Locally Recoverable Codes with Availability

A Project Report

submitted by

SOURBH NITIN BHADANE

in partial fulfilment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY AND MASTER OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

May 2017

THESIS CERTIFICATE

This is to certify that the thesis titled **Locally Recoverable Codes with Availability**, submitted by **Sourbh Nitin Bhadane**, to the Indian Institute of Technology, Madras, for the award of the degree of **Bachelor of Technology and Master of Technology**, is a bona fide record of the research work done by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Andrew Thangaraj
Research Guide
Professor
Dept. of Electrical Engineering
IIT Madras, 600 036

Place: Chennai

Date: 5th May 2017

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to Prof. Andrew Thangaraj for guiding me throughout the entire project, not only about the nitty gritties associated with the project but also about research in general. His enthusiasm towards research, intuition about possible solutions to a problem and amazing teaching skills are some of the many qualities I wish to imbibe.

I would like to thank the entire EE department faculty for providing an outstanding educational foundation that helped me find my interests and build upon it.

I would like to thank Prof. David, Prof. Hari, Prof. Nitin, Akshay and all my friends who worked on the IITMSAT project. Apart from the enriching technical experience that this project provided, I have had the opportunity to experience first-hand how much patience and perseverance is required to convert theory into practice. I will forever cherish the friendships forged through staying up late night brainstorming, debugging a weird software bug or a hardware malfunction.

I would like to thank all my friends at IIT-M. Without their unwavering companionship this 5 year journey would have been absolutely drab. I would like to thank my hockey teammates (Go Garibos !) for the wonderful time we had together.

Lastly, I want to thank my parents Sujata Bhadane, Nitin Bhadane and my sister Yuktha whose constant support and love throughout my life has been the sole reason I have managed to remain sane.

ABSTRACT

KEYWORDS: Coding theory, Erasure-correcting codes, Codes for distributed storage, Locally recoverable codes

Modern distributed storage systems are prone to node failures. Codes for distributed storage are found to efficiently repair node failures as compared to traditional methods. Locally Recoverable Codes (LRCs) are one such class of codes that minimize the number of nodes required in the repair process. A code is said to be a Locally Recoverable Code (LRC) with availability if every coordinate can be recovered from multiple disjoint sets of other coordinates called recovering sets. The sizes of recovering sets of a coordinate is called its recovery profile. In this work, we consider LRCs with availability under two different settings: (1) irregular recovery: non-constant recovery profile that remains fixed for all coordinates, (2) unequal locality: regular recovery profile that can vary with coordinates. For each setting, we derive bounds for the minimum distance that generalize previously known bounds to the cases of irregular or varying recovery profiles. For the case of regular and fixed recovery profile, we show that a specific Tamo-Barg polynomialevaluation construction is optimal for all-symbol locality, and we provide improved parity-check matrix constructions over smaller fields for information locality.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS **i**

ABSTRACT **ii**

ABBREVIATIONS **vi**

NOTATION **vii**

1 Introduction **1**

1.1 Locally Recoverable Codes (LRCs) 2

1.1.1 LRCs with Availability 3

1.1.2 Codes with Unequal Locality 3

1.2 Organization of the thesis and Contributions 4

2 Problem Setup and Preliminaries **6**

2.1 Upper Bounds on Distance for LRCs with Availability 6

2.2 Polynomial-evaluation Construction for LRCs 7

2.2.1 Example 8

2.3 Parity-Check Matrix Approach to LRCs 10

3 Bounds on Codes with Irregular Recovery and Unequal Locality with Availability **12**

3.1 Irregular Recovery with Availability 12

3.1.1 Bounds for Information Locality 12

3.1.2 Bounds for All-symbol Locality 15

3.2 Unequal Locality with Availability 19

3.2.1 Bounds for Information Locality 19

4 Optimal Constructions for LRCs with Availability **22**

4.1 Regular Recovery and Locality with Availability 22
4.1.1 All-symbol Locality 22
4.1.2 Parity-check Matrix Construction for Information Locality . . . 25
4.2 Irregular Recovery with Availability 28

5 Miscellanous Results using the Parity-Check Matrix approach 30
5.1 Simpler Proof of Distance Upper Bound for (r, δ) codes 30
5.2 Linear Codes cannot achieve Tamo-Barg Rate Upper Bound 31

6 Conclusions and Scope for Future Work 34

A Distance Upper Bound Algorithm 35

LIST OF TABLES

1. Table 4.1: Lower and upper bound on minimum distance for Example 1 21

ABBREVIATIONS

LRC	Locally Recoverable Code
RS	Reed-Solomon
HDFS	Hadoop Distributed File System
GFS	Google File System
I/O	Input/Output
MDS	Maximum Distance Separable

NOTATION

$supp(x)$	Support of the vector x
d	Distance of a code
$\mathbb{F}$	Finite field
$\mathbb{F}_q^n$	Extension field of order q^n, characteristic q
$[n]$	Set of natural numbers from 1 to n
$\lceil x \rceil$	Smallest integer greater than or equal to x
$\lfloor x \rfloor$	Greatest integer lesser than or equal to x
$V(G)$	Set of vertices of the graph G
$deg(f)$	Degree of the polynomial f

CHAPTER 1

Introduction

In today's digital age, organizations and individuals generate massive amounts of data every day. For example, recent estimates show that 2.5 exabytes of data are produced across the world on a daily basis. With the advent of numerous data-driven applications and social media platforms in this modern era of big data, large-scale data storage has become increasingly critical. Since dependence on data is inevitable, end users of such data-driven applications cannot afford to lose data, making reliability of data storage a major necessity.

Data centers these days have the capability of storing exabytes of data. For various reasons such as scalability, cost effectiveness and failure resilience, modern storage systems predominantly follow the distributed storage paradigm, wherein a massive amount of data is stored across a large number of inexpensive and unreliable storage devices resulting in a highly reliable storage system. Typically, a distributed storage system contains multiple clusters that house hundreds of servers interconnected by switches. Data is stored on a such a system using a distributed file system such as Google File System (GFS), Apache Hadoop etc.

Modern distributed storage systems are prone to node failures. Sources of these failures include power outage, hardware and software failures, maintenance related shutdowns. Bringing about a high degree of reliability in a distributed storage system requires introducing some form of redundancy. Traditionally, data centers have been using replication of data as a means to provide reliability and availability. For instance, data centers based on the Hadoop Distributed File System (HDFS) employed 3x replication for all of their data. Naturally, data replication is resource-intensive and involves a lot of space overhead.

Erasure coding techniques provide a much more efficient way to introduce redundancy as compared to data replication. In a distributed storage system, erasure codes are used in the following manner : A file is divided into k blocks which are encoded into n blocks and stored in n different nodes. In such a system, data stored in a single node can be recovered by accessing k other nodes in case of a failure. Reed-Solomon (RS) codes are a popular choice of erasure codes used in storage systems, because of the Maximum Distance Separable (MDS) property of these codes wherein any k coordinates out of n coordinates of an (n, k) RS code suffice to recover the entire codeword. For example, Facebook uses a $(14, 10)$ RS code for it's cold data, thus bringing down the storage overhead by 60% as compared to 2x replication.

Although traditional erasure codes reduce storage overhead, they perform poorly in terms of other metrics such as number of bits communicated during the repair process i.e repair bandwidth, number of disk I/O's, number of nodes participating in the repair process, i.e. repair locality etc.. For example, a (n, k) RS code has a repair locality of k whereas 2x replication has a repair locality of only one. Specialized codes for optimizing over each of these metrics have been designed and studied by coding theorists. There has been recent active interest in two classes of codes, namely *Regenerating Codes* which seek to minimize repair bandwidth and *Locally Recoverable Codes* that seek to minimize repair locality.

1.1 Locally Recoverable Codes (LRCs)

An $[n, k, d, r]$ code is said to be a Locally Recoverable Code (LRC) if every coordinate can be recovered from atmost r other coordinates. These r other coordinates are called the recovering set or the repair group of the failed coordinate and the parameter r is termed as the locality of the coordinate. An LRC is said to have *information locality* r if all information coordinates have locality r. If all coordinates have locality r, an LRC is said to have *all-symbol locality*. LRCs were first introduced in the seminal paper Gopalan *et al.* (2012) and a Singleton-like upper bound on the minimum distance was derived. Constructions meeting this bound with exponential field size were proposed in Silberstein *et al.* (2013), Tamo *et al.* (2013), Hao and Xia (2016). An alphabet-size dependent bound was obtained in Cadambe and Mazumdar (2015). A parity-check matrix approach was used in Hao and Xia (2016) to construct optimal LRC codes. An elegant algebraic optimal construction with field size linear in blocklength was presented in Tamo and Barg (2014).

