Technische Universität München

Fakultät für Elektrotechnik und Informationstechnik

Data Integrity and Privacy in Distributed Storage

Lukas Holzbaur

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

genehmigten Dissertation.

Vorsitzender:		Prof. Dr. Sebastian Steinhorst
Prüfende der Dissertation:	1.	Prof. Dr.-Ing. Antonia Wachter-Zeh
	2.	Prof. Salim El Rouayheb, Ph.D.

Die Dissertation wurde am 14.06.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 21.09.2021 angenommen.

Abstract

The increase in data stored by cloud-based storage systems has led to a high demand for efficient solutions for preserving its integrity and the users' privacy. This work investigates different concepts related to these problems, starting from codes with locality properties. New results on the set of erasure patterns correctable by the class of codes for grid-like topologies are obtained, along with a generic method for adding global redundancy symbols. In an effort to facilitate efficient repair in terms of both locality and required bandwidth, novel constructions of regenerating partial MDS codes are introduced. Then, a new bounded distance decoder for the class of lifted affine-invariant codes is introduced. Further, it is shown that this class can correct almost all error patterns in the high-error regime.

The second part of this work analyzes the application of interleaving, a powerful method for increasing the error decoding radius, to the popular class of alternant codes. New upper and lower bounds on the probability of successfully decoding this class of codes with the decoding algorithm by Schmidt et al. are derived, thereby making it the only decoder for interleaved alternant codes for which such a theoretical analysis is known.

Finally, new bounds on the rate of private information retrieval in the coded storage setting are derived. The concepts of full support-rank and strongly linear PIR are introduced and the respective capacities proved.

Acknowledgments

This dissertation is based on the work conducted during my time at the *Institute for Communications Engineering* of the *Technical University of Munich* (TUM) in the group for *Coding and Cryptography* led by Antonia Wachter-Zeh. I want to take this opportunity to thank the many people that were part of the great journey I was allowed to undertake these past years.

First and foremost, I would like to thank Antonia Wachter-Zeh for giving me this opportunity, supporting my scientific and personal development, and providing a very pleasant working environment. The many valuable discussion with Antonia contributed significantly to the quality and presentation of the results in this work and I greatly appreciate the freedoms I had scientifically, methodologically, and in the allocation of my time and efforts. Also, I am particularly grateful for the many conferences, workshops, and seminars I was allowed to attend in various parts of the world, made possible by the provided funding and encouragement to present any new results.

I am also very thankful to Camilla Hollanti for the fruitful collaborations, helpful advice, and inviting me to Helsinki twice for a total of almost half a year. During these visits she made sure that I always felt welcome and part of the group, which made them great experiences, not only scientifically, but also personally. Many thanks also go to Alexey Frolov for inviting me to Moscow for one month and the interesting collaboration that ensued.

I would also like to thank Salim El Rouayheb for agreeing to be a reviewer of this dissertation and Sebastian Steinhorst for serving as the committee chair.

For every project that filled the past four and a half years I was lucky enough to work with fantastic, talented researchers. These collaborations ultimately led to the results presented in this dissertation and shaped my own development as a researcher. For that, I am deeply thankful to my co-authors Matteo Allaix, Hannes Bartz, Ragnar Freij-Hollanti, Alexey Frolov, Masahito Hayashi, Camilla Hollanti, Stanislav Kruglik, Jie Li, Hedongliang Liu, Alessandro Neri, Tefjol Pllaha, Rina Polyanskaya, Nikita Polyanskii, Sven Puchinger, Johan Rosenkilde, Vladimir Sidorenko, Seunghoan Song, Ilya Vorobyev, Antonia Wachter-Zeh, and Eitan Yaakobi.

Aside from the more directed efforts resulting in publications, I also learned a lot from the frequent discussions, seminars, and meetings held at the institute and, in particular, within the group for Coding and Cryptography. These interactions sharpened the tools required to tackle the problems presented in the following and I would like to thank all members of the institute for that. Special thanks go out to my long-term

office-mates Andreas Lenz and Julian Renner. Our many discussions, the conferences we attended together, and the overall atmosphere in our office significantly contributed to making the past years so enjoyable.

Last but certainly not least, I would like to thank my family and friends for supporting me and always lending an open ear. In particular, I am extremely grateful for the encouragement and long-term support my parents, Barbara and Thomas Holzbaur, have provided in all aspects of my life.

Lukas Holzbaur
Munich, December 2021

Contents

1 Introduction **1**

1.1 Outline 4

2 Preliminaries **7**

2.1 Notation 7

2.2 Linear Codes 10

2.3 Codes with Locality 20

I Advances in Codes with Locality

3 Correctable Erasure Patterns in Product Topologies **29**

3.1 Introduction 29

3.2 Regular Erasure Patterns 30

3.3 Negative Results on Correctable Erasure Patterns 32

3.4 Connection between Product and Tensor-Product Codes 42

3.5 Global Redundancy $s > 0$ 44

3.6 Summary and Open Problems 47

4 Partial MDS Codes with Regeneration **49**

4.1 Introduction 49

4.2 Regenerating Codes 53

4.3 PMDS Codes with Nontrivial Global Regeneration 56

4.4 PMDS Codes with Local Regeneration 64

4.5 Discussion and Comparison of PMDS Code Constructions with Local Regeneration 78

4.6 Summary and Open Problems 80

5 Decoding of Lifted Affine-Invariant Codes **85**

5.1 Introduction 85

5.2 Bounded Distance Decoding 87

5.3 High-Error Randomized Decoding 94

5.4 Summary and Open Problems 100

6 Other Results on Codes with Locality 101

6.1 Error Decoding of Locally Repairable and Partial MDS Codes 101

6.2 Lifted Reed–Solomon Codes and Lifted Multiplicity Codes 102

6.3 Secure Codes with Accessibility for Distributed Storage 103

II Decoding of Interleaved Alternant Codes

7 Decoding of Interleaved Alternant Codes 107

7.1 Introduction 107

7.2 Decoding Algorithms for Interleaved Alternant Codes 110

7.3 Technical Preliminary Results 116

7.4 The Success Probability of Decoding Interleaved Alternant Codes . . . 121

7.5 An Upper Bound on the Probability of Successful Decoding 124

7.6 Generalization of an Upper Bound on the Probability of Miscorrection . 127

7.7 Discussion and Numerical Results 129

7.8 Summary and Open Problems 137

III Private Information Retrieval

8 Towards the Capacity of PIR from Coded and Colluding Servers 141

8.1 Introduction 141

8.2 Problem Setup 147

8.3 Preliminary Lemmas 151

8.4 The Capacity of Linear, Full Support-Rank MDS-TPIR 153

8.5 Capacity of MDS-coded TBSPIR for Schemes with Additive Randomness167

8.6 Strongly-Linear PIR Capacity 171

8.7 Summary and Open Problems 176

9 Other Results on PIR 179

9.1 Private Streaming with Convolutional Codes 179

9.2 Computational Code-Based Single-Server Private Information Retrieval 180

9.3 Quantum Private Information Retrieval from Coded and Colluding Servers181

10 Conclusion and Outlook 183

List of Figures

2.1 Lifted affine-invariant code. 26

3.1 Regularity property of erasure patterns. 33
3.2 Extension of regular erasure patterns to larger topologies. 41

4.1 DSS encoded with a locally MSR PMDS array code. 55
4.2 Codeword structure of a universal PMDS code. 71
4.3 Illustration of the local regeneration procedure. 73
4.4 Comparison of the field sizes of different regenerating PMDS code constructions. 81
4.5 Comparison of the field sizes of different regenerating PMDS code constructions. 81
4.6 Comparison of the field sizes of different regenerating PMDS code constructions. 82

5.1 Illustration of the different decoding events occurring in the local decoding of lifted affine-invariant codes. 89

7.1 Majorization of a set of integers 117
7.2 Comparison of the bounds on the success probability of decoding interleaved alternant codes for different parameters. 132
7.3 Comparison of the bounds on the success probability of decoding interleaved alternant codes for different parameters. 133
7.4 Comparison of the bounds on the success probability of decoding interleaved alternant codes for different parameters. 134
7.5 Comparison of the bounds on the success probability of decoding interleaved alternant codes for different parameters. 135

8.1 Rank properties of the Khatri-Rao product of matrices. 155

List of Tables

3.1 Correctable erasure patterns in different grid-like topologies. 31
3.2 Constructions of MR codes for different grid-like topologies. 31

4.1 Overview of the notation used in related works on regenerating and PMDS codes. 53
4.2 Comparison of field sizes of locally MSR PMDS array code constructions. 79

7.1 Overview of the different bounds on the success probability of decoding interleaved alternant codes. 130

8.1 Asymptotic PIR capacity results and conjectures. 145

Abbreviations

BCH	Bose–Ray-Chaudhuri–Hocquenghem
BD	Bounded Distance
BMD	Bounded Minimum Distance
DRGP	Disjoint Repair Group Property
DSS	Distributed Storage System
GRS	Generalized Reed–Solomon
LRC	Locally Recoverable Code
MBR	Minimum Bandwidth Regenerating
MDS	Maximum Distance Separable
ML	Maximum Likelihood
MR	Maximally Recoverable
MRD	Maximum Rank Distance
MSR	Minimum Storage Regenerating
PIR	Private Information Retrieval
PMDS	Partial Maximum Distance Separable
q-SC	q-ary Symmetric Channel
RM	Reed–Muller
RRC	Rack-Aware Regenerating Code
RS	Reed–Solomon
SD	Sector-Disk
TP	Tensor-Product

Nomenclature

Basics

$[a, b]$	Set of integers $\{i \mid a \leq i \leq b\}$
$[b]$	Set of integers $\{i \mid 1 \leq i \leq b\}$
$\mathbf{A}$	Matrix
$\mathbf{A}[l, j]$	Element in l-th row and j-th column of $\mathbf{A}$
$\mathbf{A}[l, :]$	l-th row of $\mathbf{A}$
$\mathbf{A}[:, j]$	j-th column of $\mathbf{A}$
$\mathbf{A}[\mathcal{I}, :]$	Matrix $\mathbf{A}$ restricted to the rows indexed by $\mathcal{I}$
$\mathbf{A}[:, \mathcal{I}]$ or $\mathbf{A}\vert_{\mathcal{I}}$	Matrix $\mathbf{A}$ restricted to the columns indexed by $\mathcal{I}$
$\psi_\beta(\mathcal{I})$	Mapping from thick columns $\mathcal{I}$ to corresponding columns
$\langle \mathbf{A} \rangle_{\textsf{row}}$	Row span of the matrix $\mathbf{A}$
$\langle \mathbf{A} \rangle_{\textsf{col}}$	Column span of the matrix $\mathbf{A}$
$\mathrm{colsupp}(\mathbf{A})$	The set of indices of nonzero columns of $\mathbf{A}$
$\mathbf{I}_a$	The $a \times a$ identity matrix
$\mathbf{u}$	Vector
u_i	i-th element of the vector $\mathbf{u}$
$\mathbb{Z}_q$	Ring of integers $\mod q$
$\mathbb{F}_q$ or $\mathbb{F}$	Finite field with q elements
$\mathbb{F}_q^\star$	Multiplicative subgroup of $\mathbb{F}_q$
$\mathrm{Gr}(\mathbb{F}_q^n, k)$	Set of k-dimensional subspaces of $\mathbb{F}_q^n$
$\mathcal{I} = \{i_1, i_2, i_3, \ldots\}$	Set
$\lvert\mathcal{I}\rvert$	Cardinality of set $\mathcal{I}$
$\{\{s_1, s_1, \ldots, s_2, \ldots\}\}$	Multiset
$\mathrm{supp}(\mathcal{S})$	Underlying set of the multi-set $\mathcal{S}$
$\delta_{\mathcal{S}}^{s_i}$	Multiplicity of s_i in the multiset $\mathcal{S}$
X	Random variable
$X \sim \mathbb{F}_q$	Random variable uniformly distributed over $\mathbb{F}_q$
$X_{\mathcal{I}}$	Set of random variables $\{X_j \mid j \in \mathcal{N}\}$
$\mathrm{supp}(X)$	Set of realizations of X with nonzero probability
$H(X)$	Entropy of random variable X
$I(X; Y)$	Mutual information of random variables X, Y

Matrix and Vector Products

$\mathbf{A} \cdot \mathbf{B}$	(Matrix) product
$\mathbf{A} \times \mathbf{B}$	Cartesian product
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product
$\mathbf{A} \star \mathbf{B}$	Hadamard / Star product
$\mathbf{A} \odot \mathbf{B}$	Column-wise Khatri-Rao product
$\mathbf{A} \ast \mathbf{B}$	Row-wise Khatri-Rao product
$\langle \mathbf{u}, \mathbf{v} \rangle$	Inner product

Codes

$\mathcal{C}$	Code (set of vectors/matrices)
n	Length of a code / number of servers
k	Dimension of a code
$d_{\min}$	Minimum distance of a code
t	Number of errors
$\langle \mathcal{C} \rangle_{\mathsf{row}}$	Span of the elements of $\mathcal{C}$
$\mathcal{C}^{\times \ell}$	ℓ-fold Cartesian product of $\mathcal{C}$, arranged as matrices
$A_w^{\mathcal{C}}$	Number of words of weight w in $\mathcal{C}$, i.e., w-th weight enumerator of code $\mathcal{C}$
$[n, k, d_{\min}]_q$	Linear code of length n, dimension k, and min. distance $d_{\min}$ over $\mathbb{F}_q$
$[n, k, d_{\min}; \ell]_q$	Array code with ℓ rows
ℓ	Interleaving order
$\mathbf{G}$	Generator matrix
$\mathbf{H}$	Parity-check matrix
$\mathsf{GRS}(n, d_{\min}, \boldsymbol{\beta}, \boldsymbol{\nu})$	GRS code of length n and distance d with (dual) code locators $\boldsymbol{\beta}$ and dual column multipliers $\boldsymbol{\nu}$
$\mathbb{G}(n, d_{\min}, \boldsymbol{\beta})$	Multiset of all GRS codes of length n and distance d with (dual) code locators $\boldsymbol{\beta}$
$\mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$	Multiset of all subfield subcodes of $\mathbb{G}(n, d_{\min}, \boldsymbol{\beta})$
$\mathsf{Gab}(n, d_{\min}, \boldsymbol{\beta})$	Gabidulin code of length n and distance d with dual code locators $\boldsymbol{\beta}$

Codes with Locality

n_1	Number of rows / length of column code in grid-like topology
n_2	Number of columns / length of row code in grid-like topology
b_1	Number of erasures correctable in each column

b_2	Number of erasures correctable in each row
μ	Number of local repair sets
$\mathbb{W}$	Partition of $[n]$ into local repair sets
s	Number of global parities
b	Number of local parities in LRC or PMDS code
r	Locality
ϱ	Local distance
n_l	Local length
$\mathsf{PMDS}(\mu, n_l, b, s, \mathbb{W})$	Partial MDS code
$T_{n_1 \times n_2}(b_1, b_2, s)$	Grid-like topology
$\mathbb{C}_{n_1 \times n_2}(b_1, b_2, s)$	Set of all codes for the topology $T_{n_1 \times n_2}(b_1, b_2, s)$
$\mathbb{C}^{\mathsf{MR}}_{n_1 \times n_2}(b_1, b_2, s)$	Set of all MR codes for the topology $T_{n_1 \times n_2}(b_1, b_2, s)$
$\mathbb{E}_{n_1 \times n_2}(b_1, b_2, s)$	Set of all erasure patterns correctable in the topology $T_{n_1 \times n_2}(b_1, b_2, s)$
$\mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, s)$	Set of all maximal erasure patterns correctable in the topology $T_{n_1 \times n_2}(b_1, b_2, s)$
$\mathbb{E}_q^{(a,b)}$	Set of matrices $\mathbf{E} \in \mathbb{F}_q^{a \times b}$ with at least one non-zero element in each column

Lifted Affine-Invariant Codes

z	Dimension of evaluation space of affine-invariant code
m	Dimension of evaluation space of lifted affine-invariant code
$\{\mathbb{F}_Q^z \to \mathbb{F}_q\}$	Set of all functions mapping from $\mathbb{F}_Q^z$ to $\mathbb{F}_q$
$\mathrm{ev}_{\mathbb{F}_Q^z}(f(\mathbf{x}))$	Vector containing evaluations of $f(\mathbf{x})$ in elements of $\mathbb{F}_Q^z$
$\varphi_{\mathcal{V}}$	Mapping from element of z-dimensional subspace of $\mathbb{F}_Q^m$ to element of $\mathbb{F}_Q^z$
$\mathcal{L}(\mathcal{F})$	Set of functions corresponding to the lift of function set $\mathcal{F}$
$d_{\mathcal{F}}$	Minimum distance of code given by function set $\mathcal{F}$
$t_{\mathcal{F}}$	Number of errors correctable in code given by function set $\mathcal{F}$
d_{low}	Lower bound on minimum distance of code given by function set $\mathcal{L}(\mathcal{F})$
t_{low}	Number of errors correctable according to lower bound d_{low}

PIR

n	Number of servers
$\mathcal{C}$	$[n, k]$ MDS storage code
m	Number of files
α	Subpacketization
$\mathbf{X}$	Matrix of random variables (files) with each entry $\mathbf{X}[l, j] \sim \mathbb{F}$

$X = \{X^1, \dots, X^m\}$	Set of files with $X^l \sim \mathbb{F}^{\alpha \times k}$
$\mathbf{X}^l$	File l interpreted as matrix of random variables
$\mathbf{Y} = \mathbf{X} \cdot \mathbf{G}$	Matrix of codewords stored in DSS
Y^l	Codewords corresponding to file l with $\mathrm{supp}(Y^l) = \mathcal{C}^{\times \alpha}$
$\mathbf{Y}^l$	Codewords corresponding to file l interpreted as matrix
$Y = \{Y^1, \dots, Y^m\}$	Set of codewords corresponding to all files
Y_j	Part of Y stored on server j
β	Number of thin columns per thick column / Number of iterations
b	Number of adversarial servers
t	Number of colluding servers
r	Number of nonresponsive servers
α	Number of stripes of each file
$\mathcal{Q}$	Set of query realizations $\mathrm{supp}(Q)$
$Q^i = (Q^i_1, \dots, Q^i_n)$	Query when the i-th file is requested
Q^i_j	Query sent to the j-th server when the i-th file is requested
$A^i = (A^i_1, \dots, A^i_n)$	Responses when the i-th file is requested
A^i_j	Response from the j-th server when the i-th file is requested
$\mathcal{S}$	Vector space of shared randomness
$S = (S_1, \dots, S_n) \in \mathcal{S}^n$	Randomness shared by the servers
$S_j \in \mathcal{S}$	Part of the shared randomness available to the j-th server

1 Introduction

In the seminal work of [Sha48] Shannon proposed a mathematical framework of communications that went on to severely impact the technological world and shape what is commonly referred to as the age of information. The use of channel codes, i.e., the introduction of redundancy into a communication system, is shown to allow for communication with vanishing probability of error over a variety of channels. Significant effort has since been made to allow for ever more efficient communication at rates approaching the theoretical limits and fulfilling the numerous requirements of different communication systems. The most common approach of introducing redundancy is mapping the information symbols, regarded as a vector over a finite field, to an element of a linear subspace of longer vectors, called codewords, over the same finite field. This set of viable codewords is referred to as the code and, if they form a linear subspace, a linear code. The fundamental challenge of channel coding lies in achieving a favorable trade-off between the length of the codewords, number of information symbols that can be mapped to each codeword, and some notion of distance between all viable codewords, where the latter determines the resilience against error events introduced by the respective channel.

