, no. 3, pp. 1544–1550, 2023.

[33] S. Ballentine, A. Barg, and S. Vlăduţ, "Codes with hierarchical locality from covering maps of curves," *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6056–6071, 2019.

[34] G. Luo and X. Cao, "Optimal cyclic codes with hierarchical locality," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3302–3310, Jun. 2020.

[35] S. Ballentine, A. Barg, and S. Vlăduţ, "Codes with hierarchical locality from covering maps of curves," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6056–6071, Oct. 2019.

[36] M. Grezet and C. Hollanti, "The complete hierarchical locality of the punctured simplex code," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 2833–2837.

[37] A. M. Nair and V. Lalitha, "Maximally recoverable codes with hierarchical locality," in *2019 National Conference on Communications (NCC)*, 2019, pp. 1–6.

[38] S. Gopi, V. Guruswami, and S. Yekhanin, "Maximally recoverable LRCs: A field size lower bound and constructions for few heavy parities," *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6066–6083, 2020.

[39] M. Blaum, J. L. Hafner, and S. Hetzler, "Partial-MDS codes and their application to RAID type of architectures." *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4510–4519, Jul. 2013.

[40] G. Calis and O. O. Koyluoglu, "A general construction for PMDS codes," *IEEE Communications Letters*, vol. 21, no. 3, pp. 452–455, 2017.

[41] M. Blaum, J. S. Plank, M. Schwartz, and E. Yaakobi, "Construction of partial MDS and sector-disk codes with two global parity symbols," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2673–2681, May 2016.

[42] J. Chen, K. W. Shum, Q. Yu, and C. W. Sung, "Sector-disk codes and partial MDS codes with up to three global parities," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 1876–1880.

[43] R. Gabrys, E. Yaakobi, M. Blaum, and P. H. Siegel, "Constructions of partial MDS codes over small fields," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3692–3701, Jun. 2019.

[44] G. Hu and S. Yekhanin, "New constructions of SD and MR codes over small finite fields," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1591–1595.

[45] V. Guruswami, L. Jin, and C. Xing, "Constructions of maximally recoverable local reconstruction codes via function fields," *IEEE Trans. Inf. Theory*, vol. 66, no. 10, pp. 6133–6143, Oct. 2020.

[46] U. Martínez-Peñas and F. R. Kschischang, "Universal and dynamic locally repairable codes with maximal recoverability via sum-rank codes," *IEEE Transactions on Information Theory*, vol. 65, no. 12, pp. 7790–7805, 2019.

[47] S. Gopi and V. Guruswami, "Improved maximally recoverable LRCs using skew polynomials," *IEEE Trans. Inf. Theory*, early access, Sep. 30, 2022, doi: 10.1109/TIT.2022.3176807.

[48] H. Cai, Y. Miao, M. Schwartz, and X. Tang, "A construction of maximally recoverable codes with order-optimal field size," *IEEE Transactions on Information Theory*, vol. 68, no. 1, pp. 204–212, 2021.

[49] H. Cai and M. Schwartz, "A bound on the minimal field size of LRCs, and cyclic MR codes that attain it," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 2625–2630.

[50] V. Lalitha and S. V. Lokam, "Weight enumerators and higher support weights of maximally recoverable codes," in *Proc. 53rd Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep./Oct. 2015, pp. 835–842.

[51] M. Grezet, T. Westerbäck, R. Freij-Hollanti, and C. Hollanti, "Uniform minors in maximally recoverable codes," *IEEE Commun. Lett.*, vol. 23, no. 8, pp. 1297–1300, Aug. 2019.

[52] L. Holzbaur, S. Puchinger, E. Yaakobi, and A. Wachter-Zeh, "Correctable erasure patterns in product topologies," in *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 2054–2059.

[53] V. Gandikota, E. Grigorescu, C. Thomas, and M. Zhu, "Maximally recoverable codes: The bounded case," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2017, pp. 1115–1122.