1.1.1 LRCs with Availability

LRCs were originally intended to minimize the number of nodes accessed to recover from a single node failure. Although single node failures are most frequent, LRCs with multiple disjoint *recovering sets* are useful for recovering from multiple concurrent node failures. Moreover, this property could also be exploited for the storage of "hot" data, which may be served to several users simultaneously using the recovering sets in parallel. As a result, LRCs with multiple disjoint recovering sets are also referred to as LRCs with *availability*. LRCs with availability have been studied in Tamo and Barg (2014); Pamies-Juarez *et al.* (2013); Rawat *et al.* (2014); Wang and Zhang (2014); Tamo *et al.* (2016); Huang *et al.* (2015). An upper bound on the minimum distance for an $[n, k, d]$ LRC with information locality r and availability t is given in Wang and Zhang (2014). For $n \geq k\,(tr + 1)$, Wang and Zhang (2014) proves existence of codes that meet the above bound. However, to the best of our knowledge, no explicit constructions meeting this bound are known. An alphabet-dependent upper bound for linear codes with information locality and availability was derived in Huang *et al.* (2015). In Tamo *et al.* (2016), an upper bound on minimum distance was derived for the all-symbol locality and availability case. To the best of our knowledge, no general constructions that attain this bound are known. Motivated by recovering from multiple erasures, Prakash *et al.* (2012) define the notion of (r, δ) codes wherein each coordinate is protected by a local code of length $r + \delta - 1$ and minimum distance atleast δ. They also derive Singleton-like upper bounds on the minimum distance of (r, δ) codes. Another line of work Song *et al.* (2016) is recovering from multiple erasures in a sequential manner. Note that although all these different approaches recover from multiple erasures only LRCs with availability enable parallel access.

1.1.2 Codes with Unequal Locality

Recently, Kadhe and Sprintson (2016), Zeh and Yaakobi (2016), Chen *et al.* (2017), Kim and Lee (2017) studied codes with unequal locality. Codes with unequal locality are practically significant in scenarios wherein different locality requirements are needed for different blocks of data. For example, since hot data needs to be accessed quickly, one might need a small locality for hot data whereas cold data can tolerate larger locality. Upper bounds on minimum distance for codes with unequal information locality and unequal all-symbol locality were obtained in Kadhe and Sprintson (2016). Constructions based on an adaptation of Pyramid codes and rank-metric codes were proposed to attain these bounds, respectively. Chen *et al.* (2017), Kim and Lee (2017) considered an extension of (r, δ) localities to unequal locality. However, to the best of our knowledge, none of these works consider LRCs with availability.

1.2 Organization of the thesis and Contributions

In an LRC with availability, the sizes of recovering sets of a particular coordinate is called its recovery profile, which is said to be regular if all sizes are equal, and irregular otherwise. In this work, we extend codes with availability to include irregular and varying recovery profiles. Specifically, we study the following two settings, which have not been studied before:

a) *Irregular recovery*: recovery profile can have varying recovering set sizes but remains fixed for all coordinates, i.e. t disjoint recovering sets with sizes $r_1, r_2, \ldots, r_t$ for all coordinates,

b) *Unequal locality*: regular recovery profile that may vary over coordinates, i.e. t disjoint recovering sets each of size r_i for coordinate i, $1 \leq i \leq n$.

Upper bounds on minimum distance are obtained for both settings under information and all-symbol locality. We also present a generalization of an existing construction from Tamo and Barg (2014) and prove that it meets the distance bound for arbitrary t and $r = k - 1$. For information locality and availability, we extend the construction in Hao and Xia (2016) and provide an explicit parity-check matrix construction that meets the distance upper bound for $n \geq k\,(tr + 1)$ and improves upon the implicit construction in Wang and Zhang (2014) in terms of field size for some parameters.

The rest of the thesis is organized in the following manner.

Chapter 2 provides a technical definition of LRCs and notation that we will be using throughout the thesis. We also provide brief details of a polynomial evaluation construction by Tamo and Barg (2014) that will form the foundation of Chapter 4.

Chapter 3 focuses on codes with irregular recovery and unequal locality and availability. We derive upper bounds on the distance for the case of information locality and all-symbol locality for the case of codes with irregular recovery. We derive distance upper bounds for the case of information locality for codes with unequal locality with availability and provide a brief note on the difficulty of the problem setup for the all-symbol locality case

Chapter 4 contains an optimal parity-check matrix construction for LRCs with information locality and availability subject to rate constraints. In addition, we prove optimality of an existing construction in the case of LRCs with all-symbol locality and availability.

Chapter 5 presents a couple of results that use the parity-check matrix viewpoint of LRCs. However, these are not connected to the main theme of the thesis.

Chapter 6 summarizes Chapters 3,4 and presents future directions that could be pursued.

CHAPTER 2

Problem Setup and Preliminaries

We first provide a technical definition of codes with locality r and availability t.

Consider an $[n, k, d]$ linear code $\mathcal{C} \subseteq \mathbb{F}_q^n$, where q is a prime power and $\mathbb{F}_q$ is the finite field with q elements. Suppose a subset $D \subseteq [n] \triangleq \{1, 2, \ldots, n\}$ is the support of a dual codeword. Then, for every $i \in D$, the i-th coordinate of a codeword of $\mathcal{C}$ is a linear combination of the coordinates in $D \setminus \{i\}$. The code $\mathcal{C}$ is said to have locality r and availability t if, for $i \in [n]$, there exist t dual-codeword support sets $D_j^{(i)}$, $j \in [t]$ such that

1. $i \in \left|D_j^{(i)}\right|$
2. $R_j^{(i)} = D_j^{(i)} \setminus \{i\}$ are disjoint
3. $\left|R_j^{(i)}\right| \leq r$.

The sets $R_j^{(i)}$ are called the *recovering sets* for i because the coordinate i can be recovered from the coordinates in any of its recovering sets. For $i \in [n]$, we denote $\Gamma_a(i) = \{i\} \cup R_1^{(i)} \cup R_2^{(i)} \cdots \cup R_a^{(i)}$ for $1 \leq a \leq t$. A code with locality is referred to as a Locally Recoverable Code (LRC).

2.1 Upper Bounds on Distance for LRCs with Availability

For LRCs with no availability constraints, Gopalan *et al.* (2012) derived the following Singleton-like upper bound on the distance for information locality.

$$d \leq n - k - \left\lceil \frac{k}{r} \right\rceil + 2 \tag{2.1}$$

It is interesting to note that the above upper bound also holds for LRCs with all-symbol locality as shown in Tamo and Barg (2014).

LRCs with information locality and availability were considered in Wang and Zhang (2014) and the following distance upper bound was obtained for a $[n, k, d]$ LRC with locality r and availability t

$$d \leq n - k - \left\lceil \frac{t\,(k-1)+1}{t\,(r-1)+1} \right\rceil + 2 \tag{2.2}$$

Wang and Zhang (2014) prove existence of optimal codes that meet (2.2) for $n \geq k\,(tr+1)$. In Chapter 5, we provide an explicit parity-check matrix construction that meets the above bound with a field size smaller than that in Wang and Zhang (2014).

For LRCs with all-symbol locality and availability, Tamo *et al.* (2016) derived the following distance upper bound using concepts of recovery graphs and expansion ratios

$$d \leq n - \sum_{i=0}^{t} \left\lfloor \frac{k-1}{r^i} \right\rfloor \tag{2.3}$$

No codes that meet the above construction over a non-trivial range of parameters are known. In Chapter 5, we prove that an extension of the polynomial evaluation construction of Tamo and Barg (2014) meets the above upper bound with equality for $k = r + 1$ and arbitrary t. Note that for $t = 1$, (2.2) and (2.3) reduce to (2.1).

2.2 Polynomial-evaluation Construction for LRCs

We now describe the polynomial-evaluation construction of LRCs with availability from Tamo and Barg (2014).

Let $A \subseteq \mathbb{F}$ ($\mathbb{F}$ is a finite field), $|A| = n$. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two partitions of A into m_1 and m_2 sets such that for any two sets $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$, we have $|A_1| = r_1$, $|A_2| = r_2$, and the size of their intersection $|A_1 \cap A_2| \leq 1$. Such partitions are called orthogonal partitions. Define for $\mathcal{A}_i$, $i = 1, 2$,

$$\mathbb{F}_{\mathcal{A}_i}[x] = \{f \in \mathbb{F}[x] : f \text{ is constant on } A_j \in \mathcal{A}_i, j \in [m_i]\,, \text{deg } f \leq |A_j|\}\,.$$

Further, define two families of polynomials

$$\mathcal{F}_{\mathcal{A}_1}^{r_1} = \oplus_{i=0}^{r_1-1}\mathbb{F}_{\mathcal{A}_1}[x]x^i, \quad \mathcal{F}_{\mathcal{A}_2}^{r_2} = \oplus_{i=0}^{r_2-1}\mathbb{F}_{\mathcal{A}_2}[x]x^i.$$

Let $\mathcal{P}_m$ be the space of polynomials of degree less than or equal to m. Define

$$V_m = \mathcal{F}_{\mathcal{A}_1}^{r_1} \cap \mathcal{F}_{\mathcal{A}_2}^{r_2} \cap \mathcal{P}_m$$

Let $\phi : \mathbb{F}^k \to V_m$ be an injective linear mapping with basis $g_0, \cdots, g_{k-1}$ in V_m such that

$$\phi(a) = \sum_{i=0}^{k-1} a_i g_i(x)$$

Note that $\phi(a)$ is a polynomial in x and denote it by $f(x)$. A codeword of a length-n LRC with availability $t = 2$ is obtained by evaluating f on all n points of A. If the number of such polynomials of degree at most m is $|\mathbb{F}|^k$, we obtain an (n, k, d) availability-2 LRC with minimum distance $d \geq n - m$. A set $A_i \in \mathcal{A}_i$ is a dual codeword support set because the r_i points of A_i pass through a polynomial of degree at most $r_i - 1$. Therefore, by the above construction each coordinate has two disjoint recovering sets because the partitions are orthogonal.