One channel model that has received significant attention in the history of channel coding is the erasure channel. Here, the channel output is given by the channel input with some number of symbols replaced by a special erasure symbol that is not part of the input alphabet. In other words, the values of a number of symbols are lost in positions known to the receiver. It is easy to see that these erasures are correctable if no two codewords coincide in all nonerased positions. The metric that reflects this property is the Hamming metric, defined as the number of positions in which two vectors differ. For a code to provide a guarantee on the correctability of a specific number of erasures its minimum distance, i.e., the minimum over the Hamming distance between any pair of codewords, must exceed the number of erasures. Significant effort has been made towards designing codes with large minimum distance and high rate resulting in codes such as Reed–Solomon (RS) [RS60], Bose–Ray-Chaudhuri–Hocquenghem (BCH) [Hoc59; BRC60], Goppa codes [Gop70], and many more (see, e.g., [MS77;

Rot06]).

While the minimum distance of a code is an important parameter of a code, some communication systems pose more involved challenges to the used code. Consider a distributed storage system (DSS) consisting of multiple nodes, such as servers, hard drives, or other storage media. A prime objective of any such system is to guarantee the integrity of the stored data, which entails the necessity of protecting against the loss of data when nodes fail, an event that is not uncommon if the number of nodes is large. The simplest solution to address this issue is replication, where each data symbol is stored multiple times on different nodes. However, this incurs a substantial storage overhead directly proportional to the number of failed nodes the system is required to resist . For this reason, (non-trivial) channel codes can now be found in real-world systems, such as the RS codes employed in Facebook's f4 storage system [MLR$^+$14] and the Google File System [Fik10]. By storing each symbol of a codeword on a different node, a node failure now corresponds to the erasure of a symbol. The storage overhead is then determined by the number of mapped information symbols over the length of the codewords, referred to as the rate, and the guaranteed resilience against node failures by the minimum distance. These are the primary characteristics that are to be considered in the design of a storage code and the codes that achieve the optimal trade-off between them are called maximum distance separable (MDS). Assuming each node fails independently and with the same probability, employing a code of this class maximizes the mean time to data loss, i.e., the expected time until a failure event occurs that cannot be recovered.

However, as the number of storage nodes in DSSs grows, so does the frequency of node failures. In classical codes, such as RS codes, a large number of nodes needs to be involved in the recovery process, even if only a single node failed. To remedy this problem, codes with locality were introduced [CHL07; HSX$^+$12; GHSY12; HCL13], which allow for the recovery of a single or small number of nodes from only a small subset of other nodes. This improvement comes at the cost of increased storage overhead for a given minimum distance and considerable effort has been directed towards optimizing various aspects of these locally recoverable codes (LRCs) [SAP$^+$13; KPLK14; RKSV13; BPSY16; BHH13; GHJY14; PD14; TB14a; SRV15]. A class of LRCs with particularly strong erasure correction capabilities is given by partial MDS (PMDS) codes [CHL07; HCL13; BHH13; GHJY14; BK15; BPSY16; CK16; HY16; GHK$^+$17; HN20; GYBS18; MPK19], which guarantee to correct any pattern of erasures that is theoretically correctable given the locality constraints or, in other words, maximize the mean time to data loss in this setting. While these codes cover an important case of locality, real world systems can rely on more involved constraints. For example, aside from an RS code used to compensate failures within a data center, Facebook's f4 storage system also employs an additional code across data centers [MLR$^+$14]. This generalization of the concept of locality, referred to as the system's topology, has also attracted the attention of researches in recent years [GHSY12; GHK$^+$17; SRLS18; KMG19; KLR19].

Codes that are optimal with regard to a given topology, i.e., that are able to correct any erasure pattern that is theoretically correctable given the topology constraints, are referred to maximally recoverable (MR) codes.

Aside from the number of nodes involved in recovery, another major concern is the amount of data traffic caused by these events. With storage capacity in the tera- or even petabytes, this traffic can consume a lot of bandwidth on the interconnects in a DSS. For example, in Facebook's data warehouse cluster over 180 terabyte are transferred each day for compensating unavailable/failed nodes [RSG$^+$13; RSG$^+$14]. The coding theoretic solution that addresses this problem are regenerating codes, as introduced in the seminal work of [DGW$^+$10]. These codes aim to minimize the required repair bandwidth in the more likely case of a single or very few node failures and an array of constructions under different models have been proposed [SR10; RSK11; SRKR11; TWB12; PLD$^+$12; KSP$^+$13; CJM$^+$13; GPV13; PYGP13; LC14; KK16b; KGØ16; GFV17; YB17a; YB17b; SCM18; SCYM18; LTT18; HLSH19; HLH20].

Another type of channel event, referred to as a substitution error or simply error, is the replacement of a number of symbols of the channel input by some other symbol of the code alphabet. While the notion of locality fulfilled by LRCs and PMDS codes has received a lot of attention in recent years due to its applicability to DSSs, locality properties are not only beneficial for erasure correction, but also have a long history in error correction. Early error-decoding algorithms, such as the majority logic decoding algorithm for Reed–Muller (RM) codes [MS77, Ch. 13], rely on linear dependencies within many disjoint subsets of codeword positions. The number of such disjoint sets allowing for the recovery of a specific symbol is referred to as the availability of the code. Lifted affine-invariant codes are a class of codes that naturally provides strong locality and availability properties. Well-known examples include lifted RS [GKS13; Guo15; PV19] and lifted multiplicity codes [KSY14; Wu15; LW19], which can be viewed as generalizations of q-ary RM codes [DGMW70; KLP68; MCJ73]. For some parameter regimes, these classes include the best known constructions of batch codes [IKOS04] and codes with the disjoint repair group property (DRGP) [LW19].

However, exploiting the locality and availability properties of a code is only one of many possibilities to provide efficient decoding algorithms. Building on their algebraic structure, many codes without these properties can be decoded up to half their minimum distance or even beyond, when allowing for a small failure probability or a nonunique result. Interestingly, considering codes consisting of codeword matrices, where each row is a codeword of a linear code, can provide significant benefits in this context, in particular in systems with parallel transmission or where errors occur in bursts. Different instances of this class of codes, commonly referred to as interleaved codes, have been studied extensively [MK90; HV00; KL97; BKY03; BMS04; SSB09b; Nie13; YL18; CS03; PV04; Par07; SSB07; CH13; WZB14; PR17]. These results show that interleaving can allow for decoding of a very large number of errors with high success probability. A theoretical analysis of this success probability, however, poses a difficult problem and in some cases it is still necessary to rely on simulation results [CS03; PV04; Par07; SSB07; CH13; WZB14; PR17].

While channel codes can provide resilience against data loss in the event of errors or node failures, this is not the only concern in modern DSSs. As the amount of data stored in these systems increases, so does the concern for the privacy of that data. However, not only the data itself, but also knowledge of which files are requested by a given user can be delicate information. The obvious approaches to resolve this privacy concern are to hide the content of the files or the user's identity from the storage system. However, in settings where the data is offered by the storage system, as commonly the case with, e.g., movies or stock market prices, the former is not an option. Further, if the service is not publicly accessible, the required authentication eliminates the latter as an option. These shortcomings motivate the problem of private information retrieval (PIR), where only the identity of the requested file is hidden. The initial study of this problem by Chor et al. [CGKS95] was followed by considerable advances regarding the derivation of the maximal achievable information rate [SJ17; BU18; SJ18b; HKS18; HGK$^+$18; SJ18c; WS17b; WS17a; WS17c; WS19; FGH$^+$19; KLRA19] and practical PIR schemes [TGE18; TGK$^+$19; FGHK17; DE19; ZTSL20; LKH20; TSC19; ZYQT19], particularly in recent years.

1.1 Outline

This work explores several aspects of coding theory, which can be roughly divided into the three areas of codes with locality, interleaved codes, and PIR.

To begin, Chapter 2 first introduces the notation used in this work and then proceeds to give formal definitions of concepts required in the following chapters. With these preliminaries established, we move on to the core of the dissertation.

Part I investigates different aspects of codes with locality, with a focus on maximally recoverable codes. This part consists of four chapters. Chapter 3 considers MR codes for grid-like topologies and shows that a previous conjecture on the set of erasure patterns correctable by codes of this class is false. We then introduce a generic method of adding global redundancy symbols to any MR code. Chapter 4 proposes the first known construction of PMDS codes with regenerating properties. First, we consider global regeneration, where the code obtained from puncturing the local redundancy in an PMDS code is an optimal regenerating code. Then, we turn to constructing PMDS codes with local regeneration, i.e., PMDS codes where each local code is an optimal regenerating code. While the focus in Chapter 3 and Chapter 4 was on the (efficient) correction of erasures, Chapter 5 considers error decoding in the class of lifted affine-invariant codes. We introduce a new bounded minimum distance decoder that is guaranteed to succeed if the number of errors is less than half of an asymptotically tight bound on the distance. Then, we show that lifted affine-invariant codes can correct errors of very high weight with vanishing probability of decoding failure. We conclude this part of the dissertation by briefly recalling some other results on codes with locality in Chapter 6.

Part II considers the decoding of interleaved alternant codes. First, we recall a known algorithm for the decoding of interleaved RS codes, which also applies to interleaved alternant codes. The focus of Chapter 7 is then on deriving the first known lower bounds on the success probability of this decoder, when applied to interleaved alternant codes. To complement these lower bounds and to evaluate their performance, we further introduce an upper bound on the probability of successful decoding.

Part III explores different notions of PIR, focusing on the setting with coded storage and colluding nodes. The main result of Chapter 8 is the derivation of the capacity for an important subclass of linear schemes in this setting, which includes all known schemes that are flexible in terms of the applicable system parameters. Further, we derive the capacity for two more subclasses of PIR schemes. Finally, Chapter 9 provides a brief overview of other new results on PIR.

2 Preliminaries

This chapter formally introduces and briefly reviews the main concepts treated in this work.

2.1 Notation

Denote the set of integers $[a, b] \coloneqq \{i \mid a \leq i \leq b\}$ and write $[1, b] = [b]$. The cardinality of a set $\mathcal{S} = \{s_1, s_2, \ldots\}$ is denoted by $|\mathcal{S}|$. For a multiset $\mathcal{S} = \{\{s_1, s_1 \ldots, s_2, s_2, \ldots\}\}$ denote by $\delta_{\mathcal{S}}^{s_i}$ the multiplicity of s_i in $\mathcal{S}$ and by $\mathrm{supp}(\mathcal{S})$ the (non-multi) set of elements in $\mathcal{S}$. For sets $\mathcal{I}, \mathcal{J}$ denote by $\mathcal{I} \times \mathcal{J} = \{(i, j) \mid i \in \mathcal{I}, j \in \mathcal{J}\}$ their Cartesian product.

Consider an $a \times b$ matrix $\mathbf{A}$. For integers $i \in [a]$ and $j \in [b]$ denote by $\mathbf{A}[i, j]$ the element in row i and column j. Similarly, for sets $\mathcal{I} \subseteq [a]$ and $\mathcal{J} \subseteq [b]$ the submatrix obtained from restricting the matrix $\mathbf{A}$ to the rows/columns indexed by $\mathcal{I}$ and $\mathcal{J}$, respectively, is given by $\mathbf{A}[\mathcal{I}, \mathcal{J}]$. To restrict to a given subset of rows write $\mathbf{A}[\mathcal{I}, :]$ and for the restriction to a subset of columns $\mathbf{A}[:, \mathcal{J}]$ or $\mathbf{A}|_{\mathcal{J}}$. For the linear span of the rows/columns of $\mathbf{A}$ write $\langle \mathbf{A} \rangle_{\mathsf{row}}$ and $\langle \mathbf{A} \rangle_{\mathsf{col}}$, respectively. The transpose of $\mathbf{A}$ is denoted $\mathbf{A}^\top$. The set of indices of the non-zero columns of $\mathbf{A}$ is denoted by $\mathrm{colsupp}(\mathbf{A})$. Write $\mathrm{diag}(\mathbf{A}_1, \mathbf{A}_2, \ldots)$ to denote the block diagonal matrix with matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots$ on the diagonal. By slight abuse of notation, write $\mathrm{diag}(\mathbf{a})$ for the diagonal matrix with the entries of the vector $\mathbf{a}$ on the main diagonal.

Write $\mathbf{0}_b$ and $\mathbf{1}_b$ to denote the all-zero and all-one vector of length b.

For a prime power q, the finite (extension) field of size q is denoted by $\mathbb{F}_q$ and its multiplicative subgroup by $\mathbb{F}_q^\star \coloneqq \mathbb{F}_q \setminus \{0\}$. If the field size q is not of interest, it is omitted from the notation. For an element $\alpha \in \mathbb{F}$ denote its order, i.e., the smallest non-zero integer i such that $\alpha^i = 1$, by $\mathrm{order}(\alpha)$. The ring of s-variate polynomials in the variables $\mathbf{x} = (x_1, \ldots, x_s)$ with coefficients in $\mathbb{F}_q$ is denoted by $\mathbb{F}_q[x_1, \ldots, x_s]$ or $\mathbb{F}_q[\mathbf{x}]$. For the ring of integers modulo q, write $\mathbb{Z}_q$.

Write $\mathbb{F}_q^n$ for the vector space consisting of all length n vectors over $\mathbb{F}_q$. Similarly, the matrix space of all $a \times b$ matrices over $\mathbb{F}_q$ is denoted $\mathbb{F}_q^{a \times b}$.

The dimension of a linear subspace $\mathcal{V} \subseteq \mathbb{F}_q^n$ is denoted by $\dim_q(\mathcal{V})$. The field size q is omitted if obvious from context. The set of all linear subspaces of $\mathbb{F}_q^n$ of dimension k, i.e., the Grassmannian, is denoted by $\mathrm{Gr}(\mathbb{F}_q^n, k)$. The number of such subspaces is given by the Gaussian binomial coefficient

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \begin{cases} \frac{(1-q^n)(1-q^{n-1})\dots(1-q^{n-k+1})}{(1-q)(1-q^2)\dots(1-q^k)}, & k \leq n, \\ 0, & k > n. \end{cases}$$

For a random variable X denote the set of realizations with non-zero probability by $\mathrm{supp}(X)$. If X is uniformly distributed over $\mathrm{supp}(X)$, write $X \sim \mathrm{supp}(X)$. For a set of integers $\mathcal{I}$ denote $X_\mathcal{I} = \{X_j \mid j \in \mathcal{I}\}$. The expected value of a random variable X is denoted $\mathbb{E}(X)$.

The indicator function is defined to be

$$\mathbb{1}\{\mathit{statement}\} := \begin{cases} 1, & \text{if } \mathit{statement} \text{ is true}, \\ 0, & \text{if } \mathit{statement} \text{ is false}. \end{cases}$$

2.1.1 Vector and Matrix Multiplication

In the following chapters several different notions of matrix products are used. For completeness, we introduce them here and restate some known results on their relation.

Regular matrix/vector/scalar product: For $\mathbf{A} \in \mathbb{F}^{m \times n}$ and $\mathbf{B} \in \mathbb{F}^{n \times n'}$ we have

$$\mathbf{A} \cdot \mathbf{B} = \begin{pmatrix} \langle \mathbf{A}[1,:], \mathbf{B}[:,1] \rangle & \langle \mathbf{A}[1,:], \mathbf{B}[:,2] \rangle & \cdots & \langle \mathbf{A}[1,:], \mathbf{B}[:,n'] \rangle \\ \langle \mathbf{A}[2,:], \mathbf{B}[:,1] \rangle & \langle \mathbf{A}[2,:], \mathbf{B}[:,2] \rangle & \cdots & \langle \mathbf{A}[2,:], \mathbf{B}[:,n'] \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{A}[m,:], \mathbf{B}[:,1] \rangle & \langle \mathbf{A}[m,:], \mathbf{B}[:,2] \rangle & \cdots & \langle \mathbf{A}[m,:], \mathbf{B}[:,n'] \rangle \end{pmatrix} \in \mathbb{F}^{m \times n'} ,$$

where the *inner product* between two vectors is

$$\langle \mathbf{A}[i,:], \mathbf{B}[:,j] \rangle = \sum_{l=1}^{n} \mathbf{A}[i,l] \cdot \mathbf{B}[l,j] .$$

If obvious from context, we omit the $\cdot$ symbol.

Star-product / Hadamard product: For $\mathbf{A} \in \mathbb{F}^{m \times n}$ and $\mathbf{B} \in \mathbb{F}^{m \times n}$ we have

$$\mathbf{A} \star \mathbf{B} = \begin{pmatrix} \mathbf{A}[1,1] \cdot \mathbf{B}[1,1] & \mathbf{A}[1,2] \cdot \mathbf{B}[1,2] & \cdots & \mathbf{A}[1,n] \cdot \mathbf{B}[1,n] \\ \mathbf{A}[2,1] \cdot \mathbf{B}[2,1] & \mathbf{A}[2,2] \cdot \mathbf{B}[2,2] & \cdots & \mathbf{A}[2,n] \cdot \mathbf{B}[2,n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}[m,1] \cdot \mathbf{B}[m,1] & \mathbf{A}[m,2] \cdot \mathbf{B}[m,2] & \cdots & \mathbf{A}[m,n] \cdot \mathbf{B}[m,n] \end{pmatrix} \in \mathbb{F}^{m \times n} .$$

Kronecker product: For $\mathbf{A} \in \mathbb{F}^{m \times n}$ and $\mathbf{B} \in \mathbb{F}^{m' \times n'}$ we have

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} \mathbf{A}[1,1] \cdot \mathbf{B} & \mathbf{A}[1,2] \cdot \mathbf{B} & \cdots & \mathbf{A}[1,n] \cdot \mathbf{B} \\ \mathbf{A}[2,1] \cdot \mathbf{B} & \mathbf{A}[2,2] \cdot \mathbf{B} & \cdots & \mathbf{A}[2,n] \cdot \mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}[m,1] \cdot \mathbf{B} & \mathbf{A}[m,2] \cdot \mathbf{B} & \cdots & \mathbf{A}[m,n] \cdot \mathbf{B} \end{pmatrix} \in \mathbb{F}^{mm' \times nn'} .$$

Column-wise Khatri-Rao product [KR68]: For $\mathbf{A} \in \mathbb{F}^{m \times n}$ and $\mathbf{B} \in \mathbb{F}^{m' \times n}$ we have

$$\mathbf{A} \odot \mathbf{B} = \begin{pmatrix} \mathbf{A}[:,1] \otimes \mathbf{B}[:,1] & \mathbf{A}[:,2] \otimes \mathbf{B}[:,2] & \cdots & \mathbf{A}[:,n] \otimes \mathbf{B}[:,n] \end{pmatrix} \in \mathbb{F}^{mm' \times n} .$$

Row-wise Khatri-Rao product / face-splitting product [KR68; Sly97]: For $\mathbf{A} \in \mathbb{F}^{m \times n}$ and $\mathbf{B} \in \mathbb{F}^{m \times n'}$ we have

$$\mathbf{A} * \mathbf{B} = \begin{pmatrix} \mathbf{A}[1,:] \otimes \mathbf{B}[1,:] \\ \mathbf{A}[2,:] \otimes \mathbf{B}[2,:] \\ \vdots \\ \mathbf{A}[m,:] \otimes \mathbf{B}[m,:] \end{pmatrix} \in \mathbb{F}^{m \times nn'} .$$

For sets of vectors (codes) define

$$\begin{aligned} \langle \mathbf{A} \rangle_{\mathsf{row}} \star \langle \mathbf{B} \rangle_{\mathsf{row}} &:= \langle \{ \mathbf{a} \star \mathbf{b} \mid \mathbf{a} \in \langle \mathbf{A} \rangle_{\mathsf{row}}, \mathbf{b} \in \langle \mathbf{B} \rangle_{\mathsf{row}} \} \rangle_{\mathsf{row}} \\ \langle \mathbf{A} \rangle_{\mathsf{row}} \odot \langle \mathbf{B} \rangle_{\mathsf{row}} &:= \langle \{ \mathbf{a} \odot \mathbf{b} \mid \mathbf{a} \in \langle \mathbf{A} \rangle_{\mathsf{row}}, \mathbf{b} \in \langle \mathbf{B} \rangle_{\mathsf{row}} \} \rangle_{\mathsf{row}} \\ \langle \mathbf{A} \rangle_{\mathsf{row}} \otimes \langle \mathbf{B} \rangle_{\mathsf{row}} &:= \langle \mathbf{A} \otimes \mathbf{B} \rangle_{\mathsf{row}} . \end{aligned}$$

It is easy to check that these properties are independent of the choice of the bases $\mathbf{A}$ and $\mathbf{B}$ of the respective row spaces.