[54] X. Kong, J. Ma, and G. Ge, "New bounds on the field size for maximally recoverable codes instantiating grid-like topologies," *Journal of Algebraic Combinatorics*, vol. 54, no. 2, pp. 529–557, 2021.

[55] J. Brakensiek, S. Gopi, and V. Makam, "Lower bounds for maximally recoverable tensor codes and higher order MDS codes," *IEEE Transactions on Information Theory*, vol. 68, no. 11, pp. 7125–7140, 2022.

[56] S. Gopi, V. Guruswami, and S. Yekhanin, "Maximally recoverable LRCs: A field size lower bound and constructions for few heavy parities," *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6066–6083, 2020.

[57] U. Martínez-Peñas, "Skew and linearized Reed–Solomon codes and maximum sum rank distance codes over any division ring," *Journal of Algebra*, vol. 504, pp. 587–612, 2018.

[58] D. Shivakrishna, A. M. Nair, and V. Lalitha, "Maximally recoverable codes with hierarchical locality: Constructions and field-size bounds," *IEEE Transactions on Information Theory*, vol. 69, no. 2, pp. 842–859, 2023.

[59] A. Neri and A.-L. Horlemann-Trautmann, "Random construction of partial MDS codes," *Designs, Codes and Cryptography*, vol. 88, no. 4, pp. 711–725, Dec. 2019. [Online]. Available: https://doi.org/10.1007/s10623-019-00705-x

[60] M. Chen, C. Huang, and J. Li, "On the maximally recoverable property for multi-protection group codes," in *2007 IEEE International Symposium on Information Theory*, 2007, pp. 486–490.

[61] F. MacWilliams and N. Sloane, *The theory of error correcting codes*. Elsevier, 1977.

[62] J. Hirschfeld, *Projective geometries over finite fields. Oxford Mathematical Monographs*. Oxford University Press New York, 1998.

[63] S. H. Dau, W. Song, Z. Dong, and C. Yuen, "Balanced sparsest generator matrices for MDS codes," in *2013 IEEE International Symposium on Information Theory*, July 2013, pp. 1889–1893.

[64] S. H. Dau, W. Song, and C. Yuen, "On the existence of MDS codes over small fields with constrained generator matrices," in *2014 IEEE International Symposium on Information Theory*, June 2014, pp. 1787–1791.

[65] A. Heidarzadeh and A. Sprintson, "An algebraic-combinatorial proof technique for the GM-MDS conjecture," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 11–15.

[66] W. Halbawi, Z. Liu, and B. Hassibi, "Balanced Reed-Solomon codes," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 935–939.

[67] W. Halbawi, Z. Liu, I. M. Duursma, H. Dau, and B. Hassibi, "Sparse and balanced Reed–Solomon and Tamo–Barg codes," *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 118–130, Jan 2019.

[68] H. Yildiz and B. Hassibi, "Gabidulin codes with support constrained generator matrices," *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3638–3649, 2020.

[69] W. Halbawi, Z. Liu, I. M. Duursma, H. Dau, and B. Hassibi, "Sparse and balanced Reed–Solomon and Tamo–Barg codes," *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 118–130, 2019.

[70] L. Pamies-Juarez, H. D. Hollmann, and F. Oggier, "Locally repairable codes with multiple repair alternatives," in *2013 IEEE international symposium on information theory*. IEEE, 2013, pp. 892–896.

[71] A. Wang and Z. Zhang, "Repair locality with multiple erasure tolerance," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6979–6987, 2014.

[72] I. Tamo and A. Barg, "Bounds on locally recoverable codes with multiple recovering sets," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 691–695.

[73] S. Bhadane and A. Thangaraj, "Unequal locality and recovery for locally recoverable codes with availability," in *2017 Twenty-third National Conference on Communications (NCC)*. IEEE, 2017, pp. 1–6.

[74] T. Westerbäck, R. Freij-Hollanti, and C. Hollanti, "Applications of polymatroid theory to distributed storage systems," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2015, pp. 231–237.

[75] R. Roth, *Introduction to coding theory*. Cambridge, U.K.: Cambridge Univ. Press, 2006.