Following Tamo and Barg (2014), a partition is naturally formed by a subgroup H of the multiplicative or additive group of $\mathbb{F}$ and cosets of H. A degree-$|H|$ polynomial constant on such partitions is

$$g(x) = \prod_{h \in H} (x - h).$$

Such a polynomial is called the annihilator polynomial of H. If H is a multiplicative subgroup of $\mathbb{F}_q^*$, then $g(x) = x^{|H|}$ is constant on each coset of H.

2.2.1 Example

We illustrate the above construction using the following example of a $(12, 4, [2, 3])$ LRC wherein partitions are generated using multiplicative subgroups.

Let $A = \mathbb{F}_{13} \backslash \{0\}$ and let $\mathcal{A}_1$ and $\mathcal{A}_2$ be generated by cosets of the multiplicative cyclic groups generated by 5 and 3 respectively.

$$\mathcal{A}_1 = \{\{1, 5, 12, 8\}, \{2, 10, 11, 3\}, \{4, 7, 9, 6\}\}$$

$$\mathcal{A}_2 = \{\{1, 3, 9\}, \{2, 6, 5\}, \{4, 12, 10\}\} \{7, 8, 11\}$$

Using the annihilator polynomial of each of the cyclic subgroups,

$$\mathbb{F}_{\mathcal{A}_1}[x] = \langle 1, x^4, x^8 \rangle, \mathbb{F}_{\mathcal{A}_2}[x] = \langle 1, x^3, x^6, x^9 \rangle$$

It is easy to check that

$$\mathcal{F}^{r_1}_{\mathcal{A}_1} \cap \mathcal{F}^{r_1}_{\mathcal{A}_2} = \langle 1, x, x^4, x^6, x^9, x^{10} \rangle$$

Since, $k = 4$, $V_m = \langle 1, x, x^4, x^6 \rangle$. Therefore,

$$f_a(x) = a_o + a_1 x + a_2 x^4 + a_3 x^6$$

This same polynomial can be represented in the following two different ways,

$$\begin{aligned} f_a(x) &= \left(a_0 + a_2 x^4\right) + a_1 x + \left(a_3 x^4\right) x^2 \\ &= \left(a_0 + a_3 x^6\right) + \left(a_1 + a_2 x^2\right) x \end{aligned}$$

The above two representations illustrate how a failed coordinate can be recovered using two disjoint set of symbols. For example, let the failed coordinate be $f_a(1)$. Then from the first representation we can see that a unique degree 2 polynomial passes through $\{5, 12, 8\}$. From this polynomial, we can recover the value at $\{1\}$ by evaluating the value of the degree two polynomial at 1. Similarly, a unique degree 1 polynomial passes through $\{3, 9\}$ and $f_a(1)$ can be recovered by evaluating this polynomial at 1.

2.3 Parity-Check Matrix Approach to LRCs

A parity-check matrix approach was adopted to study bounds and constructions for LRCs in Hao and Xia (2016). We briefly present a proof of the Singleton-like upper bound on the minimum distance of LRCs with all-symbol locality Gopalan *et al.* (2012) through a parity-check matrix viewpoint given in Hao and Xia (2016). We will use a similar approach to construct LRCs with information locality and availability in Chapter 4.

Theorem 1. *For a (n, k, d, r) linear LRC with all-symbol locality,*

$$d \leq n - k - \left\lceil \frac{k}{r} \right\rceil + 2$$

Proof. Let H be the parity-check matrix for a (n, k, d, r) LRC. We characterize H as given in the following algorithm

Algorithm 1 Characterization of parity check matrix H of a (n, k, r) LRC

```
Set S_0 = φ, i = 1
while S_{i-1} ≠ [n] do
    Pick j ∈ [n] \ S_{i-1}
    Choose the minimal weight parity check equation h_i
    Set S_i = S_{i-1} ∪ supp(h_i)
    i = i + 1
Set l = i - 1, H_1 = [h_1; ... ; h_l]
Choose additional n - k - l rows from C^⊥ such that these rows and H_1 form a full
    rank n - k × n matrix
```

We know that for a (n, k, d) linear code, the parity-check matrix should satisfy the following properties :

1. Any $d - 1$ columns of the parity-check matrix should be linearly independent
2. There exists some d columns that are linearly dependent.

Therefore, it is sufficient to prove that H has $n - k - \left\lceil \frac{k}{r} \right\rceil + 2$ columns that are linearly dependent. Consider the first $t = \left\lfloor \frac{k-1}{r} \right\rfloor$ rows. Let γ be the number of columns in which the non-zero elements of the t rows lie in. This implies $\gamma \leq t\,(r + 1)$. Let $\eta = n - k - t$.

Delete the t rows and the γ columns. Therefore, we are left with a $\eta \times n - \gamma$ matrix denoted by H'. Note that

$$n - \gamma \geq n - t\,(r+1) \geq n - k - t + 1 = \eta + 1$$

Therefore out of the $n - \gamma$ columns, we can find $\eta + 1$ columns which are linearly dependent in H'. Note that, these columns are also linearly dependent in H since the remaining entries of the column in H and not in H' are 0. Thus, there exists $\eta + 1 = n - k - t + 1 = n - k - \left\lceil \frac{k}{r} \right\rceil + 2$ columns that are linearly dependent in H. $\square$

CHAPTER 3

Bounds on Codes with Irregular Recovery and Unequal Locality with Availability

3.1 Irregular Recovery with Availability

In this section, we consider locally recoverable codes (LRCs) whose coordinates have an irregular recovery profile. This extends the notion of (r, t) locality in Wang and Zhang (2014) to the case where sizes of recovering sets of each coordinate are not equal. A precise definition is as follows.

Definition 1. *Let $\mathcal{C} \subseteq \mathbb{F}_q^n$ be an $[n, k, d]$ code. The i-th coordinate has $(\mathbf{r}, t)$ locality, where $\mathbf{r} = (r_1, r_2 \ldots r_t)$, if there are t disjoint recovering sets $R_1^{(i)}, R_2^{(i)} \ldots R_t^{(i)}$ for i such that*

$$\left| R_j^{(i)} \right| \leq r_j \quad \forall j \in [t] \, .$$

The code $\mathcal{C}$ has $(\mathbf{r}, t)$ information locality if all information coordinates have $(\mathbf{r}, t)$ locality. The code $\mathcal{C}$ has $(\mathbf{r}, t)$ all-symbol locality if all coordinates have $(\mathbf{r}, t)$ locality.

3.1.1 Bounds for Information Locality

First we consider bounds on minimum distance of codes with $(\mathbf{r}, t)$ information locality. We follow a similar proof technique as Gopalan *et al.* (2012) Wang and Zhang (2014) but with adaptations for unequal recovery.

Theorem 2. *If $\mathcal{C}$ has $(\mathbf{r}, t)$ information locality, then*

$$d \leq n - k - \left\lceil \frac{t(k-1)+1}{\sum_{j=1}^{t} (r_j - 1) + 1} \right\rceil + 2$$

Proof. In the proof, we will assume that $r_1 \leq r_2 \leq \cdots \leq r_t$. Let l be the number of iterations in Algorithm 1 when run on the code $\mathcal{C}$ resulting in a subset S. Denote the rank increment and size increment in the i-th iteration of Algorithm 1 (refer Appendix A) by $m_i = \text{rank}(S_i) - \text{rank}(S_{i-1})$ and $s_i = |S_i| - |S_{i-1}|$, respectively. We find a lower bound for $|S|$, which in turn leads to an upper bound on the distance $d \leq n - |S|$. Consider two cases depending on how Algorithm 1 terminates.

Case 1: S_l is a union of $\Gamma_t(j)$'s i.e, S_l is formed in line 5.

Since each of the recovering sets contribute at least one linear dependency to S_i, we have $m_i \leq s_i - t$ for $i \in [l]$. Further, we have

$$|S| = \sum_{i=1}^{l} s_i \geq \sum_{i=1}^{l} (m_i + t) = k - 1 + tl. \tag{3.1}$$

To find a lower bound on $|S|$, we find a lower bound on l, the number of iterations. Since every recovering set adds at least 1 linear equation, we have

$$\text{rank}\left(\Gamma_a(j)\right) \leq 1 + \sum_{j'=1}^{a} (r_{j'} - 1).$$

Since S is the union of l sets $\Gamma_t(j)$, we have

$$\text{rank}(S) = k - 1 \leq l\big(1 + \sum_{j=1}^{t} (r_j - 1)\big). \tag{3.2}$$

Using (3.2) in (3.1) and $d \leq n - |S|$, we get

$$\begin{aligned} d &\leq n - \left(k - 1 + t\left\lceil \frac{k-1}{1 + \sum_{j=1}^{t} (r_j - 1)} \right\rceil\right) \\ &\leq n - k - t\left\lceil \frac{k-1}{1 + \sum_{j=1}^{t} (r_j - 1)} \right\rceil + 1 \qquad (3.3) \\ &\leq n - k - \left\lceil \frac{t(k-1)+1}{1 + \sum_{j=1}^{t} (r_j - 1)} \right\rceil + 2, \qquad (3.4) \end{aligned}$$

where, to get from (3.3) to (3.4), we use the facts $\lfloor tx \rfloor \leq tx \leq t\lceil x \rceil$ for a real number x and $\lfloor \frac{a}{b} \rfloor = \lceil \frac{a+1}{b} \rceil - 1$ for positive integers a, b.

Case 2: S_l is formed in line 9.