The following proposition collects some well-known properties of the introduced matrix products.

Proposition 2.1 (See, *e.g.*, [Sly97] and [HJ91, Lemma 4.2.10.]). *Consider matrices*

$\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ *and a row vector* $\mathbf{z}$. *Then it holds that*

$$(\mathbf{A} \cdot \mathbf{B}) \star (\mathbf{C} \cdot \mathbf{D}) = (\mathbf{A} * \mathbf{C}) \cdot (\mathbf{B} \odot \mathbf{D}) \tag{2.1}$$

$$(\mathbf{A} \cdot \mathbf{B}) \otimes \mathbf{z} = \mathbf{A} \cdot (\mathbf{B} \otimes \mathbf{z}) \tag{2.2}$$

$$\langle \mathbf{A} \rangle_{\mathsf{row}} \star \langle \mathbf{B} \rangle_{\mathsf{row}} = \langle \mathbf{A} \odot \mathbf{B} \rangle_{\mathsf{row}} = \langle \mathbf{A} \rangle_{\mathsf{row}} \odot \langle \mathbf{B} \rangle_{\mathsf{row}} \ . \tag{2.3}$$

For $\mathbf{A} \in \mathbb{F}^{m \times n}$ *we have*

$$\begin{aligned} &\mathbf{1}_m \cdot (\mathbf{A} * \mathbf{I}_m) \\ &\quad = \Big(\mathbf{A}[1,1], \mathbf{A}[2,1], \dots, \mathbf{A}[m,1], \mathbf{A}[1,2], \mathbf{A}[2,2], \dots, \mathbf{A}[m,2], \dots, \mathbf{A}[m,n]\Big) \in \mathbb{F}^{1 \times mn} \ , \end{aligned} \tag{2.4}$$

where $\mathbf{I}_m$ *denotes the* $m \times m$ *identity matrix. Moreover, if the matrix* $\mathbf{A}$ *is uniformly distributed over* $\mathbb{F}^{k \times m}$, *then* $\mathbf{1}_m \cdot (\mathbf{A} * \mathbf{I}_m)$ *is uniformly distributed over* $\mathbb{F}^{1 \times km}$.

2.2 Linear Codes

By its most general definition, a code is simply a set of elements. However, most literature on coding theory considers codes where these elements, called *codewords*, are vectors of equal length n over a given field, also referred to as *scalar codes*. If this subset is a k-dimensional linear subspace of $\mathbb{F}_q^n$, the code is *linear* and we denote it by $[n,k]_q$ or $[n,k]$, if the field size is obvious from context or not of interest.

Consider a basis of a k-dimensional subspace/code, i.e., a set of k codewords that spans the code. A matrix $\mathbf{G} \in \mathbb{F}_q^{k \times n}$ that contains such a basis as its rows is called a *generator matrix* and the code is given by

$$\mathcal{C} = \{\mathbf{u} \cdot \mathbf{G} \mid \mathbf{u} \in \mathbb{F}_q^k\} \ .$$

Here, the vector $\mathbf{u}$ is commonly referred to as the *message vector* and the mapping $\mathbf{c} = \mathbf{u} \cdot \mathbf{G}$ as the *encoding* of the message. As each linear subspace has a unique dual space, the code can equivalently be described by

$$\mathcal{C} = \{\mathbf{c} \mid \mathbf{c} \in \mathbb{F}_q^n, \mathbf{c} \cdot \mathbf{H}^\top = \mathbf{0}\} \ ,$$

where the rows of $\mathbf{H}$ form a basis of the dual space of $\mathcal{C}$, also referred to as the dual code $\mathcal{C}^\perp$.

The desired properties of a code depend on the metric under consideration. In this work, we only consider the Hamming metric. The *Hamming weight* of a vector $\mathbf{c} \in \mathbb{F}_q^n$ is the number of non-zero positions

$$\mathrm{wt}(\mathbf{c}) = |\operatorname{colsupp}(\mathbf{c})| \ .$$

For two vectors $\mathbf{c}, \mathbf{c}' \in \mathbb{F}_q^n$ their *Hamming distance* is defined to be the number of positions in which the vectors differ and formally given by

$$\mathrm{d}_{\mathsf{H}}(\mathbf{c},\mathbf{c}') = |\operatorname{colsupp}(\mathbf{c}-\mathbf{c}')| \ .$$

As we only consider the Hamming metric in the following, we simply write *weight* and *distance*, respectively. A code parameter of particular interest in coding theory is the *minimum distance* of a linear code $\mathcal{C}$, i.e., the minimal number of positions in which *any* two codewords differ, formally defined to be

$$d_{\min} = \min_{\substack{\mathbf{c},\mathbf{c}'\in\mathcal{C}\\ \mathbf{c}\neq\mathbf{c}'}} \mathrm{d}_{\mathsf{H}}(\mathbf{c},\mathbf{c}') = \min_{\mathbf{c}\in\mathcal{C}\setminus\{\mathbf{0}\}} \mathrm{wt}(\mathbf{c}) \ ,$$

where the second equality holds by linearity of the code. When the minimum distance is of interest and not obvious from context, it is include it in the notation and we write $[n,k,d_{\min}]$ to denote a corresponding linear code. In the asymptotic analysis of codes one commonly considers the behaviour of the distance relative to the code length ${}^{d_{\min}}/_{n}$, also referred to as the *relative* or *normalized* distance.

Generally, it is advantageous for the minimum distance to be as high as possible. There is a substantial body of work investigating the bounds on the minimum distance given the other code parameters. The bound of highest importance for this work is the Singleton bound (cf. [MS77, Chapter 17]), which states that for any $[n,k,d_{\min}]$ code it holds that

$$d_{\min} \leq n-k+1 \ . \tag{2.5}$$

Codes that attain this bound with equality are called *maximum distance separable* (MDS) codes. One interesting property of MDS codes is that the number of codewords of Hamming weight w in any MDS code $\mathcal{C}$, i.e., its *weight enumerators*

$$A_w^{\mathcal{C}} := |\{\mathbf{c} \mid \mathrm{wt}(\mathbf{c}) = w, \mathbf{c}\in\mathcal{C}\}| \ ,$$

are completely determined by its length and distance.

Theorem 2.1 (Weight Enumerators of MDS Codes [MS77, Ch. 11, Theorem 6])**.** *Let $\mathcal{C}$ be an $[n,k,d_{\min}]_q$ MDS code. The w-th weight enumerator A_w^{MDS} of $\mathcal{C}$ is*

$$A_w^{\mathcal{C}} = \begin{cases} 1, & \text{if } w=0,\\ \binom{n}{w}\sum_{j=0}^{w-d_{\min}}(-1)^j\binom{w}{j}(q^{w-d_{\min}+1-j}-1), & \text{else.}\end{cases}$$

To emphasize the fact that the values of $A_w^{\mathcal{C}}$ are independent of the specific MDS code $\mathcal{C}$, we simply write A_w^{MDS} in the following.

MDS codes are known to exist for any combination of length n and dimension k, however, only if the field size is sufficiently large[1]. To remedy this shortcoming, a

[1] Extended RS codes are MDS, exist for any prime power q, and require $q \geq n-1$. Determining the

multitude of bounds also takes the field size into account. Among the best known are the Elias, Griesmer, Hamming, Linear Programming, and Plotkin bound. For the sake of brevity, we do not recall these bounds here, but instead refer the interested reader to [MS77, Chapter 17].

In storage applications, which are the focus of this work, it is common to consider a generalization of scalar codes where the elements are matrices, all with the same number of rows and columns, instead of vectors. In other words, these *array codes*[2] are subsets of $\mathbb{F}_q^{\ell \times n}$. The parameter ℓ is also referred to as the *subpacketization* of the code. Note that the definition of the length, dimension, weight, and distance as given above also translate to this setting, with the only difference being that we consider *columns* of the codeword matrices in place of the positions of the codeword vectors of scalar codes. We denote these codes by $[n, k, d_{\min}; \ell]_q$. Again, if the minimum distance $d_{\min}$ and/or the alphabet size q are not of interest or clear from context, we write $[n, k; \ell]$, $[n, k, d_{\min}; \ell]$, or $[n, k; \ell]_q$, respectively.

For an $[n, k, d_{\min}; \ell]$ code $\mathcal{C}$ and a set of integers $\mathcal{I} \subseteq [n]$ we write $\mathcal{C}|_{\mathcal{I}}$ for the code obtained by restricting each codeword $\mathbf{C} \in \mathcal{C}$ to the positions/columns indexed by $\mathcal{I}$, i.e., $\mathcal{C}|_{\mathcal{I}} = \{\mathbf{C}|_{\mathcal{I}} \mid \mathbf{C} \in \mathcal{C}\}$.

2.2.1 Erasure Correction

The main purpose of linear codes is to correct errors and/or erasures that occurred during the transmission over some channel. When erasures occur in specific positions of a codeword, these (columns of) symbols are replaced by an erasure symbol, which we denote by $\ast$. Clearly, a code $\mathcal{C}$ can correct a set of erasures $\mathcal{E} \subset [n]$, also called *erasure pattern*, if and only if there do not exist two (or more) codewords that coincide in all non-erased (or *surviving*) positions $[n] \setminus \mathcal{E}$. As any two codewords differ in at least $d_{\min}$ positions, a code of minimum distance $d_{\min}$ therefore guarantees to correct any combination of up to $|\mathcal{E}| \leq d_{\min} - 1$ erasures, independent of their positions in the codeword. However, this is only a sufficient condition for correctability, not a necessary one. In general, an erasure pattern is correctable if the mapping from the message $\mathbf{u} \in \mathbb{F}_q^k$ to the surviving positions, given by $\mathbf{c}|_{[n]\setminus\mathcal{E}} = \mathbf{u} \cdot \mathbf{G}|_{[n]\setminus\mathcal{E}}$, is still injective, where $\mathbf{G}$ denotes a generator matrix of the code $\mathcal{C}$. It is easy to see that this is the case if and only if the submatrix $\mathbf{G}|_{[n]\setminus\mathcal{E}}$ is of full rank k. The minimal sets of columns for which this holds, i.e., the sets $\mathcal{I} \subset [n]$ with $|\mathcal{I}| = k$ such that $\operatorname{rank}(\mathbf{G}|_{\mathcal{I}}) = k$, are called *information sets* of the code $\mathcal{C}$. Consequently, in order to show that a code can correct a given erasure pattern, it suffices to show that the complement of this pattern

field size q for which an MDS code exists for a given n and k is one of the big open problems in coding theory referred to as the *MDS conjecture*, which was posed in [Seg55].

[2] In the literature, these codes are also referred to as *vector codes*. This term stems from the fact that in storage applications each node commonly stores a column of a given codeword and therefore, from the perspective of a single node, codes where the codewords are matrices/arrays imply storing a *vector*. To avoid this misleading term, we only refer to these codes as *array codes* in this work.

contains an information set of the code. A similar condition for the correctability of an erasure pattern can be given in terms of the dual code $\mathbf{C}^{\perp}$ and the parity check matrix $\mathbf{H}$. By simple linear algebra arguments, it is easy to show that the erased positions are only correctable if they contain an information set of the dual code, i.e., $\mathrm{rank}(\mathbf{H}|_{\mathcal{E}}) = n-k$.

We formally define two operations on linear codes that are particularly useful in the context of erasure correction.

Definition 2.1. *Let $\mathcal{C}$ be an $[n,k]$ code and $\mathcal{I} \subseteq [n]$ be a set of integers. Define the* shortening *operator as*

$$\mathrm{short}_{\mathcal{I}}(\mathcal{C}) = \{\mathbf{c}|_{[n]\setminus\mathcal{I}} \mid \mathbf{c} \in \mathcal{C}, \mathbf{c}|_{\mathcal{I}} = \mathbf{0}_{|\mathcal{I}|}\}$$

and the puncturing *operator as*

$$\mathrm{punct}_{\mathcal{I}}(\mathcal{C}) = \{\mathbf{c}|_{[n]\setminus\mathcal{I}} \mid \mathbf{c} \in \mathcal{C}\} \ .$$

The duality of the shortening and puncturing operations is well-known, namely, for any given set $\mathcal{I} \subset [n]$ it holds that $\mathrm{short}_{\mathcal{I}}(\mathcal{C})^{\perp} = \mathrm{punct}_{\mathcal{I}}(\mathcal{C}^{\perp})$. Note that in our notation we have $\mathrm{punct}_{\mathcal{I}}(\mathcal{C}) = \mathcal{C}|_{[n]\setminus\mathcal{I}}$.

2.2.2 Error Decoding

In contrast to erasures, the main difficulty when considering errors is that their position is generally unknown. Assume a codeword $\mathbf{c}$ of a q-ary code $\mathcal{C}$ is transmitted over a channel that introduces errors, i.e., corrupts some number of positions (columns) of the codeword. The output of this channel, or *received word*, can be written as $\mathbf{r} = \mathbf{c} + \mathbf{e}$, where the support of $\mathbf{e}$ gives the set of *error positions*. In this case, the receiver is interested in *decoding* this received word, i.e., determining the codeword that is most likely to be the transmitted codeword. Assuming that the transmitted codeword $\mathbf{c}$ is drawn from the code $\mathcal{C}$ uniformly at random, the decoder that optimally solves this problem is called *maximum likelihood* (ML) decoder [MS77, Chapter 1]. For the q-ary symmetric channel (q-SC) with $\epsilon < {}^{q-1}/_{q}$, where each position of $\mathbf{e}$ is either 0 with probability $1-\epsilon$ or any other field element $\alpha \in \mathbb{F}_q^{\star}$ with probability ${}^{\epsilon}/_{q-1}$, the result of an ML decoder is well-known to be

$$\mathbf{c}' = \arg\min_{\mathbf{c}' \in \mathcal{C}}(\mathrm{d}_{\mathsf{H}}(\mathbf{c}+\mathbf{e}, \mathbf{c}')) \ .$$

We say the code can correct the error $\mathbf{e}$ uniquely if $\mathbf{c}' = \mathbf{c}$. From a purely mathematical perspective, an ML decoder optimally solves the decoding problem under the given constraints. However, actually finding an efficiently implementable ML decoder for a given code is extremely difficult. To circumvent these difficulties and allow for practical algorithms to be designed, the problem can be relaxed. One such relaxation, that is of particular interest for this work, is *bounded minimum distance* (BMD) decoding. There, given a code of minimum distance $d_{\min}$, the decoder corrects any error up to weight $\mathrm{wt}(\mathbf{e}) = \lfloor {}^{d_{\min}-1}/_2 \rfloor$, also referred to as the *unique decoding radius* of the code. Since any two codewords of this code differ in at least $d_{\min}$ positions, an error of this weight is guaranteed to be correctable. Efficient BMD decoders exist for many classes of linear codes, however, note that this is generally not a simple problem and it is unknown how to efficiently decode a random linear code up to its unique decoding radius. In fact, a whole branch of cryptography is based on the assumption that this is a hard problem and cryptographic schemes, such as the well-known McEliece cryptosystem [McE78], rely on this hardness assumption.

While BMD decoding has the advantage of providing a decoding guarantee, it is limited to half the minimum distance. One approach to remedy this is *list decoding*. Here, the decoding radius is increased beyond the unique decoding radius and the goal of the decoder is to return a list of all codewords that are within this extended radius. Hence, as long as the number of errors that occurred is below the list decoding radius, the correct codeword will be in this list. Formally, a q-ary code of length n is called (τ, L)-list-decodable if the Hamming sphere of radius τ centered at any vector $\mathbf{v} \in \mathbb{F}_q^n$ always contains at most L codewords $\mathbf{c} \in \mathcal{C}$. One question of interest in this context is the maximal radius that still guarantees to result in a list of size polynomial in the code length. Interestingly, it has been shown that *any* linear $[n, k, d_{\min}]_q$ can be decoded up to the q-ary Johnson radius [Joh62; Bas65].

2.2.3 Interleaved Codes

One class of codes that will be of particular interest in Chapter 7 are array codes where each row of the codeword array is a codeword of the same $\mathbb{F}_q$-linear code. In other words, codes of this class, also referred to as *homogeneous interleaved codes*, are direct sums of a *constituent code*.

Definition 2.2 (Homogeneous Interleaved Code [MK90; KL97]). *Let $\mathcal{C}$ be a linear $[n, k, d_{\min}]_q$ code and $\ell \in \mathbb{N}$. The corresponding ℓ-interleaved code is defined to be*

$$\mathcal{IC}[n, k, d_{\min}; \ell] := \mathcal{C}^{\times \ell} = \{\mathbf{C} \mid \mathbf{C}[i, :] \in \mathcal{C} \ \forall i \in [\ell]\} \ .$$

The parameter ℓ is referred to as the interleaving order *and the code $\mathcal{C}$ as the* constituent code.