Since $m_i \le s_i - t$, $1 \le i \le l-1$, and $m_l \le s_l - a$, we have

$$\begin{aligned} |S| = \sum_{i=1}^{l} s_i &\ge \sum_{i=1}^{l-1} (m_i + t) + m_l + a \\ &= k - 1 + t\,(l-1) + a. \end{aligned} \tag{3.5}$$

Since rank $(S_{l-1} \cup \Gamma_{a+1}(l)) = k$, and S_{l-1} is the union of $l-1$ sets $\Gamma_t(j)$, we have

$$k \le (l-1)\left(1 + \sum_{j=1}^{t} (r_j - 1)\right) + \left(1 + \sum_{j=1}^{a+1} (r_j - 1)\right). \tag{3.6}$$

Using the lower bound for $l-1$ from (5.2) in (3.5), we get

$$|S| \ge k - 1 + t\frac{k - 1 - \sum_{j=1}^{a+1}(r_j - 1)}{1 + \sum_{j=1}^{t}(r_j - 1)} + a. \tag{3.7}$$

Let $\Omega = 1 + \sum_{j=1}^{t}(r_j - 1)$. Since the r_j are in increasing order, the average of the first $a+1$ of the $(r_j - 1)$ is smaller than the average of all t resulting in the inequality

$$\frac{\sum_{j=1}^{a+1}(r_j - 1)}{a+1} \le \frac{\Omega - 1}{t}. \tag{3.8}$$

Using (3.8) in (3.7) and simplifying, we get

$$|S| \ge k + \frac{t(k-1) + a + 1}{\Omega} - 2 \tag{3.9}$$

$$\ge k + \left\lceil \frac{t(k-1) + 1}{\Omega} \right\rceil - 2. \tag{3.10}$$

Using the above in $d \le n - |S|$, we get the statement of the theorem. $\square$

For the case of equal recovery with availability, $\mathbf{r}$ is a constant vector with $r_i = r$, and the bound of Theorem 1 reduces to (2.2).

3.1.2 Bounds for All-symbol Locality

We now derive a minimum distance upper bound for codes with $(\mathbf{r}, t)$ all-symbol locality. Our proof is similar in outline toTamo *et al.* (2016), and for equal recovery the bound reduces to (2.3). We present two lemmas required for the proof of the upper bound. For the rest of this section, we assume that $r_1 \leq r_2 \leq \cdots \leq r_t$. We use the notions of recovering graph and expansion ratio from Tamo *et al.* (2016).

Recovering Graph and Expansion Ratio

The recovering graph of a length-n LRC code with $(\mathbf{r}, t)$ locality has vertex set $[n]$ and edges of color j from i to i' if $i' \in R_j(i)$ for $j \in [t]$. More generally, a t-edge-colored directed graph is said to be an $(\mathbf{r}, t)$ recovering graph if, for $i \in [t]$, every vertex has at least one and at most r_i outgoing color-i edges. The set of vertices with incoming edges of the same color from a given vertex i is said to be a recovery set for i. Note that the recovering graph of an LRC code with $(\mathbf{r}, t)$ locality is indeed an $(\mathbf{r}, t)$ recovering graph with $R_j^{(i)}$ being the recovery set of vertex i.

Consider an arbitrary subset of vertices S in a recovering graph. Color all the vertices in S in some fixed color, say red. Color every vertex with at least one fully colored recovery set. Continue this procedure until no more vertices can be colored. The final set of colored vertices thus obtained is defined to be the *closure* of S, denoted by $\mathrm{Cl}(S)$. The *expansion ratio* with respect to S, denoted $e(S)$, is defined as the ratio $e(S) = |\mathrm{Cl}(S)|/|S|$. Observe that all vertices in $\mathrm{Cl}(S)$ can be recovered from vertices in S.

Lemma 1. *Let G be an $(\mathbf{r}, t)$ recovering graph. For a vertex $v \in G$, there exists a subset of the vertices S such that $v \in \mathrm{Cl}(S)$ and*

$$|S| \leq \prod_{i=1}^{t} r_i \text{ and } e(S) \geq e_t \triangleq 1 + \sum_{j=1}^{t} \frac{1}{\prod_{i=1}^{j} r_i}. \tag{3.11}$$

Proof. Our proof follows the proof of Lemma 3 in Tamo *et al.* (2016) closely, except we use the following key insight to construct S : smaller recovering sets result in larger expansion ratios. Therefore, while constructing S we give a higher preference to smaller recovering sets as compared to larger recovering sets. We proceed by induction on t. For $t = 0$, we take S as the single vertex v, and get $e(S) = 1$. Assuming the statement is true for $t = k$, we prove the statement for $t = k + 1$. Remove vertex v from G. For every other vertex, $u \neq v$, remove the edges that correspond to the recovering set with size r_1 unless there is an edge from u to v; in such a case, remove the edges in that recovering set and call the resulting graph G_1. We remove the r_1 sized recovering sets because of the aforementioned principle of giving higher preference to smaller sized recovering sets. Note that the remaining vertices in the graph have $t - 1$ recovering sets. It is easy to see that for a vertex $u \in G_1$ with recovery profile $\{\widetilde{r_1}, \widetilde{r_2} \cdots \widetilde{r_k}\}$ where $\widetilde{r_i} \leq r_{i+1}$. Let $v_1, v_2 \cdots v_l$ be the vertices in the r_{k+1} sized recovering set of v, where $l \leq r_{k+1}$. We omit the construction of the set S since it is the exact same as that of Tamo *et al.* (2016). We now derive the upper bound on the size of S

$$|S| \leq r_1 \prod_{i=1}^{k} \widetilde{r_i} \leq \prod_{i=1}^{k+1} r_i$$

As a result, $e\left(S\right)$ satisfies

$$e\left(S\right) \geq e_{k+1} \triangleq \frac{1}{\prod\limits_{i=1}^{k+1} r_i} + e_k$$

□

The increasing order $r_1 \leq \cdots \leq r_t$ ensures that the largest possible expansion ratio can be obtained using $\{r_1, r_2, \ldots, r_t\}$. Note that Lemma 1 reduces to Lemma 3 in Tamo *et al.* (2016) for the case of equal recovery when $r_i = r$.

The radix-r representation of an integer plays a role in Tamo *et al.* (2016). The analog for unequal recovery is the representation of an integer in the unequal radix

$$\{1, r_1, r_1 r_2, \ldots, r_1 r_2 \cdots r_t\}$$

. The next lemma concerns such representations.

Lemma 2. *Let m be an integer with the following representation*

$$m = \beta r_t \prod_{i=1}^{t} r_i + \sum_{i=1}^{t} \alpha_i \prod_{j=1}^{i} r_j + \alpha_0,$$

where $0 \leq \alpha_i < r_{i+1}$, $0 \leq i \leq t-1$, $0 \leq \alpha_t < r_t$, and β is an integer. Let $\tilde{e}_i = 1 + \sum_{j=1}^{i} \frac{1}{\prod_{l=1}^{j} r_l}$. Then,

$$\left\lfloor \frac{m}{\prod_{i=1}^{t} r_i} \right\rfloor e_t \prod_{i=1}^{t} r_i + \sum_{i=0}^{t-1} \alpha_i \tilde{e}_i \prod_{j=1}^{i} r_j = \sum_{i=0}^{t} \left\lfloor \frac{m}{\prod_{j=1}^{i} r_j} \right\rfloor$$

Proof. For $i \in [t]$, we have

$$\tilde{e}_i \prod_{j=1}^{i} r_j = \sum_{j=1}^{i+1} \prod_{l=j}^{i} r_l. \tag{3.12}$$

Next, observe that

$$\sum_{i=0}^{t} \left\lfloor \frac{m}{\prod_{j=1}^{i} r_j} \right\rfloor = \sum_{i=0}^{t} (\beta r_t + \alpha_t) \prod_{j=i+1}^{t} r_j + \sum_{j=i}^{t-1} \alpha_j \prod_{l=i+1}^{j} r_l.$$

Also, from (3.12) and the above representation of m,

$$\left\lfloor \frac{m}{\prod_{i=1}^{t} r_i} \right\rfloor e_t \prod_{i=1}^{t} r_i = (\beta r_t + \alpha_t) \sum_{i=0}^{t-1} \prod_{j=i+1}^{t} r_j.$$

Therefore, it suffices to prove that

$$\sum_{i=0}^{t-1} \alpha_i \sum_{j=1}^{i+1} \prod_{l=j}^{i} r_l = \sum_{i=0}^{t} \sum_{j=i}^{t-1} \alpha_j \prod_{l=i+1}^{j} r_l.$$

The above equation can be verified to be true by comparison of coefficients of α_i, $i \in$

$[t-1]$, thus completing the proof. □

Theorem 3. *If an* $[n, k, d]$ *code* $\mathcal{C}$ *has* $(\mathbf{r}, t)$ *all-symbol locality, then*

$$d \leq n - k + 1 - \sum_{i=1}^{t} \left\lfloor \frac{k-1}{\prod_{j=1}^{i} r_j} \right\rfloor .$$

Proof. The proof is similar to that of Tamo *et al.* (2016), and only key ideas are presented. We obtain a $k-1$ sized subset of vertices S and prove a lower bound on $|\mathrm{Cl}\,(S)\,|$ by applying Lemma 1 repeatedly. Consider the recovering graph G of $\mathcal{C}$. From Lemma 1, there exists a set of vertices S_0 such that its expansion ratio is atleast e_t. Let the induced subgraph on $V \setminus \mathrm{Cl}\,(S_0)$ be G_1, which is an $(\mathbf{r}, t)$ recovering graph. Apply Lemma 1 on G_1 and continue this process until the number of remaining vertices in the graph G_l after l steps is lesser than $\prod_{j=1}^{t} r_j$.

Now, continue by viewing the graph G_l as an $([r_1, \ldots, r_{t-1}], t-1)$ recovering graph. By Lemma 1, there exists a set of vertices S_l with $|S_l| \leq \prod_{j=1}^{t-1} r_j$ and expansion ratio at least $\tilde{e}_{t-1}$ (see Lemma 2 for definition). Continue the coloring process going through $([r_1, \ldots, r_i], i)$ recovering graphs containing sets of vertices of size at most $\prod_{j=1}^{i} r_j$ and expansion ratio at least $\tilde{e}_i$ for $i = t-2, \ldots, 1$ till $k-1$ vertices are colored.

By keeping track of the expansion ratios and the number of applications of Lemma 1, we get the following lower bound on $|\mathrm{Cl}\,(S)\,|$:

$$|\mathrm{Cl}\,(S)\,| \geq \left\lfloor \frac{k-1}{\prod\limits_{i=1}^{t} r_i} \right\rfloor e_t \prod_{i=1}^{t} r_i + \sum_{i=0}^{t-1} \alpha_i \tilde{e}_i \prod_{j=1}^{i} r_j$$

where $k-1 = \sum_i \left(\alpha_i \prod\limits_{j=1}^{i} r_j \right)$. Using Lemma 2,

$$|\mathrm{Cl}\,(S)\,| \geq \sum_{i=0}^{t} \left\lfloor \frac{k-1}{\prod\limits_{j=1}^{i} r_i} \right\rfloor .$$

Since rank$(\mathrm{Cl}\,(S)) = $ rank$(S) < k$, $d \leq n - |\mathrm{Cl}\,(S)\,|$, which results in the bound of the theorem. □

3.2 Unequal Locality with Availability

We now consider the case of unequal locality where different coordinates have possibly different, but regular recovery profiles with availability t. That is, the i-th coordinate has a length-t recovery profile of the form $[r_i \; r_i \; \cdots \; r_i]$. We will consider only information locality here. At the end of the following section, we will give a brief intuition as to why this particular case is ill-posed to study all-symbol locality.

3.2.1 Bounds for Information Locality

We first consider information locality and prove a lower bound on minimum distance. We use the notion of locality profile from Kadhe and Sprintson (2016).

Definition 2. *An $[n, k, d]$ code $\mathcal{C}$ has information locality profile $\{k_1, k_2, \dots, k_r\}$ with availability t if k_i is the number of information coordinates with locality i and availability t.*

A modified version of Algorithm 1, which we refer to as Algorithm 2, is used in the proof. Algorithm 2 is identical to Algorithm 1 except for Step 3, which becomes

$$3 : \text{Set } i = i + 1, \text{ Choose } j \in [n] \setminus S_{i-1} \text{ with minimal locality}$$

Theorem 4. *If $\mathcal{C}$ is an $[n, k, d]$ linear code with information locality profile $\{k_1, k_2, \dots, k_r\}$ with availability t, then*

$$d \leq n - k + 2 - t\left(\sum_{j=1}^{r-1} \left\lceil \frac{k_j}{t\,(j-1)+1} \right\rceil\right) - \left\lceil \frac{t\,(k_r - 1) + 1}{t\,(r-1)+1} \right\rceil .$$

Proof. We use Algorithm 2 with the code $\mathcal{C}$. Let l be the number of iterations of Algorithm 2. Consider two cases depending on how Algorithm 2 terminates.

Case 1: S_l is formed in line 5.

As in the proof of Theorem 2, we get

$$|S| \geq k - 1 + tl. \tag{3.13}$$

Let l_j be the number of iterations in which coordinates with locality j are chosen. In these l_j iterations, the rank of S increases by k_j for $0 \leq j \leq r-1$. Since rank $(S_l) = k-1$ and coordinates with least locality are preferred in Step 3, for $j = r$, the rank increment for the l_r iterations is $k_r - 1$. Now, as in the proof of Theorem 2,

$$\begin{aligned} k_j &\leq l_j \left(1 + t\left(j-1\right)\right) \quad \forall j \in \left[r-1\right], \\ k_r - 1 &\leq l_r \left(1 + t\left(r-1\right)\right). \end{aligned}$$

Since $l = \sum_{j=1}^{r} l_j$, we have

$$l \geq \sum_{j=1}^{r-1} \left\lceil \frac{k_j}{t\left(j-1\right)+1} \right\rceil + \left\lceil \frac{k_r - 1}{t\left(r-1\right)+1} \right\rceil. \tag{3.14}$$

Plugging (3.14) in (3.13), we get

$$\begin{aligned} |S| &\geq k - 1 + t\left(\sum_{j=1}^{r-1} \left\lceil \frac{k_j}{t\left(j-1\right)+1} \right\rceil + \left\lceil \frac{k_r - 1}{t\left(r-1\right)+1} \right\rceil\right) \\ &\geq k - 2 + t\left(\sum_{j=1}^{r-1} \left\lceil \frac{k_j}{t\left(j-1\right)+1} \right\rceil\right) + \left\lceil \frac{t\left(k_r - 1\right)+1}{t\left(r-1\right)+1} \right\rceil, \end{aligned}$$

where the manipulations for the last step are same as before. Using $d \leq n - |S|$, the distance bound follows.

Case 2 S_l is formed in line 9.

Note that (3.5) holds in this case. Since rank $\left(S_{l-1} \cup \Gamma_{a+1}\left(c_l\right)\right) = k$, in the last l_r iterations, the rank increment is now k_r instead of $k_r - 1$ in Case 1 above. Lower bounds on l_j, $1 \leq j \leq r-1$ are the same as in Case 1. For l_r, we get a lower bound from the following inequality

$$k_r \leq (l_r - 1)(1 + t(r-1)) + 1 + (a+1)(r-1).$$

Adding the lower bounds for l_j,

$$l - 1 \geq \sum_{j=1}^{r-1} \left\lceil \frac{k_j}{t(j-1)+1} \right\rceil + \left\lceil \frac{k_r - ar - r + a}{t(r-1)+1} \right\rceil. \tag{3.15}$$

Plugging (3.15) in (3.5), we have

$$\begin{aligned} |S| \geq k - 1 + a + \\ t\left(\sum_{j=1}^{r-1} \left\lceil \frac{k_j}{t(j-1)+1} \right\rceil + \left\lceil \frac{k_r - ar - r + a}{t(r-1)+1} \right\rceil \right). \end{aligned} \tag{3.16}$$

Let $\Omega = 1 + i(r-1)$. Using $t\lceil x \rceil \geq \lceil tx \rceil$,

$$t\left\lceil \frac{k_r - \Omega_{a+1}}{\Omega_t} \right\rceil \geq \left\lceil \frac{t(k_r - \Omega_{a+1})}{\Omega_t} \right\rceil \tag{3.17}$$

$$= \left\lceil \frac{t(k_r-1) + 1 - t(\Omega_{a+1} - 1) - 1}{\Omega_t} \right\rceil$$

$$= \left\lceil \frac{t(k_r-1)+1}{\Omega_t} - (a+1) + \frac{a}{\Omega_t} \right\rceil. \tag{3.18}$$

Substituting (3.18) in (3.16), and using $d \leq n - |S|$, we get the desired bound.

□

CHAPTER 4

Optimal Constructions for LRCs with Availability

In this chapter, we present constructions of LRCs with availability and compare their minimum distances with the derived distance upper bounds. In some cases, we obtain optimal constructions where the minimum distance meets the upper bound.

4.1 Regular Recovery and Locality with Availability

In this section, we revert to the notion of equal recovery and locality, and consider LRCs having (r, t) locality with availability.

4.1.1 All-symbol Locality

The upper bound on minimum distance for LRCs with availability t and all-symbol locality r is given by (2.3). The tightness of this bound for arbitrary t has not been fully settled. We provide a generalization of the polynomial-evaluation construction of LRC codes in Example 6 of Tamo and Barg (2014) and show optimality for some specific cases by computational methods. Later, we prove optimality for the case of $r = k - 1$ and arbitrary t.

For the sake of clarity and completeness, we briefly outline Example 6 of Tamo and Barg (2014) below. Although this example is similar to that of the example in Chapter 2, we illustrate the use of additive subgroups and make an important observation that helps prove optimality of this construction.