The great advantage of this class of codes is that it allows for efficient decoding algorithms that correct errors beyond the unique decoding radius, where the number of errors is given as the number of corrupted *columns* of the codeword array. We note that in coding theory literature errors and erasures are more commonly defined as the corruption/erasing of a symbol of the base field $\mathbb{F}_q$ instead of a column. However, there are many practical motivations for considering this type of error, also referred to as a "burst errors", such as replicated file disagreement location [MK90], data-storage applications [KL97], suitable outer codes in concatenated codes [MK90; KL98; HV99; JTH04; SSB05; SSB09b], ALOHA-like random-access [HV99], decoding scalar codes beyond half-the-minimum distance by power decoding [SSB10; Kam14; Ros18; PRB19], and recently code-based cryptography [EWZ18; HLPW19].

One decoder that allows for increasing the decoding radius by collaboratively decoding the rows of an interleaved codes, was introduced in [MK90], rediscovered in [HV99], and generalized in [HV00; RV14]. The remarkable property of this algorithm is that it can be applied to decode up to $t \leq d_{\min} - 2$ errors in an interleaved codes with an *arbitrary* constituent code $\mathcal{C}$ with high probability, provided that the interleaving order is sufficiently large.

Other decoding algorithms for interleaved codes can also be applied in settings with smaller interleaving order, however, only for specific classes of constituent codes. The first such algorithm was given in [KL97] for interleaved RS codes and corrects up to ${}^{\ell}/_{\ell+1}(n-k)$ errors. Since then, many decoders with better complexity and larger decoding radius, as well as some bounds on the probability of decoding failure have been derived [BKY03; CS03; PV04; BMS04; Par07; SSB07; SSB09b; CH13; Nie13; WZB14; PR17; YL18]. Other code classes that have been considered as constituent codes of interleaved codes are one-point Hermitian codes [Kam14; PRB19] and, more generally, algebraic-geometry codes [BMS05].

One interesting property of homogeneous interleaved codes is that they can also be viewed as codes over a larger field with the same parameters.

Corollary 2.1. *Let $\{\gamma_1, \dots, \gamma_\ell\}$ be a basis of $\mathbb{F}_{q^\ell}$ over $\mathbb{F}_q$ and $\mathcal{C}$ be an $[n, k, d_{\min}]_q$. Then the code*

$$\left\{(\gamma_1, \dots, \gamma_\ell) \cdot \mathbf{C} \mid \mathbf{C} \in \mathcal{C}^{\times \ell}\right\} \simeq \langle \mathcal{C} \rangle_{\mathbb{F}_{q^\ell}}$$

is an $[n, k, d_{\min}]_{q^\ell}$ code.

To simplify the notation when considering errors in interleaved codes, we denote by $\mathbb{E}_q^{(a,b)}$ the set of matrices $\mathbf{E} \in \mathbb{F}_q^{a \times b}$ with at least one non-zero element in each column. Note that the matrices $\widetilde{\mathbf{E}}$ with $|\operatorname{colsupp}(\widetilde{\mathbf{E}})| =: t$ fulfill $\widetilde{\mathbf{E}}|_{\operatorname{colsupp}(\widetilde{\mathbf{E}})} \in \mathbb{E}_q^{(\ell,t)}$. Further, the set obtained by mapping such matrices to the corresponding vectors $\tilde{\mathbf{e}} \in \mathbb{F}_{q^\ell}^n$ by the bijective mapping $\mathbb{F}_q^{\ell \times n} \mapsto \mathbb{F}_{q^\ell}^n$ as in Corollary 2.1, is the subset of all vectors of weight exactly t. Specifically, for an arbitrary basis $\{\gamma_1, \dots, \gamma_\ell\}$ of $\mathbb{F}_{q^\ell}$ over $\mathbb{F}_q$ we have

$$\{(\gamma_1, \dots, \gamma_\ell) \cdot \widetilde{\mathbf{E}} \mid \widetilde{\mathbf{E}} \in \mathbb{F}_q^{\ell \times n}, \widetilde{\mathbf{E}}|_{\operatorname{colsupp}(\widetilde{\mathbf{E}})} \in \mathbb{E}_q^{(\ell,t)}\} = \{\tilde{\mathbf{e}} \mid \tilde{\mathbf{e}} \in \mathbb{F}_{q^\ell}^n, \operatorname{wt}(\tilde{\mathbf{e}}) = t\} .$$

Corollary 2.1 applies to arbitrary constituent codes. However, if the constituent codes are RS codes, the statement can be refined, as in this case, it is well-known that the resulting code is also an RS codes (cf. [SSB08]) with the same evaluation points, which are in a subfield $\mathbb{F}_q$ of the interleaved code's field $\mathbb{F}_{q^\ell}$.

2.2.4 Generalized Reed–Solomon Codes

Generalized Reed–Solomon (GRS) codes are a particularly popular class of MDS codes that is easy to construct, can be efficiently decoded, acts well as a constituent code of interleaved codes, and possesses many more desirable properties. There are several different possibilities to define GRS codes, below we collect those required in this work.

Definition 2.3 (Generalized Reed–Solomon Codes [MS77, Chapter 10])**.** *For positive integers n and $d_{\min}$, let $\boldsymbol{\beta} \in \mathbb{F}_q^n$ be a vector of distinct* code locators *and $\boldsymbol{\nu} \in (\mathbb{F}_{q^m}^{\star})^n$ be a vector of* column multipliers. *We define an $[n, k = n - d_{\min} + 1]$* generalized Reed–Solomon *code $\mathsf{GRS}(n, d_{\min}, \boldsymbol{\beta}, \boldsymbol{\nu})_{q^m}$ as*

$$\mathsf{GRS}(n, d_{\min}, \boldsymbol{\beta}, \boldsymbol{\nu})_{q^m} := \{\mathbf{c} \in \mathbb{F}_{q^m}^n \;\mid\; \mathbf{H} \cdot \operatorname{diag}(\boldsymbol{\nu}) \cdot \mathbf{c}^\top = \mathbf{0}\} \;,$$

with

$$\mathbf{H} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \beta_1 & \beta_2 & \dots & \beta_n \\ \vdots & \vdots & & \vdots \\ \beta_1^{d_{\min}-2} & \beta_2^{d_{\min}-2} & \dots & \beta_n^{d_{\min}-2} \end{pmatrix} \in \mathbb{F}_{q^m}^{(d_{\min}-1)\times n} \;.$$

Equivalently, the code is defined by

$$\begin{aligned}\mathsf{GRS}(n, d_{\min}, \boldsymbol{\beta}, \boldsymbol{\nu})_{q^m} := \{&(\hat{\nu}_1 f(\beta_1), \hat{\nu}_2 f(\beta_2), \dots, \hat{\nu}_n f(\beta_n)) \\ &\mid f(x) \in \mathbb{F}_{q^m}[x], \deg(f(x)) < n - d_{\min} + 1\}\end{aligned}$$

with (cf. [Rot06, Problem 5.7])

$$\hat{\nu}_i = \nu_i^{-1} \prod_{\substack{j\in[n]\\ j\neq i}} (\beta_i - \beta_j)^{-1}, \; i \in [n] \;.$$

Denote by $\mathbb{G}(n, d_{\min}, \boldsymbol{\beta})_{q^m}$ the multi-set

$$\mathbb{G}(n, d_{\min}, \boldsymbol{\beta})_{q^m} = \{\{\mathsf{GRS}(n, d_{\min}, \boldsymbol{\beta}, \boldsymbol{\nu})_{q^m} \;\mid\; \boldsymbol{\nu} \in (\mathbb{F}_{q^m}^{\star})^n\}\} \;.$$

If the specific choice of the column multipliers $\boldsymbol{\nu}$, the code locators $\boldsymbol{\beta}$, and/or the field size q^m are not important, we omit them and write $\mathsf{GRS}(n, d_{\min}, \boldsymbol{\beta})$ or $\mathsf{GRS}(n, d_{\min})$, respectively. GRS codes are well-known to be MDS, i.e., they fulfill the Singleton bound, as given in Eq. (2.5), with equality.

2.2.5 Alternant Codes

By design, GRS codes must be defined over fields $\mathbb{F}_{q^m}$ with $q^m \geq n$. In many applications it is desirable to work with codes of smaller field size, which can be obtained, e.g., by taking subcodes of codes defined over larger fields.

Definition 2.4 (Subfield Subcode). *Let $\mathcal{C}$ be an $[n,k,d_{\min}]_{q^m}$ code. We define the $\mathbb{F}_q$-subfield subcode of $\mathcal{C}$ as*

$$\mathcal{C} \cap \mathbb{F}_q^n = \{\mathbf{c} \mid \mathbf{c} \in \mathcal{C}, c_i \in \mathbb{F}_q \ \forall \ i \in [n]\} \ .$$

Equivalently, let $\mathbf{H} \in \mathbb{F}_{q^m}^{(n-k)\times n}$ be a parity check matrix of $\mathcal{C}$. Then $\mathcal{C} \cap \mathbb{F}_q^n$ is given by the $\mathbb{F}_q$ kernel of $\mathbf{H}$, i.e.,

$$\mathcal{C} \cap \mathbb{F}_q^n = \{\mathbf{c} \mid \mathbf{H} \cdot \mathbf{c}^\top = \mathbf{0}, \mathbf{c} \in \mathbb{F}_q^n\} \ .$$

A class of subfield subcodes that has received considerable attention is that of subfield subcodes of GRS codes.

Definition 2.5 (Alternant Code [MS77, Ch. 12.2]). *The subfield subcode of a GRS code is referred to as an* alternant code. *For a fixed set of code locators $\boldsymbol{\beta}$ as in Definition 2.3 and designed distance $d_{\min}$, we define the multi-set of alternant codes as*

$$\mathbb{A}(n, d_{\min}, \boldsymbol{\beta}) = \{\{\mathcal{C} \cap \mathbb{F}_q^n \mid \mathcal{C} \in \mathbb{G}(n, d_{\min}, \boldsymbol{\beta})\}\} \ .$$

Note that the parameter $d_{\min}$ is not (necessarily) the actual minimum distance of the alternant code. In fact, for specific alternant codes it is known that the distance is larger (see Remark 2.1). However, it serves as a lower bound that applies to all alternant codes obtained by taking subfield subcodes of the codes in $\mathbb{G}(n, d_{\min}, \boldsymbol{\beta})$.

We define $\mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$ as a multiset, since the multiplicities will be important in the following. One further advantage is that for a given code length n, we know its cardinality to be

$$|\mathbb{A}(n, d_{\min}, \boldsymbol{\beta})| = |\mathbb{G}(n, d_{\min}, \boldsymbol{\beta})| = (q^m - 1)^n \ . \tag{2.6}$$

For GRS codes it is known [Del75] that for a fixed set of code locators $\boldsymbol{\beta}$, it holds that $\mathsf{GRS}(n, d_{\min}, \boldsymbol{\beta}, \boldsymbol{\nu}) = \mathsf{GRS}(n, d_{\min}, \boldsymbol{\beta}, \boldsymbol{\nu}')$ if and only if $\boldsymbol{\nu}'$ is an $\mathbb{F}_{q^m}$-multiple of $\boldsymbol{\nu}$, i.e., any code $\mathcal{C} \in \mathbb{G}(n, d_{\min}, \boldsymbol{\beta})$ occurs with multiplicity exactly $\delta^{\mathcal{C}}_{\mathbb{G}(n,d_{\min},\boldsymbol{\beta})} = q^m - 1$ in $\mathbb{G}(n, d_{\min}, \boldsymbol{\beta})$. This gives a lower bound on the multiplicity of alternant codes by

$$\delta^{\mathcal{A}}_{\mathbb{A}(n,d_{\min},\boldsymbol{\beta})} \geq q^m - 1 \ \forall \ \mathcal{A} \in \mathbb{A}(n, d_{\min}, \boldsymbol{\beta}) \ . \tag{2.7}$$

We give some general well-known bounds on the dimension of the $\mathbb{F}_q$-subcode of an $\mathbb{F}_{q^m}$-linear code $\mathcal{C}$ in terms of the parameters of $\mathcal{C}$.

Lemma 2.1. *Let $\mathcal{C}$ be an $[n,k,d_{\min}]_{q^m}$ code. Then*

$$\max\{n - m(n-k), 0\} \leq \dim_q(\mathcal{C} \cap \mathbb{F}_q^n) \leq \min\{k, k_q^{\mathsf{opt.}}(n, d_{\min})\} \ ,$$

where $k_q^{\mathsf{opt.}}(n, d_{\min})$ is an upper bound on the dimension of a q-ary linear code of length n and minimum distance $d_{\min}$.

Proof. The lower bound of 0 is trivial. The lower bound of $n - m(n-k)$ follows from expanding the $n-k$ rows of any parity-check matrix of $\mathcal{C}$ over some basis of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$. The resulting $m(n-k) \times n$ matrix is a parity check matrix of the $\mathbb{F}_q$-subcode of $\mathcal{C} \cap \mathbb{F}_q^n$ and the bound follows.

The upper bound of $k_q^{\mathsf{opt.}}(n, d_{\min})$ follows from the fact that the distance of the code $\mathcal{C} \cap \mathbb{F}_q^n$ is at least that of $\mathcal{C}$. Finally, if elements in $\mathbb{F}_q^n$ are $\mathbb{F}_q$-linearly independent, then they are also $\mathbb{F}_{q^m}$-linearly independent for every extension field $\mathbb{F}_{q^m}$ of $\mathbb{F}_q$. Therefore, $\dim_q(\mathcal{C} \cap \mathbb{F}_q^n) \leq k$. □

Remark 2.1 (Dimension vs. Distance of Binary BCH and Wild Goppa Codes)**.** *Wild Goppa codes [SKHN76; Wir88], which include* binary square-free Goppa codes *[Gop70; Gop71; Ber73], are a subclass of Goppa Codes. Along with BCH codes [Hoc59; BRC60], Goppa codes are the best known class of alternant codes, due to their good distance properties. Now, consider the binary BCH and q-ary wild Goppa codes that are subfield subcodes of a GRS code in $\mathbb{G}(n, d_{\min}, \boldsymbol{\beta})$ for some $\boldsymbol{\beta}$ and $d_{\min}$.*

For binary BCH codes, it is well-known (cf. [MS77, Ch. 7]) that their dimension is $k_{\mathsf{BCH}} \geq n - m\frac{n-k}{2}$, for length n and dimension $k := n - d_{\min} + 1$ of the corresponding GRS code. Therefore, the dimension of binary BCH codes exceeds the generic lower bound of Lemma 2.1.

Wild q-ary Goppa codes on the other hand are often considered as alternant codes of $\mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$, but with an increased minimum distance $d_{\mathsf{Goppa}} \approx \frac{q}{q-1} d_{\min}$. However, for the purpose of this work it is more convenient to view them as alternant codes of $\mathbb{A}(n, d_{\mathsf{Goppa}}, \boldsymbol{\beta})$ with a larger dimension *than guaranteed by the lower bound in Lemma 2.1 instead of alternant codes in $\mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$ with increased distance. This is possible as the improvements of wild Goppa codes compared to other alternant codes can be shown by proving an equivalence between the Goppa codes obtained from different Goppa polynomials (cf. [SKHN76], [BLP11, Theorem 4.1]), which directly implies that $\mathcal{C}_{\mathsf{Goppa}} \in \mathbb{A}(n, d_{\min}, \boldsymbol{\beta}) \cap \mathbb{A}(n, d_{\mathsf{Goppa}}, \boldsymbol{\beta})$ for $d_{\mathsf{Goppa}} > d_{\min}$. Clearly, the "good" distance follows immediately from the code being in $\mathbb{A}(n, d_{\mathsf{Goppa}}, \boldsymbol{\beta})$, while the dimension can be shown to be large by applying the lower bound of Lemma 2.1 corresponding to $\mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$.*

2.2.6 Gabidulin Codes

Another well-known class of evaluation codes, which is closely related to GRS codes, is known as Gabidulin codes [Del78; Gab85; Rot91]. These codes are most popular for being of maximum rank distance (MRD), the equivalent to the MDS property in the rank-metric. However, they also possess properties that make them attractive for constructions in the Hamming metric, as will be evident in Chapters 3 and 4.

Definition 2.6 (Gabidulin codes)**.** *Let n and $d_{\min}$ be positive integers and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n) \in \mathbb{F}_{q^M}^n$ be such that the β_i are linearly independent over $\mathbb{F}_q$. The $[n, k, d_{\min}]_{q^M}$ Gabidulin code $\mathsf{Gab}(n, d_{\min}, \boldsymbol{\beta})_{q^M}$ is defined to be*

$$\mathsf{Gab}(n, d_{\min}, \boldsymbol{\beta})_{q^M} = \left\{ \mathbf{c} \mid \mathbf{c} \cdot \mathbf{H}^\top = \mathbf{0}, \mathbf{c} \in \mathbb{F}_{q^M}^n \right\}$$

with

$$\mathbf{H} = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_n \\ \beta_1^{q^1} & \beta_2^{q^1} & \cdots & \beta_n^{q^1} \\ \vdots & \vdots & & \vdots \\ \beta_1^{q^{d_{\min}-2}} & \beta_2^{q^{d_{\min}-2}} & \cdots & \beta_n^{q^{d_{\min}-2}} \end{pmatrix} \in \mathbb{F}_{q^M}^{d_{\min}-1 \times n} .$$

Note that the existence of linearly independent β_i implies $n \leq M$. The set $\{\beta_i\}, i \in [n]$ is referred to as the *code locators* of the Gabidulin code. While the column multipliers of GRS codes allow for the code locators in their generator and the parity-check matrix to be the same, this is generally not the case for the code locators of a Gabidulin code and its dual code. In the following, when we refer to the code locators of a Gabidulin code, we always refer to the β_i used for the *parity-check matrix*, as in Definition 2.6.

The codewords of an $\mathsf{Gab}(n, d_{\min}, \boldsymbol{\beta})_{q^M}$ Gabidulin code can be seen as matrices in $\mathbb{F}_q^{M \times n}$ by expanding elements of $\mathbb{F}_{q^M}$ into vectors in $\mathbb{F}_q^M$ using a fixed basis of $\mathbb{F}_{q^M}$ over $\mathbb{F}_q$. Thus, we can define the rank distance of two codewords as the rank of their matrix representations' difference. It is well-known that the minimum rank distance of a Gabidulin code is $n - k + 1$, i.e., it fulfills the Singleton-like bound in the rank metric with equality. Further, as the rank of the $\mathbb{F}_q$-expansion of a vector is a lower bound on its Hamming weight, Gabidulin codes are also MDS, i.e., fulfill the Singleton bound in the Hamming metric with equality.

We recall another well-known property of Gabidulin codes. For completeness we include short proof.

Lemma 2.2 (Isometries of Gabidulin Codes[Ber03, Lemma 3])**.** *Let $\mathbf{G} \in \mathbb{F}_{q^M}^{k \times n}$ be a generator matrix of the Gabidulin code $\mathsf{Gab}(n, d_{\min}, \boldsymbol{\beta})_{q^M}$. Then, for any full-rank matrix $\mathbf{A} \in \mathbb{F}_q^{n \times n}$, the code*

$$\mathcal{C}' = \langle \mathbf{G} \cdot \mathbf{A} \rangle$$

is the Gabidulin code $\mathsf{Gab}(n, d, \boldsymbol{\beta}')$ with $\boldsymbol{\beta}' = \mathbf{A}^{-1} \cdot \boldsymbol{\beta}^\top$.