Example 1. An $(n = 16, k, r = 3, t = 2)$ LRC is constructed over $\mathbb{F}_{16}$ by generating orthogonal partitions from cosets of two copies of $\mathbb{F}_4^+$ denoted $H_1 = \{0, 1, \alpha, \alpha^4\}$, $H_2 =$

$\{0, \alpha^2, \alpha^3, \alpha^6\}$, where α is the residue class of x modulo $x^4 + x + 1$. The annihilator polynomials of H_1 and H_2, denoted g_1 and g_2, respectively, are

$$g_1(x) = x^4 + \alpha^{10}x^2 + \alpha^5 x,$$
$$g_2(x) = x^4 + \alpha^{14}x^2 + \alpha^{11} x.$$

The orthogonal partitions that are generated by H_1, H_2 and their cosets are

$$\mathcal{A}_1 = \{\{0, 1, \alpha, \alpha^4\}, \{\alpha^2, \alpha^8, \alpha^5, \alpha^{10}\}, \{\alpha^3, \alpha^{14}, \alpha^9, \alpha^7\}, \{\alpha^6, \alpha^{13}, \alpha^{11}, \alpha^{12}\}\},$$
$$\mathcal{A}_2 = \{\{0, \alpha^2, \alpha^3, \alpha^6\}, \{1, \alpha^8, \alpha^{14}, \alpha^{13}\}, \{\alpha, \alpha^5, \alpha^9, \alpha^{11}\}, \{\alpha^4, \alpha^{10}, \alpha^7, \alpha^{12}\}\}.$$

The basis of $\mathcal{F}^3_{\mathcal{A}_1} \bigcap \mathcal{F}^3_{\mathcal{A}_2}$ is obtained by choosing polynomials of distinct degrees that can be expressed as a linear combination of the basis of both $\mathcal{F}^3_{\mathcal{A}_1}$ and $\mathcal{F}^3_{\mathcal{A}_2}$. We find that the basis of $\mathcal{F}^3_{\mathcal{A}_1} \bigcap \mathcal{F}^3_{\mathcal{A}_2}$ comprises of polynomials of degrees $0, 1, 2, 4, 6, 8, 9, 10, 12$. Table 4.1 summarizes the possible dimensions along with two distance lower bounds: (1) $n - \max_{f_a \in V_m} \deg(f_a)$, (2) $n - \max_{f_a \in V_m} \deg(\gcd(f_a, x^{16} - x))$. The second lower bound is evaluated computationally. The distance upper bound in (2.3) is also shown. The second lower bound is tighter than the first for $k = 6, 7$. For $k = 7$, the second lower bound meets the upper bound, thus giving an optimal code. For $k = 8, 9$, our computations did not terminate.

Table 4.1: Lower and upper bound on minimum distance for Example 1.

k	LB 1	LB 2	UB
4	12	12	12
5	10	10	11
6	8	9	10
7	7	8	8
8	6	–	7
9	4	–	6

We generalize the construction in the above example to arbitrary t and show that it is optimal for $k = r + 1$.

Construction 1. Let $r + 1 = p^l$, p: prime, $l \geq 1$, $t \geq 2$. Let $n = (r+1)^t, k = r + 1$ and $A = \mathbb{F}_{(r+1)^t}$. The additive subgroup of $\mathbb{F}_{(r+1)^t}$ can be written as

$$\mathbb{F}^+_{(r+1)^t} \cong \{[a_1, \ldots, a_t] : a_i \in \mathbb{F}^+_{(r+1)}\}.$$

Consider t subgroups $H_i = \{[0, \ldots, 0, a_i, 0, \ldots, 0] : a_i \in \mathbb{F}^+_{(r+1)}\}$ of $\mathbb{F}^+_{(r+1)^t}$ of size $r + 1$ for $i \in [t]$. Let g_i be the annihilator polynomial of H_i. Let $\mathcal{A}_i$ be the partitions of A induced by the cosets of H_i. We have

$$\mathbb{F}_{\mathcal{A}_i}[x] = \langle 1, g_i(x), g_i(x)^2 \ldots g_i(x)^{\frac{n}{r+1}-1} \rangle.$$

Since $\bigcap_{i=1}^{t} H_i = \{0\}$, $\{\mathcal{A}_i\}$ are orthogonal partitions.

Now, a crucial observation is the following. Since H_i is a copy of $\mathbb{F}^+_{r+1}$, we have $\sum_{h \in H_i} h = 0$. It follows that the coefficient of x^r in $g_i(x) = \prod_{h \in H_i}(x - h)$ is 0 $\forall i \in [t]$. Therefore, $g_i(x)$ is of degree $r + 1$ and is contained in $\bigcap_{i=1}^{t} \mathcal{F}^r_{\mathcal{A}_i}$ as degrees 1 to $r - 1$ are contained in each $\mathcal{F}^r_{\mathcal{A}_i}$. Using this, we see that

$$V_{r+1} = \bigcap_{i=1}^{t} \mathcal{F}^r_{\mathcal{A}_i} \bigcap P_{r+1} = \langle 1, x \ldots x^{r-1}, g_1(x) \rangle.$$

To encode a message $a \in \mathbb{F}^{r+1}_{(r+1)^t}$, we define the encoding polynomial

$$f_a(x) = \sum_{i=0}^{r-1} a_i x^i + a_{r+1} g_1(x).$$

The code is obtained by evaluating f_a on the points of A.

Theorem 5. *The* $((r+1)^t, r+1, r, t)$ *LRC code from Construction 1 is optimal.*

Proof. Since $r + 1$ is the maximum degree of the encoding polynomials, we have $d \geq$

$n-(r+1)$. By the bound (**??**),

$$d \leq n - \left(k - 1 + \sum_{i=1}^{t} \left\lfloor \frac{k-1}{r^i} \right\rfloor \right) = n - (r+1),$$

and the proof is complete. □

For $r < k-1$, a basis is chosen in the same way as mentioned in the above example. More formally, to obtain the basis of the intersection we define a matrix M_l corresponding to the l^{th} recovering set, $l \in [t]$ where $M_l(i,j)$ is the coefficient of x^i of the j^{th} basis (ordered in increasing order of their degrees). We say that M_l is 'truncated to m' if the degree of the last basis is atmost m. Consequently, the rows will also be truncated till degree m. Let $M(m)$ denote the matrix obtained by augmenting M_l, $l \in [t]$, such that each M_l is truncated to m. It is not hard to see that the nullspace of $M(m)$ is isomorphic to $\bigcap_{i=1}^{t} \mathcal{F}_{\mathcal{A}_i}^{r} \bigcap P_m$, thus providing a method to obtain the dimension of the code by getting the rank of $M(m)$. The distance lower bound can be obtained by identifying the maximum degree of the polynomials in the nullspace of $M(m)$.

4.1.2 Parity-check Matrix Construction for Information Locality

The upper bound on minimum distance for LRCs with information locality r and availability t is given by (2.2). We present an explicit parity-check matrix construction, which is an extension of Hao and Xia (2016), to achieve this bound for $n \geq k(tr+1)$. Let

$$\Gamma = n - k + 1 - \left\lceil \frac{t(k-1)+1}{t(r-1)+1} \right\rceil .$$

Construction 2. Let $tr+1 | n$ and $tr+1 \nmid \Gamma$. Define $v = \frac{n}{tr+1}$, $u = n-k-vt$, $\alpha_{i,j,h} \in \mathbb{F}_{q^m}, m \geq n\frac{t(r-1)+1}{tr+1}$ such that $\left\{ \alpha_{i,1,0} - \sum_{l=1}^{t} \alpha_{i,l,r}, \alpha_{i,j,h} - \alpha_{i,j,r} \mid i \in [v], j \in [t], h \in [r-1] \right\}$ are linearly independent over $\mathbb{F}_q$, then the following parity-check matrix, H defines a q^m-

ary (n, k, r, t) LRC with information locality and availability.

$$\left[\begin{array}{ccccccc} \multicolumn{7}{c}{I_v \otimes H_1} \\ \hline \alpha_{1,1,0} & \alpha_{1,1,1} & \dots & \alpha_{1,1,r} & \alpha_{1,2,1} & \dots & \alpha_{v,t,r} \\ \alpha_{1,1,0}^{q} & \alpha_{1,1,1}^{q} & \dots & \alpha_{1,1,r}^{q} & \alpha_{1,2,1}^{q} & \dots & \alpha_{v,t,r}^{q} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{1,1,0}^{q^u} & \alpha_{1,1,1}^{q^u} & \dots & \alpha_{1,1,r}^{q^u} & \alpha_{1,2,1}^{q^u} & \dots & \alpha_{v,t,r}^{q^u} \end{array}\right]$$

where

$$H_1 = \left[\begin{array}{c|c} 1 & \\ 1 & \\ \dots & I_t \otimes \underbrace{(11\dots1)}_{\text{r}} \\ 1 & \end{array}\right]$$

Theorem 6. *For $n \geq k\,(tr + 1)$, the linear code obtained using Construction 2 meets the distance upper bound* (2.2)

Proof. Choose Γ columns arbitrarily from H. We refer to the columns beginning with all ones as 'availability columns' and the rest as 'non-availability columns'. Let x be the number of availability columns chosen. Note that $x \leq \dfrac{n}{tr+1}$. We claim that the number of non-zero rows among the chosen Γ columns is greater than Γ. Each availability column ensures that t distinct rows are non-zero. Since each of these availability column cover disjoint rows we obtain the following lower bound to the number of non-zero rows among Γ arbitrarily chosen columns

$$n - k - \frac{nt}{tr+1} + xt + \left\lceil \frac{\Gamma - x - xtr}{r} \right\rceil \tag{4.1}$$

We consider two cases

Case 1 $\left\lceil \dfrac{\Gamma}{tr+1} \right\rceil \leq x \leq \dfrac{n}{tr+1}$

Therefore, (4.1) is reduced to $n - k - \frac{nt}{tr+1} + xt$. It is sufficient to prove that

$$n - k - \frac{nt}{tr+1} + t\left\lceil \frac{\Gamma}{tr+1} \right\rceil \geq \Gamma \tag{4.2}$$

We first show that

$$n-k-\frac{nt}{tr+1}+\frac{\Gamma t}{tr+1}+1>\Gamma \tag{4.3}$$
$$\iff (n-\Gamma)\left(\frac{t\left(r-1\right)+1}{tr+1}\right)>k-1$$
$$\iff n-\Gamma>k-1+\frac{t\left(k-1\right)}{t\left(r-1\right)+1}$$
$$\iff k-1+\left\lceil\frac{t\left(k-1\right)+1}{t\left(r-1\right)+1}\right\rceil>k-1+\frac{t\left(k-1\right)}{t\left(r-1\right)+1}$$

Since, the last inequality is obviously true, (4.3) holds. We now show that

$$n-k-\frac{nt}{tr+1}+\frac{\Gamma t}{tr+1}<\Gamma \tag{4.4}$$
$$\iff (n-\Gamma)\left(\frac{t\left(r-1\right)+1}{tr+1}\right)<k$$
$$\iff n-\Gamma<k+\frac{kt}{t\left(r-1\right)+1}$$
$$\iff k-1+\left\lceil\frac{t\left(k-1\right)+1}{t\left(r-1\right)+1}\right\rceil<k+\frac{kt}{t\left(r-1\right)+1}$$

Since, the last inequality is obviously true, (4.4) holds. Therefore, if $t\left\lceil\frac{\Gamma}{tr+1}\right\rceil\geq\frac{\Gamma t}{tr+1}$, then (4.2) holds. Otherwise, from (4.3) and (4.4), Γ is the integer lying between them implying (4.2) becomes an equality.