Proof. Let $\mathbf{H}, \mathbf{H}' \in \mathbb{F}_{q^M}^{d_{\min}-1 \times n}$ be the parity-check matrices, as given in Definition 2.6, of the codes $\mathsf{Gab}(n, d_{\min}, \boldsymbol{\beta})$ and $\mathsf{Gab}(n, d_{\min}, \boldsymbol{\beta}')$, respectively. By definition, we have

$$\begin{aligned} \mathbf{0} &= \mathbf{G} \cdot \mathbf{H}^\top \\ &= \mathbf{G} \cdot \mathbf{A} \cdot \underbrace{\mathbf{A}^{-1}\mathbf{H}^\top}_{\overset{(\mathsf{a})}{=}\mathbf{H}'^\top}, \end{aligned}$$

where (a) follows from the fact that $\lambda_i\beta_i^{q^l} + \lambda_j\beta_j^{q^l} = (\lambda_i\beta_i + \lambda_j\beta_j)^{q^l}$ $\forall$ $\lambda_i, \lambda_j \in \mathbb{F}_q$. As $\mathbf{A}$ is of full rank over $\mathbb{F}_q$, we have $\operatorname{rank}_q(\boldsymbol{\beta}') = \operatorname{rank}_q(\boldsymbol{\beta})$. Further, if the elements of $\boldsymbol{\beta}$ are linearly independent, so are the elements of $\boldsymbol{\beta}'$, thereby fulfilling the requirements of Definition 2.6 on the code locators. □

2.3 Codes with Locality

In the classical erasure and error decoding problems, as discussed in Sections 2.2.1 and 2.2.2, the goal of the decoder is to determine the correct codeword while being able to access *all* surviving positions. However, this is not desirable in some applications, as access to individual positions can, e.g., be costly in terms of required time, hardware utilization, or communication overhead. To remedy this problem, codes that allow for the correction of some erasures and errors from small subsets of positions have been introduced.

2.3.1 Locally Recoverable and Partial MDS Codes

The seminal work [GHSY12] established some fundamental results on a class of codes with locality, commonly referred to as locally recoverable codes (LRCs). Non-rigorously, a code is said to have locality r if every position can be recovered from at most r other codeword positions. Equivalently, for every position, there has to exist a codeword in the dual code of weight at most r that is supported on this position. If multiple erasures can be tolerated within such a *local repair set*, the code is said to have (r, ϱ)-locality.

Definition 2.7 ((r, ϱ)-locality (see, e.g., [KPLK14])). *Let $n, k, r, \varrho, \mu \in \mathbb{Z}_{>0}$ with $\mu \geq 2$. The $[n, k, d_{\min}]$ code $\mathcal{C}$ has (r, ϱ)-locality if there exists a partition $\mathbb{W} = \{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_\mu\}$ of $[n]$ into sets of cardinality $|\mathcal{W}_j| \leq r + \varrho - 1$ such that for all $j \in [\mu]$ it holds that $\mathrm{d}_{\mathsf{H}}\left(\mathcal{C}|_{\mathcal{W}_j}\right) \geq \varrho$.*

Note that, by the Singleton bound (see Eq. (2.5)), the condition on the distance of the local codes implies, that any subset of $\varrho - 1$ positions of a local repair set can be recovered by accessing at most r positions of that local repair set. The parameter r is therefore called the *locality* of the code.

In the following, $\mathcal{W}_j$ is referred to as the j-th local repair set and the code $\mathcal{C}|_{\mathcal{W}_j}$ as the j-th local code. For simplicity, we only consider codes where every local code is of the same length $n_l = |\mathcal{W}_1| = \ldots = |\mathcal{W}_\mu|$ with $n_l \mid n$.

A Singleton-like upper bound on the achievable distance of an $[n, k]$ code with (r, ϱ)-locality was derived in [GHSY12] for $\varrho = 2$ and generalized for $\varrho \geq 2$ in [KPLK14] to

$$d_{\min} \leq n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\varrho - 1) \ . \tag{2.8}$$

In the following we refer to codes achieving this bound with equality as *distance-optimal* or simply *optimal* LRCs. Several classes of optimal LRCs are known [CHL07; HCL13; BHH13; RKSV13; KPLK14; PD14; TB14a; TPD16; BPSY16; MPK19; HN20; GYBS18] for a wide range of parameters.

While LRCs that fulfill the bound of Eq. (2.8) are optimal in terms of their minimum distance, they are not necessarily optimal with respect to the set of erasure patterns they can correct. Codes that are also optimal in this regard, are called maximally recoverable LRCs [CHL07; HCL13; GHJY14; GHK$^+$17; MPK19] or partial MDS (PMDS) codes [BHH13; GYBS18; BPSY16; CK16; HN20]. The codes of this class of LRCs correct all erasure patterns that are information-theoretically correctable, given the locality constraints. These patterns are exactly those with b erasures in each local code plus s additional erasures in arbitrary positions [BHH13; GHJY14]. For consistency with literature on PMDS codes, we provide a formal definition of this special class of LRCs.

Definition 2.8 (Partial MDS codes (see, e.g., [BHH13]))**.** *Let $n, \mu, b, s \in \mathbb{Z}_{>0}$ be such that $\mu \geq 2$, $b < n_l$, and $s \leq (n_l - b)(\mu - 1)$. Let $\mathbb{W} = \{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_\mu\}$ be a partition of $[\mu n_l]$ with $|W_i| = n_l \ \forall \ i \in [\mu]$.*

Let $\mathcal{C} \subset \mathbb{F}_q^{\mu n_l}$ be a linear $[\mu n_l, (n-b)\mu - s]$ code. The code $\mathcal{C}$ is a $\mathsf{PMDS}(\mu, n_l, b, s, \mathbb{W})$ partial MDS code *if*

- *the code $\mathcal{C}|_{\mathcal{W}_i}$ is an $[n_l, n_l - b, b + 1]$ MDS code for all $i \in [\mu]$ and*
- *for any $\mathcal{E}_i \subset \mathcal{W}_i$ with $|\mathcal{E}_i| = b \ \forall \ i \in [\mu]$, the code $\mathcal{C}|_{[\mu n_l] \setminus \cup_{i=1}^{\mu} \mathcal{E}_i}$ is an $[\mu n_l - b\mu, \mu n_l - b\mu - s, s + 1]$ MDS code.*

Equivalently, a PMDS array code, denoted by $\mathsf{PMDS}(\mu, n_l, b, s, \mathbb{W}; \ell)$, is defined as a code where the corresponding restrictions are MDS array codes. Trivially, PMDS codes are optimal with respect to the Singleton-like bound of Eq. (2.8). We refer to parameters n, μ, b, s satisfying the constraints of Definition 2.8 as *valid PMDS parameters.* Note that this excludes parameters for which the definition results in a trivial PMDS code. One trivial case is given by $b = 0$ (and arbitrary μ, s), where the code obtained from "puncturing $b = 0$ positions" in each local repair set, i.e., the unpunctured code, is MDS by the second property of Definition 2.8. On the other hand, if $s = 0$ (and arbitrary μ, b) the code is just a concatenation of (independent) local MDS codes. The requirement $s \leq (n-b)(\mu-1)$ is necessary for the PMDS code definition since otherwise the dimension of the local code exceeds the overall dimension — a contradiction.

Remark 2.2. *A more general definition of PMDS codes, where the local code (including its parameters such as length and distance, i.e., number of tolerable erasures) can be different in each local repair set is sometimes considered in literature [HN20; NH20]. For simplicity, we focus on PMDS where all local codes are the same code in this work but note that the approaches taken in Chapter 4 can also be generalized to PMDS codes with distinct local codes.*

A relaxation of the PMDS properties leads to the class of sector-disk (SD) codes. While this class is not the focus of in the following, some results on PMDS codes of Chapter 4 also apply to SD codes and therefore we include a short definition here. Informally, this class differs from PMDS codes by only requiring the second property to hold for the puncturing of the same positions in each local repair set.

Remark 2.3 (Sector-Disk Code)**.** *Consider the partition $\mathbb{W} = \{\mathcal{W}_1, \dots, \mathcal{W}_\mu\}$ with $\mathcal{W}_i = [(i-1)n_l + 1, in_l]$. A Sector-Disk $\mathsf{SD}(\mu, n, b, s, \mathbb{W}; \ell)$ code is defined similar to a PMDS codes as in Definition 2.8, except that the second property only needs to holds for $\mathcal{E}_i = \{(i-1)n_l + j \mid j \in \mathcal{E}_{\mathsf{SD}}\}$ and any $\mathcal{E}_{\mathsf{SD}} \in [n_l]$ with $|\mathcal{E}_{\mathsf{SD}}| = b$.*

Remark 2.4. *In [BPSY16; GYBS18] each codeword of the PMDS and SD codes is regarded as a $\mu \times n$ array. As we will construct PMDS and SD codes with local regeneration in Chapter 4, we require subpacketization, i.e., each node does not store a symbol, but a vector of multiple symbols. To avoid having different types of rows, we adopt the terminology to the one commonly used in the LRC literature and view the codewords of a PMDS or sector-disk (SD) code as vectors. Hence, what we refer to as* local codes *is equivalent to the* rows *of [BPSY16; GYBS18].*

2.3.2 Grid-Like Topologies

While LRCs and PMDS codes represent an important notion of locality, namely, the case where the subsets of codeword positions that need to fulfill the locality constraints are disjoint, some applications require a more sophisticated structure of the locality. Consequently, the general concept of codes for topologies was introduced in [GHJY14]. In general, a topology is defined to be a restriction on the support of the parity-check matrix of a code, i.e., a set of positions that are fixed to be zero. Given such a topology T, we say code $\mathcal{C}$ is a *code for the topology* T if there exists a matrix $\mathbf{H}$ with $\mathcal{C} = \langle \mathbf{H} \rangle_{\mathsf{row}}^{\perp}$ that fulfills these support restrictions.

One important special case is given by grid-like topologies $T_{n_1 \times n_2}(b_1, b_2, 0)$, for which several important results were established in the seminal paper [GHK$^+$17]. In such a topology, each column/row of a codeword, interpreted as an $n_1 \times n_2$ array, is a codeword of a column/row code with b_1 and b_2 parity equations, respectively. This class of codes is well-known in coding theory and also referred to as *product codes*. The topology $T_{n_1 \times n_2}(b_1, b_2, s)$ augments these local parity checks by s additional parity-checks that are not restricted in their support. Hence, codes for this topology are product codes with s additional *global* parity constraints. We formally define codes for grid-like topologies by adapting the notation of [GHK$^+$17].

Definition 2.9 (Code for grid-like topology [GHK$^+$17, Definition 2.1])**.** *Let $\mathcal{C}_{\mathsf{col}}$ be an $[n_1, \geq n_1 - b_1]$ code and $\mathcal{C}_{\mathsf{row}}$ be an $[n_2, \geq n_2 - b_2]$ code. We define a code for the topology $T_{n_1 \times n_2}(b_1, b_2, s)$ to be any code with parity-check matrix*

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{\mathsf{local}} \\ \mathbf{H}_{\mathsf{global}} \end{pmatrix} ,$$

where $\mathbf{H}_{\mathsf{local}}$ is a parity-check matrix of the code $\mathcal{C}_{\mathsf{col}} \otimes \mathcal{C}_{\mathsf{row}}$ and $\mathbf{H}_{\mathsf{global}}$ is an arbitrary $s \times n_1 n_2$ matrix.

We denote the set of all such codes by $\mathbb{C}_{n_1 \times n_2}(b_1, b_2, s)$.

By definition, a code $\mathcal{C} \in \mathbb{C}_{n_1 \times n_2}(b_1, b_2, s)$ is a subset of $\mathbb{F}^{n_1 n_2}$. However, we frequently use the equivalent interpretation as a subset of $\mathbb{F}^{n_1 \times n_2}$, where each column/row is a codeword of $\mathcal{C}_{\mathsf{col}}$ and $\mathcal{C}_{\mathsf{row}}$, respectively.

Throughout this work, let n_1, n_2, b_1, b_2, s be non-negative integers which satisfy $n_1 > b_1$ and $n_2 > b_2$. Further, to exclude trivial cases where the dimension of the code $\mathcal{C}$ is smaller than the dimension of the codes $\mathcal{C}_{\mathsf{col}}$ and/or $\mathcal{C}_{\mathsf{row}}$, we assume that $s \leq (n_1 - b_1)(n_2 - b_2) - \max\{n_1 - b_1, n_2 - b_2\}$ for the remainder of this work.

It is easy to see that PMDS codes and LRCs, as defined in Section 2.3.1, are a special case of codes for grid-like topologies given by $T_{\mu \times n_l}(0, b, s) = T_{\mu \times n_l}(0, \varrho - 1, d_{\min} - \varrho)$.

While product codes have been studied extensively (see, e.g., [MS77, Chapter 18] and the references therein), the main difference between these works and [GHK$^+$17], aside from the additional global parities, is the goal of determining a characterization of all erasure patterns that are correctable in these topologies.

Definition 2.10 (Correctable erasure pattern [GHK$^+$17, Definition 2.2])**.** *Let $\mathcal{E} \subseteq [n_1] \times [n_2]$ denote a set of erased positions. We say the* erasure pattern *$\mathcal{E}$ is* correctable in the topology *$T_{n_1 \times n_2}(b_1, b_2, s)$ if and only if there exists a code in $\mathbb{C}_{n_1 \times n_2}(b_1, b_2, s)$ that can correct this erasure pattern.*

We denote the set of erasure patterns which are correctable in $T_{n_1 \times n_2}(b_1, b_2, s)$ by $\mathbb{E}_{n_1 \times n_2}(b_1, b_2, s)$ and by $\mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, s)$ those that are not a proper subset of any other correctable pattern.

Similar to the definition of a PMDS code, an MR code for a given topology must be able to correct all theoretically correctable erasures patterns.

Definition 2.11 (Maximally recoverable code [GHK+17, Definition 2.3])**.** *We say a code $\mathcal{C} \in \mathbb{C}_{n_1 \times n_2}(b_1, b_2, s)$ is* maximally recoverable *(MR) if it corrects every erasure pattern in $\mathbb{E}_{n_1 \times n_2}(b_1, b_2, s)$.*

We denote the set of codes that are MR for a topology $T_{n_1 \times n_2}(b_1, b_2, s)$ by $\mathbb{C}^{\mathsf{MR}}_{n_1 \times n_2}(b_1, b_2, s)$.

For a sufficiently large finite field, the existence of an MR code for any topology was proved in [GHJY14].

Observe that, any code that can correct an erasure pattern $\mathcal{E}$ can also correct any erasure pattern $\mathcal{E}' \subset \mathcal{E}$. Hence, an MR code is equivalently defined as being able to correct any pattern in $\mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, s)$ instead of $\mathbb{E}_{n_1 \times n_2}(b_1, b_2, s)$. By the same argument, the set of correctable erasure patterns $\mathbb{E}_{n_1 \times n_2}(b_1, b_2, s)$ is uniquely determined by $\mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, s)$.

We summarize some important properties of MR codes for grid-like topologies in terms of our notation.

Proposition 2.2 (Properties of codes for grid-like topologies [GHK+17, Proposition 2.1])**.** *For any $\mathcal{C} \in \mathbb{C}^{\mathsf{MR}}_{n_1 \times n_2}(b_1, b_2, s)$ it holds that*

- *the dimensions of $\mathcal{C}, \mathcal{C}_{\mathsf{col}}$, and $\mathcal{C}_{\mathsf{row}}$ are*

$$\begin{gathered} \dim(\mathcal{C}) = (n_1 - b_1)(n_2 - b_2) - s \\ \dim(\mathcal{C}_{\mathsf{col}}) = n_1 - b_1 \quad \textit{and} \quad \dim(\mathcal{C}_{\mathsf{row}}) = n_2 - b_2 \ , \end{gathered}$$

- *the codes $\mathcal{C}_{\mathsf{col}}$ and $\mathcal{C}_{\mathsf{row}}$ are MDS.*

2.3.3 Lifted Affine-Invariant Codes

Lifted affine-invariant codes are a class of codes that naturally provides strong locality properties, however, in a different sense than LRCs and codes for grid-like topologies. Affine-invariant codes are best described by regarding them as functions mapping between certain domains such as vector spaces over finite fields.

Denote by $\{\mathbb{F}_Q^z \to \mathbb{F}_q\}$ the set of all functions mapping from $\mathbb{F}_Q^z$ to $\mathbb{F}_q$, where Q is a power of q. In this case, for any $f \in \{\mathbb{F}_Q^z \to \mathbb{F}_q\}$, there exists a unique polynomial $f(\mathbf{x}) \in \mathbb{F}_Q[x_1, \ldots, x_z]$ of degree at most $Q-1$ in each variable corresponding to the function f.

For a function $f(\mathbf{x}) \in \{\mathbb{F}_Q^z \to \mathbb{F}_q\}$ with $\mathbf{x} = (x_1, x_2, \ldots, x_z)$, we define the evaluation map

$$\mathrm{ev}_{\mathbb{F}_Q^z}(f(\mathbf{x})) := (f(\mathbf{a}))_{\mathbf{a} \in \mathbb{F}_Q^z} \in \mathbb{F}_q^{Q^z} \ .$$

Similarly, by slight abuse of notation, we write

$$\mathrm{ev}_{\mathbb{F}_Q^z}(\mathcal{F}) := \{\mathrm{ev}_{\mathbb{F}_Q^z}(f(\mathbf{x})) \mid f(\mathbf{x}) \in \mathcal{F}\}$$

for a set of functions $\mathcal{F} \subseteq \{\mathbb{F}_Q^z \to \mathbb{F}_q\}$. If the set of functions $\mathcal{F}$ is linear, i.e., it contains $\lambda f + g$ for any $\lambda \in \mathbb{F}_Q$ and $f, g \in \mathcal{F}$, the set $\mathrm{ev}_{\mathbb{F}_Q^z}(\mathcal{F})$ is a linear code.

A function $A(\mathbf{x}) \in \{\mathbb{F}_Q^t \to \mathbb{F}_Q^t\}$ is called *affine* if it can be represented as $\mathbf{M}\mathbf{x} + \mathbf{b}$ for some matrix $\mathbf{M} \in \mathbb{F}_Q^{z \times z}$ and vector $\mathbf{b} \in \mathbb{F}_Q^z$ (if $\mathbf{b} = 0$ the function is linear). If $\mathbf{M}$ is of full rank, then A is said to be an *affine permutation*.

With these ingredients we are now ready to define affine-invariant codes.

Definition 2.12 (Affine-Invariant Code)**.** *The code $\mathrm{ev}_{\mathbb{F}_Q^z}(\mathcal{F})$ is said to be affine-invariant if for every affine permutation function $A \in \{\mathbb{F}_Q^t \to \mathbb{F}_Q^t\}$ and for every $f(\mathbf{x}) \in \mathcal{F}$, the function $f(A(\mathbf{x}))$ belongs to $\mathcal{F}$.*

Many important properties of affine-invariant codes were derived in [KS08; BSGM⁺11; GKS13].