Case 2 $0\leq x\leq\left\lfloor\frac{\Gamma}{tr+1}\right\rfloor$

(4.1) is reduced to $n-k-\frac{nt}{tr+1}+\left\lceil\frac{\Gamma-x}{r}\right\rceil$. Since, this is a decreasing function of x, it is sufficient to prove

$$n-k-\frac{nt}{tr+1}+\left\lceil\frac{\Gamma-\left\lfloor\frac{\Gamma}{tr+1}\right\rfloor}{r}\right\rceil\geq\Gamma \tag{4.5}$$

The LHS of (4.5) is lower bounded by $n-k-\frac{nt}{tr+1}+\left\lceil\frac{\Gamma t}{tr+1}\right\rceil$ which is equal to Γ from (4.4) and (4.3). Thus (4.5) follows. Since, the rest of the proof is similar to Hao and Xia (2016), we provide a brief outline. From the above, we have established that among arbitrarily chosen Γ columns, there are more than Γ rows. This selection can be reduced to a square matrix S, by deleting all all-zero rows and remaining rows beginning from the bottom. We perform the following column transformation. For all non-availability columns in S, we choose a column from one of the repair groups and subtract it from other columns in its repair group. It is easy to see that after this column transformation, for each of the locality rows, there is exactly one entry with a 1 corresponding to either a chosen non-availability column or an availability column. Also, note that an availability column has all zeros in the locality row if each of its repair groups were present among the Γ arbitrarily chosen columns. Obtaining the determinant by expanding along rows with entry 1, we obtain a matrix such that it's first row comprises of the following elements $\left\{\alpha_{i,1,0} - \sum_{l=1}^{t} \alpha_{i,l,r_l}, \alpha_{i,j,h} - \alpha_{i,j,r_j} \mid h \in [r_j - 1]\right\}$, where i and j depend on the arbitrary chosen Γ columns. By our construction and Lemma 4 of Hao and Xia (2016) this first row is linearly independent over $\mathbb{F}_q$. As a result, the determinant of this reduced matrix does not vanish implying the originally chosen arbitrary Γ columns are linearly independent. □

Note that, there exist m such that

$$n - k \geq \frac{ntr}{tr+1} \geq m \geq n\frac{t(r-1)+1}{tr+1}.$$

Therefore, $q^{n-k} > q^m$. For small enough values of q , q^{n-k} can be made smaller than $\binom{n}{k+\mu}$, which results in a smaller field size when compared to Wang and Zhang (2014). In addition, Construction 2 gives an explicit construction unlike Wang and Zhang (2014). Finally, we remark that Construction 2 can easily be extended to construct codes with irregular recovery and availability and codes with unequal information locality and availability as defined in Chapter 4

4.2 Irregular Recovery with Availability

We consider two examples of LRCs having irregular recovery with availability. The first example is the same as Example 5 of Tamo and Barg (2014). The second example is a similar construction applied to a larger field. We compare their distances with the upper bound of Theorem 3. We show the second example is optimal by showing that the distance satisfies the upper bound of Theorem 3 with equality.

Example 2. An $(n=12, k=4, r_1=3, r_2=2, t=2)$ LRC is constructed over $\mathbb{F}_{13}$ by generating orthogonal partitions, $\mathcal{A}$ and $\mathcal{A}'$, from the cosets of the multiplicative subgroups generated by 5 and 3, respectively. The partitions are as follows:

$$\mathcal{A} = \{\{1,5,12,8\}, \{2,10,11,3\} \{4,7,9,6\}\},$$

$$\mathcal{A}' = \{\{1,3,9\}, \{2,6,5\}, \{4,12,10\}, \{7,8,11\}\}.$$

Since the constant polynomials with respect to $\mathcal{A}$ and $\mathcal{A}'$ are x^4 and x^3, respectively, (see Section **??**), we have

$$\mathbb{F}_{\mathcal{A}}[x] = \langle 1, x^4, x^8 \rangle, \quad \mathbb{F}_{\mathcal{A}'}[x] = \langle 1, x^3, x^6, x^9 \rangle.$$

Finding the basis of $\mathcal{F}_{\mathcal{A}}^3 \bigcap \mathcal{F}_{\mathcal{A}'}^2$ and truncating it to obtain a $k=4$ dimensional subspace, we have

$$V_6 = \langle 1, x, x^4, x^6 \rangle.$$

Since $\max\limits_{f_a \in V_6} \deg(f_a)$ is 6, this code has distance $d \geq 6$. Evaluating the upper bound on distance from Theorem (3), we have $d \leq 8$. Therefore, this construction using multiplicative subgroups is not likely to be optimal for LRCs having irregular recovery with availability.

Example 3. An $(n=32, k=8, r_1=7, r_2=3, t=2)$ LRC is constructed over $\mathbb{F}_{32}$ by generating orthogonal partitions, $\mathcal{A}$ and $\mathcal{A}'$, from cosets of copies of $\mathbb{F}_8^+$ and $\mathbb{F}_4^+$ denoted H and H', respectively. Let $\alpha \in \mathbb{F}_{32}$ be primitive satisfying $\alpha^5 + \alpha^2 + 1 = 0$. We have

$$H = \left(0, 1, \alpha, \alpha^2, \alpha^5, \alpha^{11}, \alpha^{18}, \alpha^{19}\right), \quad H' = \left(0, \alpha^3, \alpha^4, \alpha^{21}\right).$$

$\mathcal{F}_{\mathcal{A}}^7$, $\mathcal{F}_{\mathcal{A}'}^3$ and annihilator polynomials of H, H' are obtained as defined in Chapter 2. By linear algebraic techniques, we can find the dimension and basis of $V_m = \mathcal{F}_{\mathcal{A}}^7 \bigcap \mathcal{F}_{\mathcal{A}'}^3 \cup P_m$ numerically. It can be verified numerically that for $m=9$, we have the dimension $k=8$. Thus, evaluating the distance upper bound of Theorem (3), we get $d \leq 23$. Since $\max\limits_{f_a \in V_9} \deg(f_a)$ is 9, we also $d \geq 23$. Thus, the bound of Theorem (3) is met with equality.

CHAPTER 5

Miscellanous Results using the Parity-Check Matrix approach

In this chapter, we present two results that are disconnected from the overall theme of this thesis. The first result is a parity-check matrix proof of the distance upper bound on (r, δ) codes. Prakash *et al.* (2012) proved this upper bound using generalized Hamming weights. Our proof is simpler than the latter. The second result is a negative result that pertains to the rate upper bound for LRCs with availability derived in Tamo *et al.* (2016). We prove that this upper bound is unattainable by linear codes.

5.1 Simpler Proof of Distance Upper Bound for (r, δ) codes

We recall that (r, δ) codes are codes wherein each coordinate is protected by a local code of length $r + \delta - 1$ with distance atleast δ. We make the following assumption first introduced by Tamo and Barg (2014) that this local code is an MDS code.

The distance upper bound for an $[n, k, d]$, (r, δ) code is given by the following expression

$$d \leq n - k - (\delta - 1) \left\lfloor \frac{k-1}{r} \right\rfloor + 1$$

Proof. We modify the characterization of the parity-check matrix from Algorithm (2) as follows; instead of choosing the minimum weight parity-check equation for a coordinate, we choose a $(\delta - 1) \times (r + \delta - 1)$ block representing the parity-check matrix of the local $(r, r + \delta - 1)$ MDS code it is a part of and run the rest of the algorithm as it is except H_1 is the block diagonal matrix of the parity-check matrices of the local $(r, r + \delta - 1)$ MDS codes. It is easy to see that there are $\left\lfloor \frac{n}{r+\delta-1} \right\rfloor$ such blocks.