Example 2.1 (RS codes are affine-invariant)**.** *One simple example of affine-invariant codes are RS codes. By Definition 2.3, they contain the evaluation of any univariate polynomial of restricted degree. Specifically, the $[n, k]$ RS code over $\mathbb{F}_q$ with locators $\boldsymbol{\beta} = \mathbb{F}_q$ is given by*

$$\mathcal{C} := \left\{ \mathrm{ev}_{\mathbb{F}_q}(f(x)) \mid f(x) \in \{\mathbb{F}_q \to \mathbb{F}_q\}, \deg(f(x)) < k \right\} .$$

Now consider an affine permutation. As $z = 1$ in the case of RS codes, the permutation is given by $A(x) = mx + b$ for some scalars $m \in \mathbb{F}_q^\star$ and $b \in \mathbb{F}_q$. Clearly, for any $f(x)$, the polynomial $f'(x) = f(mx + b)$ is of the same degree as $f(x)$. Hence, we have $\deg(f(A(x))) = \deg(f(x)) < k$ and therefore $\mathrm{ev}_{\mathbb{F}_q}(f(A(x))) \in \mathcal{C}$. It follows that the code is affine invariant.

On a high level, the idea of a lifted code is best described as the property that the restriction to any subspace of a certain dimension is a codeword of a given code — the code that is lifted to the higher dimension. To represent this property, we require a formal definition of the restriction of a function to a given domain.

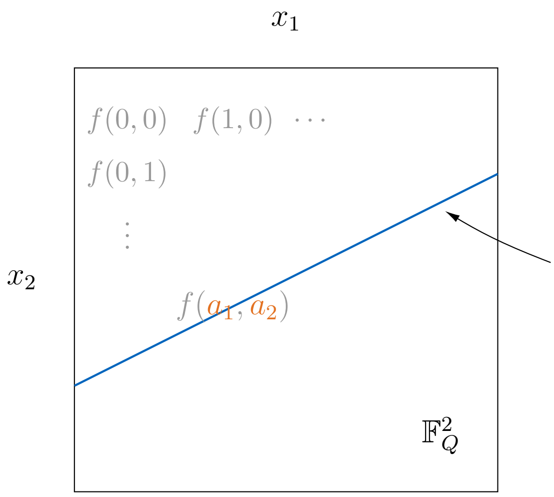
For a fixed basis $\{\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_z\}$ of a z-dimensional linear vector space $\mathcal{V}$ over $\mathbb{F}_Q^m$ define the linear map $\varphi_{\mathcal{V}} : \ \mathcal{V} \mapsto \mathbb{F}_Q^z$ by

$$\varphi_{\mathcal{V}} \left(\sum_{j=1}^{z} \lambda_j \boldsymbol{\gamma}_j \right) := (\lambda_1, \dots, \lambda_z) \ \in \mathbb{F}_Q^z.$$

For a function $g \in \{\mathbb{F}_Q^m \to \mathbb{F}_q\}$ and an affine subspace $\mathcal{V} + \mathbf{a}$, where $\mathbf{a} \in \mathbb{F}_Q^m$, define the function $g_{\mathbf{a}}^{(\mathcal{V})} \in \{\mathbb{F}_Q^z \to \mathbb{F}_q\}$ as

$$g_{\mathbf{a}}^{(\mathcal{V})}(\mathbf{y}) := g(\varphi_{\mathcal{V}}^{-1}(\mathbf{y}) + \mathbf{a}) \ . \tag{2.9}$$

Note that, for any two points $\mathbf{a}, \mathbf{a}' \in \mathcal{V}$ of the same subspace $\mathcal{V}$ we have $g_{\mathbf{a}'}^{(\mathcal{V})}(\mathbf{y}) = g_{\mathbf{a}}^{(\mathcal{V})}(\mathbf{y} + \varphi_{\mathcal{V}}(\mathbf{a}' - \mathbf{a}))$



One-dimensional affine subspace $\mathcal{V}+\mathbf{a}$ of $\mathbb{F}_Q^2$. Evaluations of $f(\mathbf{x})$ in this subspace are a codeword of $\mathcal{F}$.

Figure 2.1: Illustration of a lifted affine-invariant code for $m = 2$ and $z = 1$.

Definition 2.13 (Lifted Affine-Invariant Code [GKS13, Definition 1.1])**.** *Let* $\mathrm{ev}_{\mathbb{F}_Q^z}(\mathcal{F})$ *be an affine-invariant code with* $\mathcal{F} \subseteq \{\mathbb{F}_Q^z \to \mathbb{F}_q\}$*. Denote by* $\mathcal{L}(\mathcal{F}) \subseteq \{\mathbb{F}_Q^m \to \mathbb{F}_q\}$ *the set of functions* $f \in \mathcal{L}(\mathcal{F})$ *which fulfill that* $f_{\mathbf{a}}^{(\mathcal{V})} \in \mathcal{F}$ *for any* z*-dimensional affine subspace* $\mathcal{V} + \mathbf{a} \subset \mathbb{F}_Q^m$*. Then the lifted code is given by* $\mathrm{ev}_{\mathbb{F}_Q^m}(\mathcal{L}(\mathcal{F}))$*.*

An illustration of a lifted affine-invariant code for $m = 2$ and $z = 1$ is given in Fig. 2.1.

The minimum distance of a lifted affine-invariant code can be tightly bounded from above and below.

Lemma 2.3 (Distance of Lifted Affine-Invariant Code [GKS13, Lemma 5.7])**.** *Let* $\mathrm{ev}_{\mathbb{F}_Q^z}(\mathcal{F})$ *with* $\mathcal{F} \subseteq \{\mathbb{F}_Q^z \to \mathbb{F}_q\}$ *be an affine-invariant code of distance* $d_{\mathcal{F}}$*. Then the distance* $d_{\mathcal{L}(\mathcal{F})}$ *of the lifted code* $\mathrm{ev}_{\mathbb{F}_Q^m}(\mathcal{L}(\mathcal{F}))$*, as in Definition 2.13, is bounded by*

$$(d_{\mathcal{F}} - 1)\frac{Q^m}{Q^z - 1} < d_{\mathcal{L}(\mathcal{F})} \leq d_{\mathcal{F}} Q^{m-z} \ .$$

Part I

Advances in Codes with Locality

3
Correctable Erasure Patterns in Product Topologies

Abstract

Locality enables storage systems to recover failed nodes from small subsets of surviving nodes. The setting where nodes are partitioned into subsets, each allowing for local recovery, is well understood. This chapter considers a generalization of this setting introduced by Gopalan et al., where, viewing the codewords as arrays, constraints are imposed on the columns and rows in addition to some global constraints. Specifically, new results on the set of correctable erasure patterns are derived and a generic method of adding such global parity-checks is presented. Further, the set of correctable erasure patterns in topologies without global parities is related to those correctable in tensor-product codes.

This chapter is based on the work [HPYW21] published in the proceedings of the 2021 IEEE International Symposium on Information Theory (ISIT).

3.1 Introduction

To begin, we derive some results on the set erasure patterns correctable in grid-like topologies (see Definition 2.9). Recall, that in this setting constraints are imposed on the row and the column, i.e., each column and row are codewords of a column/row code. For the special case of PMDS codes, the characterization of this set is known [BHH13; GHJY14]. However, for grid-like topologies, which still represent a considerable simplification of the most general definition of a topology given in [GHJY14], it is a highly non-trivial problem to determine the erasure patterns that are correctable, i.e., the patterns that an MR code for the given topology must be able to correct. Interestingly, it is in general not sufficient for the column and row code to be MDS to correct all patterns in this set [GHK+17], which complicates the problem significantly. The seminal works of [GHJY14; GHK+17] initiated studies of the classification of these patterns for some restricted cases (see Table 3.1) and bounds on the required field size

[SRLS18; KMG19; KLR19]. In particular, [GHK+17] established a necessary condition for an erasure pattern to be correctable, which is then shown to also be sufficient for the case of one column constraint and no global constraints. Further, the sufficiency of this condition, referred to as *regularity*, is conjectured to also hold for the case of more column constraints.

3.1.1 Contributions and Outline

Section 3.2 recalls the definition of a regular erasure pattern and restates a conjecture made in [GHK+17] on their correctability. In Section 3.3 this conjecture is shown to be false by providing a counter-example of a regular erasure pattern that is not correctable in the corresponding topology. Specifically, we show that for *any* code of this topology, as in Definition 2.9, there exists at least one non-zero codeword that is zero in all nonerased positions of this pattern and therefore indistinguishable from the all-zero codeword. Then, we investigate the implications this counter-example has for the applicability of the conjecture to other topologies with larger parameters by establishing relations between the respective sets of correctable erasure patterns through shortening and puncturing arguments. By combining these arguments, we show that the conjecture does not hold for a large class of grid-like topologies.

Section 3.4 considers the class of tensor product codes, which are defined as the duals of codes for grid-like topologies without global parities. By exploiting a general connection between the erasure patterns correctable in a code and its dual code, the problems of MR codes for grid-like topologies and MR tensor-product codes are related.

The second main result of this chapter, provided in Section 3.5, is a generic method of adding global code constraints. To this end, the method used for the construction of PMDS codes in [RKSV13; CK16] and for the extension of binary codes in [GHJY14] is generalized. Given an MR code for a grid-like topology without global redundancy over an arbitrary field, an explicit construction that adds global redundancy is provided, at the expense of an increase in field size.

Finally, in Section 3.6 the chapter is concluded and some interesting open problems related to codes for grid-like topologies are briefly discussed.

3.2 Regular Erasure Patterns

In general, a necessary condition for the correctability of an erasure pattern is that the number of erasures[1] remaining in the code shortened in a given set of positions does not exceed the number of parity symbols. If the code has local redundancy, this condition can be refined to take into account that this local redundancy depends only

[1] Recall that codes for grid-like topologies are *scalar* codes and form a linear *vector* space (see Definition 2.9). Accordingly, despite treating the codewords as matrices here, we consider erasures of symbols, not columns.

Table 3.1: Properties of different topologies. Lines 1 to 4 collect the known results. Lines 5 and 6 summarizes the results on global redundancy of Section 3.5. Lines 7 and 8 give the results on the characterization of correctable erasure patterns presented in Section 3.3. The column "Conj." indicates whether the set of correctable erasure patterns matches Conjecture 3.1.

Topology	Correctable Patterns	Conj.	Reference
$T_{n_1\times n_2}(0, b_2, s)$	regular + any s	-	[BHH13; GHJY14]
$T_{n_1\times n_2}(1, 1, 0)$	regular	✓	[GHK+17]
$T_{n_1\times n_2}(1, b_2, 0)$	regular	✓	[GHK+17]
$T_{n_1\times n_2}(1, 1, s)$	regular + any s	-	[GHJY14]
$T_{n_1\times n_2}(1, b_2, s)$	regular + any s	-	[GHK+17] + Thm. 3.4
$T_{n_1\times n_2}(b_1, b_2, s)$	$\mathbb{E}_{n_1\times n_2}(b_1, b_2, 0)$ + any s	-	Thm. 3.5
$T_{5\times 5}(2, 2, 0)$	See Rem. 3.2	✗	Lem. 3.2
$T_{(\geq b_1+3)\times(\geq b_2+3)}(\geq 2, \geq 2, 0)$	unknown	✗	Thm. 3.2

Table 3.2: Known constructions of MR codes for different topologies. Lines 1 to 3 collect the known results. Lines 4 and 5 summarizes the results on global redundancy of Section 3.5, where the $\sim$ symbol implies that a code for the respective topology can be constructed given a code for the same topology with $s = 0$.

Topology	Construction	Reference
$T_{n_1\times n_2}(0, b_2, s)$	✓	[RKSV13; CK16]
$T_{n_1\times n_2}(1, 1, 0)$	✓	[GHK+17]
$T_{n_1\times n_2}(1, 1, s)$	✓	[GHJY14], [GHK+17] + Thm. 3.4
$T_{n_1\times n_2}(1, b_2, s)$	$\sim$	Thm. 3.4 + any $\mathcal{C} \in \mathbb{C}^{\mathsf{MR}}_{n_1\times n_2}(1, b_2, 0)$
$T_{n_1\times n_2}(b_1, b_2, s)$	$\sim$	Thm. 3.4 + any $\mathcal{C} \in \mathbb{C}^{\mathsf{MR}}_{n_1\times n_2}(b_1, b_2, 0)$

on small subset of positions. In [GHK+17] such a refinement was given for grid-like topologies and patterns that fulfill it are termed *regular erasure patterns.*

Definition 3.1 (Regular erasure pattern [GHK+17, Definition 3.1])**.** *Consider the topology $T_{n_1 \times n_2}(b_1, b_2, 0)$ and an erasure pattern $\mathcal{E} \subset [n_1] \times [n_2]$. We say that $\mathcal{E}$ is regular*[2] *if for all $\mathcal{U} \subseteq [n_1], |\mathcal{U}| = u \geq b_1$, and $\mathcal{V} \subseteq [n_2], |\mathcal{V}| = v \geq b_2$, we have*

$$|\mathcal{E} \cap (\mathcal{U} \times \mathcal{V})| \leq vb_1 + ub_2 - b_1 b_2 \ .$$

Remark 3.1. *Intuitively, the restriction of the erasure pattern to $\mathcal{U} \times \mathcal{V}$ can be interpreted as the shortening of the respective code in the positions outside of this grid. As will be shown in Section 3.3.2, this shortened code needs to be able to decode the remaining erasures for the pattern to be correctable. Definition 3.1 is a necessary (in general insufficient, see Section 3.3.1) condition for this to be possible, namely, that the number of erasures remaining in the shortened code does not exceed its redundancy.*

For example, consider shortening of the blue positions of the first row of the patterns in Fig. 3.1. Since the positions are part of an information set this reduces the code dimension by 3*. As the green parity position in the first row is a linear combination of the shortened symbols, it is fixed to be zero in the shortened code and can therefore also be shortened without further decreasing the dimension. This implies that despite the total number of shortened symbols (indicated in gray) being* 4*, the dimension of the code only decreases by* 3*. Hence, the shortened code is a* $[20, 9]$ *code, which trivially corrects at most* 11 *erasures. For the* regular *pattern* 10 *erasures remain in its restriction to* $[2, 6] \times [1, 4]$*, thereby* not *leading to a contradiction. It is easy to check that this pattern fulfills Definition 3.1 by applying this principle to all subgrids given by $\mathcal{U}$ and $\mathcal{V}$. On the other hand, in the* irregular *pattern* 12 *erasures remain in its restriction to* $[2, 6] \times [1, 4]$*, showing that this pattern cannot be corrected.*

It was shown in [GHK+17] that all erasure patterns that are *not* regular are *not* correctable in the topology $T_{n_1 \times n_2}(b_1, b_2, 0)$. On the other hand, for some cases (see Table 3.1) it is known that *all* regular patterns are correctable, which led to the following conjecture.

Conjecture 3.1 ([GHK+17, Conjecture 3.1])**.** *An erasure pattern $\mathcal{E}$ is correctable for the topology $T_{n_1 \times n_2}(b_1, b_2, 0)$ if and only if it is regular.*

In Section 3.3.1 we disprove this conjecture.

3.3 Negative Results on Correctable Erasure Patterns

This section presents negative results on the set of correctable erasure patterns for a given topology. First, we prove that a specific erasure pattern is never correctable in

[2] The original definition does not include the restriction $u \geq b_1$ and $v \geq b_2$. However, it is easy to check that this restriction is indeed necessary to exclude trivial cases.

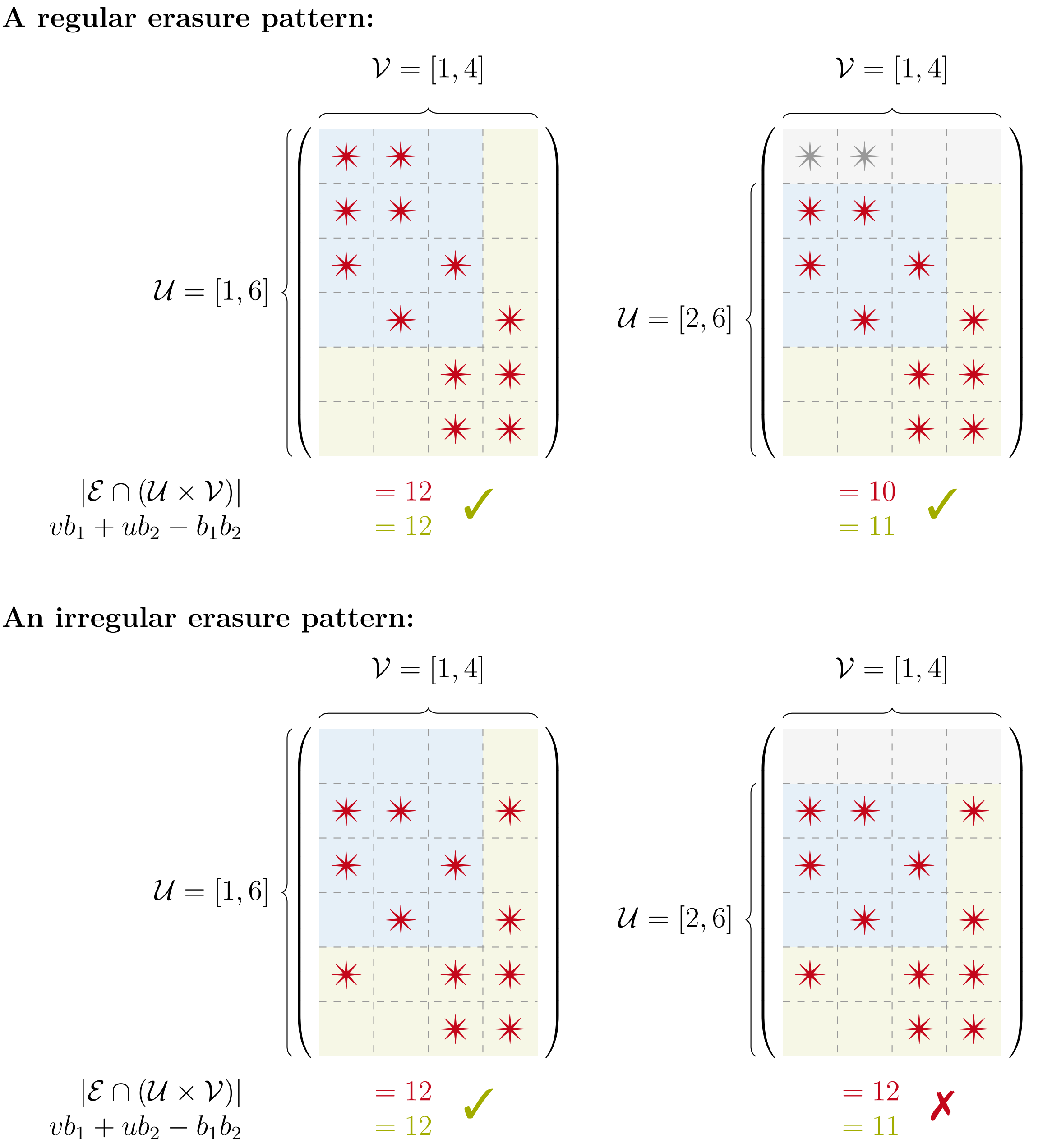


Figure 3.1: Illustration of the regularity property of an erasure pattern for the topology $T_{6\times 4}(2,1,0)$. The blue positions represent an information set. Note that such a subgrid is always an information set in an MR code for a grid-like topology [GHK+17]. The green positions represent the remaining (parity) symbols of the codeword. The value $|\mathcal{E} \cap (\mathcal{U} \times \mathcal{V})|$ is the number of erasures remaining in the subgrid spanned by $\mathcal{U}$ and $\mathcal{V}$ and the number of remaining parity symbols is given by $vb_1 + ub_2 - b_1 b_2$.

the topology $T_{5\times 5}(2,2,0)$, thereby disproving Conjecture 3.1 as given in [GHK+17]. Then, using this result, we provide generic methods of constructing incorrectable, regular erasure patterns for larger topologies.