For the above modification to be well-defined, we have to ensure that

$$\left\lfloor \frac{n}{r+\delta-1} \right\rfloor (\delta-1) > n-k$$

We prove this by giving a simple argument that

$$\frac{n}{r+\delta-1}(\delta-1) < n-k \tag{5.1}$$

Since r coordinates from the local MDS codes are sufficient to recover the entire $r+\delta-1$ sized codeword, for $\left\lfloor \frac{n}{r+\delta-1} \right\rfloor$ such local codes, $r\left\lfloor \frac{n}{r+\delta-1} \right\rfloor$ coordinates are sufficient to recover the entire codeword. Therefore,

$$k < r\left\lfloor \frac{n}{r+\delta-1} \right\rfloor \leq \frac{nr}{r+\delta-1}$$

Rearranging the terms gives (5.1)

We now proceed with the proof by picking $t = \left\lfloor \frac{k-1}{r} \right\rfloor$ local parity-check matrices. The number of rows picked are $(\delta-1)t$. Also, the number of non-zero columns are atmost $(r+\delta-1)t$. Using the same reasoning that we used in Section 2.3, we can find $n-k-\delta-1)t+1$ columns that are linearly dependent. Therefore,

$$d \leq n-k-(\delta-1)\left\lfloor \frac{k-1}{r} \right\rfloor + 1$$

□

5.2 Linear Codes cannot achieve Tamo-Barg Rate Upper Bound

Tamo *et al.* (2016) derived the following upper bound on the rate of a LRC with all-symbol locality and availability.

The rate of a (n, k, r, t) LRC with all-symbol locality and availability satisfies

$$\frac{k}{n} \leq \frac{1}{\prod_{j=1}^{t}(1+\frac{1}{jr})} \tag{5.2}$$

Lemma 3. *For all $t > 1$ and $r \geq 2$,*

$$\prod_{j=1}^{t}(1+\frac{1}{jr}) < 1 + \frac{t}{r}$$

Proof. We prove this by induction on t. First for $t = 2$

$$\begin{aligned}(1+\frac{1}{r})(1+\frac{1}{2r}) &= 1 + \frac{1}{r} + \frac{1}{2r} + \frac{1}{2r^2} \\ &< 1 + \frac{2}{r}\end{aligned}$$

as $r \geq 2$

Suppose the inequality holds for $l - 1$. Let

$$\Gamma = \prod_{j=1}^{l-1}(1+\frac{1}{jr})$$

Then,

$$\begin{aligned}\Gamma(1+\frac{1}{lr}) &< 1 + \frac{l-1}{r} + \frac{\Gamma}{lr} \\ &< 1 + \frac{l}{r}\end{aligned}$$

where the last inequality follows from $l > 1 + \frac{l-1}{r} > \Gamma$ □

Theorem 7. *A linear (n, k, r, t) LRC with all-symbol locality and availability satisfying (5.2) does not exist*

Proof. Let C be a linear (n, k, r, t) LRC with all-symbol locality and availability satisfying (5.2). According to the recovering graph terminology in the Tamo-Barg paper, this implies that there exists a subset U of $[n]$ such that

$$|U| = n(1 - \frac{1}{\prod_{j=1}^{t}(1+\frac{1}{jr})}) = n - k \tag{5.3}$$

such that all the coordinates in U have atleast one recovering set completely in $\bar{U}$, where $\bar{U}$ is the complement of U in $[n]$. Equivalently, for each parity symbol, there exists atleast one recovering set of the parity symbol that comprises only of information symbols. Since, this is a linear code, such recovering sets of each parity symbol form $n-k$ linearly independent parity checks which suffice to define a parity check matrix which is of the following form.

$$H = [H'|I_{n-k}]$$

We make a couple of observations about H'

- Each row of H' contains atmost r non-zero entries.
- Each column of H' contains atleast t non-zero entries.

To justify the second observation, assume to the contrary that there exists atleast one column corresponding to, say c, with less than t non-zero entries. Note that any recovery set can be expressed as a linear combination of the $n-k$ parity checks. The linear combination of a recovery set of c not covered by the parity checks should consist of some parity check that covers c and a distinct parity symbol p. Note that the resulting linear combination consists of both c and p thus violating the disjoint recovering sets assumption.

By the aforementioned two observations,

$$(n-k)r \geq kt \Rightarrow \frac{k}{n} \leq \frac{1}{1+\frac{t}{r}}$$

But by our assumption,

$$\frac{k}{n} = \frac{1}{\prod_{j=1}^{t}(1+\frac{1}{jr})}$$

This contradicts Lemma 3. □

CHAPTER 6

Conclusions and Scope for Future Work

We derived new upper bounds on minimum distance for codes with unequal all-symbol locality and availability. We presented a generalization of a construction for LRCs with availability, that attains the upper bound on minimum distance for arbitrary t and $r = k-1$. An explicit parity-check matrix construction that meets the upper bound on minimum distance for LRCs with information locality and availability was also obtained for $n \geq k\,(tr+1)$. Future work includes finding optimal constructions for LRCs with availability for higher values of k (or lower values of r) and finding constructions that meet the bounds proposed in this paper for a larger range of parameter values.

APPENDIX A

Distance Upper Bound Algorithm

For $S \subseteq [n]$, let $\mathcal{C}_S$ denote the code $\mathcal{C}$ restricted to the positions in S, and let rank(S) denote the dimension of $\mathcal{C}_S$. A useful bound on minimum distance of $\mathcal{C}$ is the following: if rank$(S) < k$, then $d \leq n - |S|$. For LRCs, Algorithm 2 is typically used in proofs of minimum distance bounds to find a set S for which rank$(S) < k$ Gopalan *et al.* (2012)Wang and Zhang (2014).

Algorithm 2 Construct S such that rank $(S) = k - 1$

```
Set S_0 = φ, i = 0
while rank (S_i) ≤ k − 2 do
    Set i = i + 1, Choose j ∈ [n] \ S_{i−1}
    if rank (S_{i−1} ∪ Γ_t (j)) < k then
        Set S_i = S_{i−1} ∪ Γ_t (j)
    else
        Choose a s. t. rank (S_{i−1} ∪ Γ_{a+1} (j)) = k and
        R ⊆ R^{(j)}_{a+1} s. t. rank (S_{i−1} ∪ Γ_a (j) ∪ R) = k − 1
        Set S_i = S_{i−1} ∪ Γ_a (j) ∪ R
Return S = S_i
```

LIST OF PAPERS BASED ON THESIS

1. Sourbh Bhadane, Andrew Thangaraj "Unequal Locality and Recovery for Locally Recoverable Codes with Availability", *Twenty third National Conference on Communications (NCC), 2017*.

REFERENCES

1. **Cadambe, V. R.** and **A. Mazumdar** (2015). Bounds on the size of locally recoverable codes. *IEEE Transactions on Information Theory*, **61**(11), 5787–5794.

2. **Chen, B.**, **S. Xia**, and **J. Hao** (2017). Locally repairable codes with multiple (r_i, δ_i)-localities. *CoRR*, **abs/1702.05741**. URL `http://arxiv.org/abs/1702.05741`.

3. **Gopalan, P.**, **C. Huang**, **H. Simitci**, and **S. Yekhanin** (2012). On the locality of codeword symbols. *IEEE Transactions on Information Theory*, **58**(11), 6925–6934.

4. **Hao, J.** and **S. Xia** (2016). Bounds and constructions of locally repairable codes: Parity-check matrix approach. *CoRR*, **abs/1601.05595**. URL `http://arxiv.org/abs/1601.05595`.

5. **Huang, P.**, **E. Yaakobi**, **H. Uchikawa**, and **P. H. Siegel**, Linear locally repairable codes with availability. *In 2015 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2015.

6. **Kadhe, S.** and **A. Sprintson**, Codes with unequal locality. *In 2016 IEEE International Symposium on Information Theory (ISIT)*. 2016.

7. **Kim, G.** and **J. Lee** (2017). Locally repairable codes with unequal locality requirements. *CoRR*, **abs/1701.07340**. URL `http://arxiv.org/abs/1701.07340`.

8. **Pamies-Juarez, L.**, **H. D. L. Hollmann**, and **F. Oggier**, Locally repairable codes with multiple repair alternatives. *In Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. 2013. ISSN 2157-8095.

9. **Prakash, N.**, **G. M. Kamath**, **V. Lalitha**, and **P. V. Kumar**, Optimal linear codes with a local-error-correction property. *In Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012.

10. **Rawat, A. S.**, **D. S. Papailiopoulos**, **A. G. Dimakis**, and **S. Vishwanath**, Locality and availability in distributed storage. *In 2014 IEEE International Symposium on Information Theory*. IEEE, 2014.

11. **Silberstein, N.**, **A. S. Rawat**, **O. O. Koyluoglu**, and **S. Vishwanath**, Optimal locally repairable codes via rank-metric codes. *In Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. 2013. ISSN 2157-8095.

12. **Song, W.**, **K. Cai**, and **C. Yuen** (2016). On sequential locally repairable codes. *CoRR*, **abs/1610.09767**. URL `http://arxiv.org/abs/1610.09767`.

13. **Tamo, I.** and **A. Barg** (2014). A family of optimal locally recoverable codes. *IEEE Transactions on Information Theory*, **60**(8), 4661–4676.

14. **Tamo, I.**, **A. Barg**, and **A. Frolov** (2016). Bounds on the parameters of locally recoverable codes. *IEEE Transactions on Information Theory*, **62**(6), 3070–3083.

15. **Tamo, I.**, **D. S. Papailiopoulos**, and **A. G. Dimakis**, Optimal locally repairable codes and connections to matroid theory. *In Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. 2013. ISSN 2157-8095.

16. **Wang, A.** and **Z. Zhang** (2014). Repair locality with multiple erasure tolerance. *IEEE Transactions on Information Theory*, **60**(11), 6979–6987.

17. **Zeh, A.** and **E. Yaakobi**, Bounds and constructions of codes with multiple localities. *In 2016 IEEE International Symposium on Information Theory (ISIT)*. 2016.