3.3.1 Disproving a Conjecture on the Correctability of Regular Erasure Patterns

We begin with some simple observations on the relation between the generator matrix and the entries of low-weight codewords of an arbitrary linear $[5,3]$ code.

Lemma 3.1. *Consider a linear $[5,3]$ code $\mathcal{C}_{\mathsf{row}}$ with generator matrix*

$$\mathbf{G}_{\mathsf{row}} = \begin{pmatrix} 1 & 0 & 0 & p^{\mathsf{row}}_{1,1} & p^{\mathsf{row}}_{1,2} \\ 0 & 1 & 0 & p^{\mathsf{row}}_{2,1} & p^{\mathsf{row}}_{2,2} \\ 0 & 0 & 1 & p^{\mathsf{row}}_{3,1} & p^{\mathsf{row}}_{3,2} \end{pmatrix}$$

and a codeword

$$\begin{pmatrix} 1 & \alpha_2 & \alpha_3 & 0 & 0 \end{pmatrix} \in \mathcal{C}_{\mathsf{row}} \ . \tag{3.1}$$

Then

$$p^{\mathsf{row}}_{1,1} = -(\alpha_2 p^{\mathsf{row}}_{2,1} + \alpha_3 p^{\mathsf{row}}_{3,1}) \tag{3.2}$$
$$p^{\mathsf{row}}_{1,2} = -(\alpha_2 p^{\mathsf{row}}_{2,2} + \alpha_3 p^{\mathsf{row}}_{3,2}) \ . \tag{3.3}$$

Proof. The statement follows from

$$\begin{aligned} \begin{pmatrix} 1 & \alpha_2 & \alpha_3 & 0 & 0 \end{pmatrix} &= \begin{pmatrix} 1 & \alpha_2 & \alpha_3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & p^{\mathsf{row}}_{1,1} & p^{\mathsf{row}}_{1,2} \\ 0 & 1 & 0 & p^{\mathsf{row}}_{2,1} & p^{\mathsf{row}}_{2,2} \\ 0 & 0 & 1 & p^{\mathsf{row}}_{3,1} & p^{\mathsf{row}}_{3,2} \end{pmatrix} \\ \Rightarrow \ \begin{pmatrix} 1 & \alpha_2 & \alpha_3 \end{pmatrix} &\cdot \begin{pmatrix} p^{\mathsf{row}}_{1,1} & p^{\mathsf{row}}_{1,2} \\ p^{\mathsf{row}}_{2,1} & p^{\mathsf{row}}_{2,2} \\ p^{\mathsf{row}}_{3,1} & p^{\mathsf{row}}_{3,2} \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix} \ . \end{aligned}$$

□

Observe that each row of $\mathbf{G}_{\mathsf{row}}$ is a codeword of $\mathcal{C}_{\mathsf{row}}$ and, in particular,

$$\begin{pmatrix} 1 & 0 & 0 & p^{\mathsf{row}}_{1,1} & p^{\mathsf{row}}_{1,2} \end{pmatrix} \in \mathcal{C}_{\mathsf{row}} \ . \tag{3.4}$$

We define the same notions in similar notation for a $[5,3]$ code $\mathcal{C}_{\mathsf{col}}$, i.e.,

$$\begin{pmatrix}1 & \gamma_2 & \gamma_3 & 0 & 0\end{pmatrix} \in \mathcal{C}_{\mathsf{col}} \tag{3.5}$$

$$\begin{pmatrix}1 & 0 & 0 & p^{\mathsf{col}}_{1,1} & p^{\mathsf{col}}_{1,2}\end{pmatrix} \in \mathcal{C}_{\mathsf{col}} \tag{3.6}$$

$$p^{\mathsf{col}}_{1,1} = -(\gamma_2 p^{\mathsf{col}}_{2,1} + \gamma_3 p^{\mathsf{col}}_{3,1}) \tag{3.7}$$

$$p^{\mathsf{col}}_{1,2} = -(\gamma_2 p^{\mathsf{col}}_{2,2} + \gamma_3 p^{\mathsf{col}}_{3,2}) \ . \tag{3.8}$$

Note that if $\mathcal{C}_{\mathsf{row}}$ ($\mathcal{C}_{\mathsf{col}}$) is MDS we have $p^{\mathsf{row}}_{1,1}, p^{\mathsf{row}}_{1,2} \neq 0$ ($p^{\mathsf{col}}_{1,1}, p^{\mathsf{col}}_{1,2} \neq 0$) and the codewords given by Eqs. (3.1) and (3.4) (Eqs. (3.5) and (3.6)) are unique. This follows from the well-known facts that all symbols in the parity part of a systematic generator matrix of an MDS code must be non-zero and that every codeword of a linear code of a given support and weight $d_{\min}$ is unique up to scalar multiplication with an element of $\mathbb{F}^*$.

With these general relations established, we are now ready to prove that there exists a regular erasure pattern that is never correctable in the topology $T_{5\times 5}(2,2,0)$, by constructing a non-zero codeword that is zero in all non-erased positions.

Lemma 3.2. *The regular pattern $\mathcal{E}$ given by*

$$\mathbf{E} = \begin{pmatrix} 0 & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & 0 & 0 \\ \ast & \ast & \ast & 0 & 0 \\ \ast & 0 & 0 & \ast & \ast \\ \ast & 0 & 0 & \ast & \ast \end{pmatrix}$$

$$\mathcal{E} = \{(i,j) \mid \mathbf{E}[i,j] = \ast\}$$

is not correctable in $T_{5\times 5}(2,2,0)$.

Proof. Let $\mathcal{C} \in \mathbb{C}_{n_1\times n_2}(b_1, b_2, 0)$. We prove the statement by constructing a non-zero codeword of $\mathcal{C}$ that is zero in all non-erased positions. This implies that the code $\mathcal{C}$ cannot correct the pattern uniquely since there are at least two codewords that coincide on all non-erased positions — the constructed non-zero codeword and the all-zero codeword.

To construct such a codeword, we replace the $\ast$-symbols in $\mathbf{E}$ by elements of $\mathbb{F}$ (not all zero) in the following manner:

(a) Choose the element in position $(2,1)$ to be $\gamma_2 \in \mathbb{F}^*$. Note that, since the code is linear, we can always normalize one single non-zero position to be an arbitrary element of $\mathbb{F}^*$ and w.l.o.g. we assume position $(2,1)$ to be non-zero.

(b) Choose the second row to be the γ_2-multiple of Eq. (3.1).

(c) Choose the second and third column to be the corresponding multiples of Eq. (3.5).

(d) To obtain a multiple of Eq. (3.1) as the third row, set its first position to be γ_3.

(e) Encode the first row and column with $\mathbf{G}_{\mathsf{row}}$ and $\mathbf{G}_{\mathsf{col}}$, respectively.

(f) Replace the entries according to Eqs. (3.2), (3.3), (3.7) and (3.8).

(g) Fill in the fourth and fifth rows with the corresponding multiples of Eq. (3.4).

The individual steps are given by

$$\overset{(\mathsf{a})}{\Rightarrow}\begin{pmatrix} 0 & \ast & \ast & \ast & \ast \\ \gamma_2 & \ast & \ast & 0 & 0 \\ \ast & \ast & \ast & 0 & 0 \\ \ast & 0 & 0 & \ast & \ast \\ \ast & 0 & 0 & \ast & \ast \end{pmatrix} \overset{(\mathsf{b})}{\Rightarrow}\begin{pmatrix} 0 & \ast & \ast & \ast & \ast \\ \gamma_2 & \gamma_2\alpha_2 & \gamma_2\alpha_3 & 0 & 0 \\ \ast & \ast & \ast & 0 & 0 \\ \ast & 0 & 0 & \ast & \ast \\ \ast & 0 & 0 & \ast & \ast \end{pmatrix}$$

$$\overset{(\mathsf{c})}{\Rightarrow}\begin{pmatrix} 0 & \alpha_2 & \alpha_3 & \ast & \ast \\ \gamma_2 & \gamma_2\alpha_2 & \gamma_2\alpha_3 & 0 & 0 \\ \ast & \gamma_3\alpha_2 & \gamma_3\alpha_3 & 0 & 0 \\ \ast & 0 & 0 & \ast & \ast \\ \ast & 0 & 0 & \ast & \ast \end{pmatrix} \overset{(\mathsf{d})}{\Rightarrow}\begin{pmatrix} 0 & \alpha_2 & \alpha_3 & \ast & \ast \\ \gamma_2 & \gamma_2\alpha_2 & \gamma_2\alpha_3 & 0 & 0 \\ \gamma_3 & \gamma_3\alpha_2 & \gamma_3\alpha_3 & 0 & 0 \\ \ast & 0 & 0 & \ast & \ast \\ \ast & 0 & 0 & \ast & \ast \end{pmatrix}$$

$$\overset{(\mathsf{e})}{\Rightarrow}\begin{pmatrix} 0 & \alpha_2 & \alpha_3 & \alpha_2 p^{\mathsf{row}}_{2,1}+\alpha_3 p^{\mathsf{row}}_{3,1} & \alpha_2 p^{\mathsf{row}}_{2,2}+\alpha_3 p^{\mathsf{row}}_{3,2} \\ \gamma_2 & \gamma_2\alpha_2 & \gamma_2\alpha_3 & 0 & 0 \\ \gamma_3 & \gamma_3\alpha_2 & \gamma_3\alpha_3 & 0 & 0 \\ \gamma_2 p^{\mathsf{col}}_{2,1}+\gamma_3 p^{\mathsf{col}}_{3,1} & 0 & 0 & \ast & \ast \\ \gamma_2 p^{\mathsf{col}}_{2,2}+\gamma_3 p^{\mathsf{col}}_{3,2} & 0 & 0 & \ast & \ast \end{pmatrix}$$

$$\overset{(\mathsf{f})}{\Rightarrow}\begin{pmatrix} 0 & \alpha_2 & \alpha_3 & -p^{\mathsf{row}}_{1,1} & -p^{\mathsf{row}}_{1,2} \\ \gamma_2 & \gamma_2\alpha_2 & \gamma_2\alpha_3 & 0 & 0 \\ \gamma_3 & \gamma_3\alpha_2 & \gamma_3\alpha_3 & 0 & 0 \\ -p^{\mathsf{col}}_{1,1} & 0 & 0 & \ast & \ast \\ -p^{\mathsf{col}}_{1,2} & 0 & 0 & \ast & \ast \end{pmatrix} \overset{(\mathsf{g})}{\Rightarrow}\begin{pmatrix} 0 & \alpha_2 & \alpha_3 & -p^{\mathsf{row}}_{1,1} & -p^{\mathsf{row}}_{1,2} \\ \gamma_2 & \gamma_2\alpha_2 & \gamma_2\alpha_3 & 0 & 0 \\ \gamma_3 & \gamma_3\alpha_2 & \gamma_3\alpha_3 & 0 & 0 \\ -p^{\mathsf{col}}_{1,1} & 0 & 0 & -p^{\mathsf{col}}_{1,1}p^{\mathsf{row}}_{1,1} & -p^{\mathsf{col}}_{1,1}p^{\mathsf{row}}_{1,2} \\ -p^{\mathsf{col}}_{1,2} & 0 & 0 & -p^{\mathsf{col}}_{1,2}p^{\mathsf{row}}_{1,1} & -p^{\mathsf{col}}_{1,2}p^{\mathsf{row}}_{1,2} \end{pmatrix}.$$

It is easy to see that the fourth and fifth columns are also multiples of Eq. (3.6). Hence, this array contains a codeword of the row code in every row and a codeword of the column code in every column. It is therefore a valid codeword of any code for $T_{5\times 5}(2,2,0)$. As the used properties hold for any linear code we conclude that this pattern is never correctable.

Note that if both $\mathcal{C}_{\mathsf{row}}$ and $\mathcal{C}_{\mathsf{col}}$ are MDS, the first step, i.e., choosing the element in position $(2,1)$, determines the whole matrix, as each of the subsequent steps is unique in this case. □

Remark 3.2. *The proof of Lemma 3.2 can be carried out for any column/row permutation of* $\mathbf{E}$*, i.e., none of these permutations is correctable in* $T_{5\times5}(2,2,0)$*. Further, computer search shows that these* 450 *permutations are* exactly *the regular patterns that are not correctable in this topology.*

As the erasure pattern given in Lemma 3.2 is regular but not correctable in a grid-like topology, we come to the following conclusion.

Theorem 3.1. *Conjecture 3.1 (cf. [GHK⁺17, Conjecture 3.1]) is false.*

Proof. Follows immediately from Lemma 3.2. □

3.3.2 Implications of Incorrectable Regular Patterns for Larger Topologies

This section is dedicated to showing the implications the incorrectable erasure pattern for the topology $T_{5\times5}(2,2,0)$, given in Lemma 3.2, has on the set of correctable erasure for almost all topologies. To this end, we employ arguments based on *shortening* and *puncturing*, as in Definition 2.1. We collect some well-known/basic properties of linear codes in the following proposition.

Proposition 3.1. *Let* $\mathbf{G}$ *and* $\mathbf{H}$ *denote a generator and parity-check matrix of an* $[n,k]$ *code* $\mathcal{C}$*, respectively. Then,*

1. $\mathbf{H}|_{[n]\setminus\mathcal{I}}$ *is a parity-check matrix of* $\mathrm{short}_{\mathcal{I}}(\mathcal{C})$ *and* $\mathbf{G}|_{[n]\setminus\mathcal{I}}$ *is a generator matrix of* $\mathrm{punct}_{\mathcal{I}}(\mathcal{C})$*.*

2. *an erasure pattern* $\mathcal{E}\subset[n]$ *is correctable if and only if it fulfills the equivalent conditions*

$$\dim(\mathrm{short}_{[n]\setminus\mathcal{E}}(\mathcal{C}))=0 \quad\Longleftrightarrow\quad \dim(\mathrm{punct}_{\mathcal{E}}(\mathcal{C}))=k\ .$$

3. *an erasure pattern* $\mathcal{E}\subset[n]$ *is correctable only if the pattern* $\mathcal{E}\setminus\mathcal{I}$ *is correctable in the code* $\mathrm{short}_{\mathcal{I}}(\mathcal{C})$ *for any* $\mathcal{I}\subset[n]$*.*

4. *an erasure pattern* $\mathcal{E}$ *is correctable only if the pattern* $\mathcal{E}\setminus\mathcal{I}$ *is correctable in the code* $\mathrm{punct}_{\mathcal{I}}(\mathcal{C})$ *for any* $\mathcal{I}\subseteq[\mathcal{E}]$*.*

Proof. Properties 1 and 2 are well-known.

Proof of 3): By property 2) we know that $\mathcal{E}\setminus\mathcal{I}$ is correctable in $\mathrm{short}_{\mathcal{I}}(\mathcal{C})$ if and only if $\dim(\mathrm{short}_{[n]\setminus(\mathcal{E}\setminus\mathcal{I})}(\mathrm{short}_{\mathcal{I}}(\mathcal{C})))=0$. Now observe that

$$\begin{aligned}\mathrm{short}_{[n]\setminus(\mathcal{E}\setminus\mathcal{I})}(\mathrm{short}_{\mathcal{I}}(\mathcal{C})) &= \mathrm{short}_{([n]\setminus(\mathcal{E}\setminus\mathcal{I}))\cup\mathcal{I}}(\mathcal{C})\\ &= \mathrm{short}_{([n]\setminus\mathcal{E})\cup\mathcal{I}}(\mathcal{C})\\ &= \mathrm{short}_{\mathcal{I}\setminus([n]\setminus\mathcal{E})}(\mathrm{short}_{([n]\setminus\mathcal{E})}(\mathcal{C}))\ .\end{aligned}$$

As shortening does not increase the code dimension we have

$$\dim(\mathrm{short}_{[n]\setminus(\mathcal{E}\setminus\mathcal{I})}(\mathrm{short}_{\mathcal{I}}(\mathcal{C}))) \leq \dim(\mathrm{short}_{[n]\setminus\mathcal{E}}(\mathcal{C})) \ .$$

The statement follows from the observation that $\mathcal{E}$ is correctable in $\mathcal{C}$ if and only if $\dim(\mathrm{short}_{[n]\setminus\mathcal{E}}(\mathcal{C})) = 0$.

Proof of 4): By Definition 2.1 we have

$$\begin{aligned} \dim(\mathrm{punct}_{\mathcal{E}}(\mathcal{C})) \leq \dim(\mathrm{punct}_{\mathcal{I}}(\mathcal{C})) &\leq \dim(\mathcal{C}) \\ \mathrm{punct}_{\mathcal{E}\setminus\mathcal{I}}(\mathrm{punct}_{\mathcal{I}}(\mathcal{C})) &= \mathrm{punct}_{\mathcal{E}}(\mathcal{C}) \ . \end{aligned}$$

If $\mathcal{E}$ is correctable in $\mathcal{C}$ it holds that $\dim(\mathcal{C}) = \dim(\mathrm{punct}_{\mathcal{E}}(\mathcal{C}))$ by property 2), which implies

$$\begin{aligned} \dim(\mathrm{punct}_{\mathcal{I}}(\mathcal{C})) &= \dim(\mathrm{punct}_{\mathcal{E}}(\mathcal{C})) \\ &= \dim(\mathrm{punct}_{\mathcal{E}\setminus\mathcal{I}}(\mathrm{punct}_{\mathcal{I}}(\mathcal{C}))) \ . \end{aligned}$$

By property 2) this is a necessary and sufficient condition for the pattern $\mathcal{E} \setminus \mathcal{I}$ to be correctable in the code $\mathrm{punct}_{\mathcal{I}}(\mathcal{C})$.

□

The properties of Proposition 3.1 provide two ways of relating an erasure pattern to codes with larger parameters, which correspond to the opposite operations of shortening and puncturing[3].

We begin by showing a general relation between the codes for grid-like topologies for different parameters. The following lemma shows that restricting a code for a grid-like topology to positions of a subgrid, either by shortening or puncturing the other positions, gives a code for a grid-like topology of smaller parameters.

Lemma 3.3. *Let $\mathcal{C} \in \mathbb{C}_{n_1 \times n_2}(b_1, b_2, 0)$. Denote by $\mathbb{I}_{\mathsf{col}}$ and $\mathbb{I}_{\mathsf{row}}$ the sets of information sets of the respective column and row code. Then for any $\mathcal{U} \subseteq [n_1]$ such that $([n_1]\setminus\mathcal{U}) \subseteq \mathcal{I}$ for some $\mathcal{I} \in \mathbb{I}_{\mathsf{col}}$ and $\mathcal{V} \subseteq [n_2]$ such that $([n_2] \setminus \mathcal{V}) \subseteq \mathcal{I}$ for some $\mathcal{I} \in \mathbb{I}_{\mathsf{row}}$ we have*

$$\mathrm{short}_{([n_1]\times[n_2])\setminus(\mathcal{U}\times\mathcal{V})}(\mathcal{C}) \in \mathbb{C}_{|\mathcal{U}|\times|\mathcal{V}|}(b_1, b_2, 0) \ .$$

Further, for any $\mathcal{U} \subseteq \mathcal{I}$ for some $\mathcal{I} \in \mathbb{I}_{\mathsf{col}}$ and $\mathcal{V} \subseteq \mathcal{I}$ for some $\mathcal{I} \in \mathbb{I}_{\mathsf{row}}$ we have

$$\mathrm{punct}_{([n_1]\times[n_2])\setminus(\mathcal{U}\times\mathcal{V})}(\mathcal{C}) \in \mathbb{C}_{|\mathcal{U}|\times|\mathcal{V}|}(b_1 - (n_1 - u), b_2 - (n_2 - v), 0) \ .$$

Proof. As $h = 0$, by Definition 2.9, the generator matrix of $\mathcal{C}$ is given by $\mathbf{G}_{\mathsf{col}} \otimes \mathbf{G}_{\mathsf{row}}$. Let $\mathbf{G}_{\mathsf{col}}^{\mathrm{short}_{\mathcal{J}}}, \mathbf{G}_{\mathsf{row}}^{\mathrm{short}_{\mathcal{J}}}$ and $\mathbf{G}_{\mathsf{col}}^{\mathrm{punct}_{\mathcal{J}}}, \mathbf{G}_{\mathsf{row}}^{\mathrm{punct}_{\mathcal{J}}}$ denote the generator matrices of the column/row code shortened/punctured in the positions indexed by $\mathcal{J}$. It follows directly

[3] These operations are sometimes referred to as *lengthening* and *extending*.

from the definition of the Kronecker product (see Section 2.1.1) that

$$\begin{aligned}\mathrm{short}_{([n_1]\times[n_2])\setminus(\mathcal{U}\times\mathcal{V})}(\mathcal{C}) &= \left\langle \mathbf{G}_{\mathsf{col}}^{\mathrm{short}_{\mathcal{U}}} \otimes \mathbf{G}_{\mathsf{row}}^{\mathrm{short}_{\mathcal{V}}} \right\rangle \\ \mathrm{punct}_{([n_1]\times[n_2])\setminus(\mathcal{U}\times\mathcal{V})}(\mathcal{C}) &= \left\langle \mathbf{G}_{\mathsf{col}}^{\mathrm{punct}_{\mathcal{U}}} \otimes \mathbf{G}_{\mathsf{row}}^{\mathrm{punct}_{\mathcal{V}}} \right\rangle \ .\end{aligned}$$

The statement follows from observing that shortening a position of an information set of an $[n_1, n_1 - b_1]$ code gives an $[n_1 - 1, n_1 - b_1 - 1]$ code. Similarly, puncturing a position in the complement of an information set gives an $[n_1 - 1, n_1 - b_1]$ code. □

Using this relation between codes for different grid-like topologies, we can now establish a connection between the set of correctable erasure patterns in topologies with longer row and/or column codes.

Lemma 3.4. *If there exists a regular erasure pattern $\mathcal{E}$ that is not correctable in $T_{n_1\times n_2}(b_1, b_2, 0)$ then there exists a regular erasure pattern $\mathcal{E}'$ that is not correctable in $T_{n_1+\delta\times n_2+\gamma}(b_1, b_2, 0)$ for any $\delta, \gamma \geq 0$.*

Proof. We show that $\mathcal{E}' = \mathcal{E}$ is not correctable[4]. An illustration of such a pattern is provided in Fig. 3.2a. As $\mathcal{E}'$ does not contain additional erasures compared to $\mathcal{E}$ and the restriction of Definition 3.1 does not depend on n_1 or n_2, we conclude that the pattern is regular, i.e., the regularity of a pattern in $T_{n_1\times n_2}(b_1, b_2, 0)$ directly implies the regularity of the same pattern in $T_{n_1+\delta\times n_2+\delta}(b_1, b_2, 0)$ for any $\delta, \gamma \geq 0$.

By Lemma 3.3 we have $\mathrm{short}_{([n_1+\delta]\times[n_2+\gamma])\setminus([n_1]\times[n_2])}(\mathcal{C}') \in \mathbb{C}_{n_1\times n_2}(b_1, b_2, 0)$ for any code $\mathcal{C}' \in \mathbb{C}_{n_1+\delta\times n_2+\gamma}(b_1, b_2, 0)$. By assumption, the restricted pattern $\mathcal{E}' \cap ([n_1] \times [n_2]) = \mathcal{E}$ is not correctable in $T_{n_1\times n_2}(b_1, b_2, 0)$ and the statement follows from 3) in Proposition 3.1. □

The following shows a small example to illustrate the incorrectable patterns implied by Lemma 3.4.

Example 3.1. *Consider the erasure pattern $\mathcal{E}$ as in Lemma 3.2, which is* not *correctable in the topology $T_{5\times 5}(2, 2, 0)$. Then, by Lemma 3.4, the erasure pattern $\mathcal{E}' = \mathcal{E}$*

[4] Even if $\mathcal{E}$ is maximal for $T_{n_1\times n_2}(b_1, b_2, 0)$, it is not maximal for the topology $T_{n_1+\delta\times n_2+\gamma}(b_1, b_2, 0)$ for any $\delta, \gamma > 0$. However, it is also easy to construct a non-correctable maximal pattern by having each additional column/row contain exactly b_1 or b_2 erasures, respectively.

is not correctable in the topology $T_{7\times 6}(2,2,0)$*. Illustrating* $\mathcal{E}$ *for this topology gives*

$$\mathbf{E}' = \begin{pmatrix} 0 & \ast & \ast & \ast & \ast & 0 \\ \ast & \ast & \ast & 0 & 0 & 0 \\ \ast & \ast & \ast & 0 & 0 & 0 \\ \ast & 0 & 0 & \ast & \ast & 0 \\ \ast & 0 & 0 & \ast & \ast & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathcal{E}' = \{(i,j) \mid \mathbf{E}'[i,j] = \ast\} \ .$$

Lemma 3.4 is based on increasing the length of the column and row codes while keeping the respective number of redundancy symbols constant. In contrast, the following lemma increases the length of the column and row code while holding the dimension of each code constant.

Lemma 3.5. *If there exists a regular erasure pattern* $\mathcal{E}$ *that is not correctable in* $T_{n_1\times n_2}(b_1,b_2,0)$ *then there exists a regular erasure pattern* $\mathcal{E}'$ *that is not correctable in* $T_{n_1+\delta\times n_2+\gamma}(b_1+\delta, b_2+\gamma, 0)$ *for any* $\delta,\gamma \geq 0$*.*

Proof. Denote by $\mathcal{C}'_{\mathsf{col}}$ and $\mathcal{C}'_{\mathsf{row}}$ the column and row code of a code $\mathcal{C}' \in \mathbb{C}_{n_1+\delta\times n_2+\gamma}(b_1+\delta, b_2+\gamma, 0)$. Without loss of generality assume that the first n_1-b_1 and n_2-b_2 positions are (a subset of) an information set of the code $\mathcal{C}'_{\mathsf{col}}$ and $\mathcal{C}'_{\mathsf{row}}$. Let $\mathcal{E}'$ be the erasure pattern obtained by adding δ rows and γ columns of erasures to $\mathcal{E}$, i.e.,

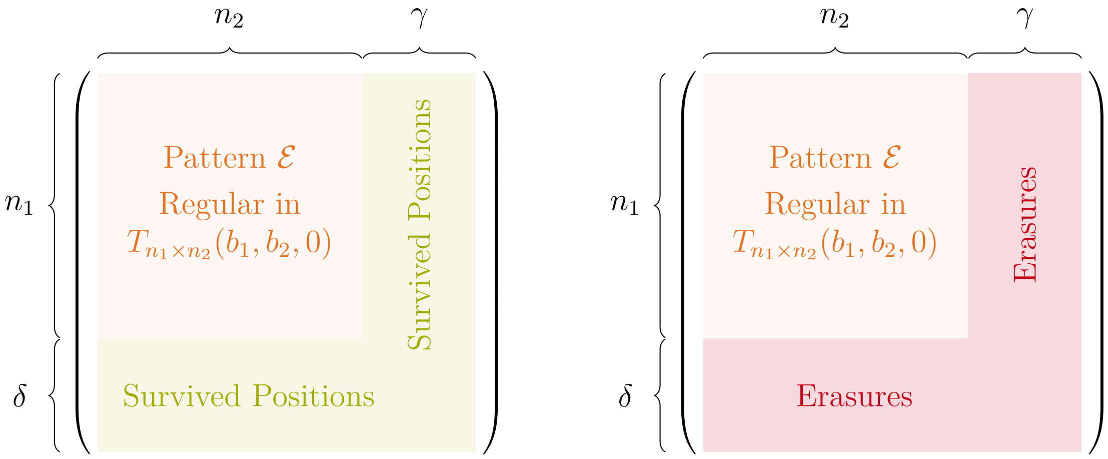
$$\mathcal{E}' = \mathcal{E} \cup \Big(([n_1+\delta] \times [n_2+\gamma]) \setminus ([n_1] \times [n_2]) \Big) \ .$$

For an illustration of this pattern, see Fig. 3.2b. We show that this pattern is regular for the topology $T_{n_1+\delta\times n_2+\gamma}(b_1+\delta, b_2+\gamma, 0)$, i.e., fulfills Definition 3.1 for any $\mathcal{U}' \subseteq [n_1+\delta], |\mathcal{U}'| = u \geq b_1+\delta$ and $\mathcal{V}' \subseteq [n_2+\gamma], |\mathcal{V}'| = v \geq b_2+\gamma$. As the δ/γ additional rows/columns consist only of erasures it suffices to show that the subsets of rows/columns with $[n_1+1, n_1+\delta] \subset \mathcal{U}$ and $[n_2+1, n_2+\gamma] \subset \mathcal{V}$ fulfill the condition. Define $\mathcal{U} = \mathcal{U}' \cap [n_1]$ and $\mathcal{V} = \mathcal{V}' \cap [n_2]$ and observe that $\mathcal{E}'$ can be partitioned into two disjoint subsets $\mathcal{E}' \cap (\mathcal{U} \times \mathcal{V})$ and $\mathcal{E}' \cap ((\mathcal{U}' \times \mathcal{V}') \setminus (\mathcal{U} \times \mathcal{V}))$. For the former we have

$$|\mathcal{E}' \cap (\mathcal{U} \times \mathcal{V})| \leq (v-\gamma)b_1 + (u-\delta)b_2 - b_1 b_2$$

because $|\mathcal{E}' \cap (\mathcal{U} \times \mathcal{V})| = \mathcal{E}$ is regular by assumption. The second subset are exactly the δ/γ additional rows/columns, which consist only of erasures by definition of $\mathcal{E}'$, and therefore

$$|\mathcal{E}' \cap ((\mathcal{U}' \times \mathcal{V}') \setminus (\mathcal{U} \times \mathcal{V}))| = v\delta + u\gamma - \delta\gamma \ ,$$



(a) Erasure pattern $\mathcal{E}'$ as in Lemma 3.4, which is regular in the topology $T_{n_1+\delta \times n_2+\gamma}(b_1, b_2, 0)$.

(b) Erasure pattern $\mathcal{E}'$ as in Lemma 3.5, which is regular in the topology $T_{n_1+\delta \times n_2+\gamma}(b_1 + \delta, b_2 + \gamma, 0)$.

Figure 3.2: Illustration of the extension of regular erasure patterns to larger topologies.

Hence, the cardinality of $\mathcal{E}' \cap (\mathcal{U}' \times \mathcal{V}')$ is bounded by

$$\begin{aligned} |\mathcal{E}' \cap (\mathcal{U}' \times \mathcal{V}')| &= |\mathcal{E}' \cap (\mathcal{U} \times \mathcal{V})| + |\mathcal{E}' \cap ((\mathcal{U}' \times \mathcal{V}') \setminus (\mathcal{U} \times \mathcal{V}))| \\ &\leq ((v - \gamma)b_1 + (u - \delta)b_2 - b_1 b_2) + (v\delta + u\gamma - \delta\gamma) \\ &= v(b_1 + \delta) + u(b_2 + \gamma) - (b_1 + \delta)(b_2 + \gamma) \end{aligned}$$

and it follows that the pattern is regular for the topology $T_{n_1+\delta \times n_2+\gamma}(b_1 + \delta, b_2 + \gamma, 0)$.

By Lemma 3.3 we have $\mathrm{punct}_{([n_1+\delta]\times[n_2+\gamma])\setminus([n_1]\times[n_2])}(\mathcal{C}') \in \mathbb{C}_{n_1 \times n_2}(b_1, b_2, 0)$. By definition, the pattern $\mathcal{E} = \mathcal{E}' \setminus \Big(([n_1 + \delta] \times [n_2 + \gamma]) \setminus ([n_1] \times [n_2])\Big)$ is not correctable in this topology and the statement follows by 4) in Proposition 3.1. □

The following shows a small example of an incorrectable erasure pattern, based on the pattern of Lemma 3.2.

Example 3.2. *Consider the erasure pattern $\mathcal{E}$ as in Lemma 3.2, which is* not *correctable in the topology $T_{5\times5}(2, 2, 0)$. Then, by Lemma 3.4, an erasure pattern that is not correctable in $T_{7\times6}(4, 3, 0)$ can be obtained by appending two rows and one column* *of erasures. Therefore, the pattern*

$$\mathbf{E}' = \begin{pmatrix} 0 & \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & 0 & 0 & \ast \\ \ast & \ast & \ast & 0 & 0 & \ast \\ \ast & 0 & 0 & \ast & \ast & \ast \\ \ast & 0 & 0 & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast & \ast \end{pmatrix}$$

$$\mathcal{E}' = \{(i,j) \mid \mathbf{E}'[i,j] = \ast\}$$

is not correctable in the topology $T_{7\times 6}(4,3,0)$.

As the approaches of Lemmas 3.4 and 3.5 can be combined arbitrarily to obtain erasure patterns that are incorrectable in larger topologies, we arrive at the following main statement.

Theorem 3.2. *Let* $b_1, b_2 \geq 2$. *For any* $n_1 \geq b_1 + 3$ *and* $n_2 \geq b_2 + 3$ *there exist regular erasure patterns that are not correctable in the topology* $T_{n_1 \times n_2}(b_1, b_2, 0)$.

Proof. Follows from applying Lemmas 3.4 and 3.5 to the topology $T_{5\times 5}(2,2,0)$, for which an incorrectable regular erasure pattern exists by Lemma 3.2. □

Observe that with the presented results, the only topologies $T_{n_1\times n_2}(b_1, b_2, 0)$ for which it is not known whether Conjecture 3.1 applies are the cases $b_1, b_2 \geq 2$ and $n_1 \leq b_1 + 2$ or $n_2 \leq b_2 + 2$, i.e., settings where the column and/or the row code are of very low rate.

3.4 Connection between Product and Tensor-Product Codes

Before moving on to codes for grid-like topologies with global parities, we provide a connection between codes for grid-like topologies with $s = 0$ and another class of codes with application to storage, so called tensor-product (TP) codes [Wol65]. These codes are defined as the duals of product codes and are interesting due to their small storage overhead and good protection against some types of correlated erasures and errors. However, there is no general classification of erasure patterns correctable by a TP code. In this section, a connection between this problem and the corresponding problem for grid-like topologies is established by showing how the maximal erasure patterns of codes and their dual codes are connected. This implies that for certain TP codes, which are duals of MR codes of a grid-like topology with $s = 0$, their correctable erasure patterns can be exactly characterized by relying on the results summarized in Table 3.1.

Definition 3.2 (Tensor-Product Code)**.** *Consider an $[n_1, b_1]$ code $\mathcal{C}_{\mathsf{col}}$ and an $[n_2, b_2]$ code $\mathcal{C}_{\mathsf{row}}$. Then the tensor-product code $\mathsf{TP}(\mathcal{C}_{\mathsf{col}}, \mathcal{C}_{\mathsf{row}})$ is the $[n_1n_2, n_1n_2-(n_1-b_1)(n_2-b_2)]$ code defined as*

$$\mathsf{TP}(\mathcal{C}_{\mathsf{col}}, \mathcal{C}_{\mathsf{row}}) = \langle \mathbf{H}_{\mathsf{col}} \otimes \mathbf{H}_{\mathsf{row}} \rangle^{\perp} \ ,$$

where $\mathbf{H}_{\mathsf{col}}$ and $\mathbf{H}_{\mathsf{row}}$ denote parity-check matrices of the codes $\mathcal{C}_{\mathsf{col}}$ and $\mathcal{C}_{\mathsf{row}}$, respectively.

We begin by formally establishing the connection between TP codes and codes for grid-like topologies.

Lemma 3.6. *Let $\mathcal{C}_1$ be an $[n_1, b_1]$ code and $\mathcal{C}_2$ be an $[n_2, b_2]$ code. Then*

$$\mathsf{TP}(\mathcal{C}_1, \mathcal{C}_2)^{\perp} \in \mathbb{C}_{n_1 \times n_2}(b_1, b_2, 0) \ .$$

Proof. Denote by $\mathbf{H}_{\mathsf{col}}$ and $\mathbf{H}_{\mathsf{row}}$ parity-check matrices of the codes $\mathcal{C}_{\mathsf{col}}$ and $\mathcal{C}_{\mathsf{row}}$, respectively. We have

$$\mathsf{TP}(\mathcal{C}_{\mathsf{col}}, \mathcal{C}_{\mathsf{row}})^{\perp} = \langle \mathbf{H}_{\mathsf{col}} \otimes \mathbf{H}_{\mathsf{row}} \rangle \ .$$

The statement follows from interpreting the matrices $\mathbf{H}_{\mathsf{col}}$ and $\mathbf{H}_{\mathsf{row}}$ as generator matrices of $[n_1, n_1 - b_1]$ and $[n_2, n_2 - b_2]$ codes, respectively. □

Now, to apply the results on correctable erasures patterns given in Table 3.1 to TP codes, we only need to relate the set of erasure patterns correctable in a given code to those correctable in its dual code.

Lemma 3.7. *Consider any $[n, k]$ code $\mathcal{C}$ and denote by $\mathbb{E}$ the set of erasure patterns that are correctable in this code. Then the set $\mathbb{E}^{\perp}$ of erasure patterns correctable in the dual code is given by*

$$\mathbb{E}^{\perp} = \{\mathcal{E} \mid \mathcal{E} \subseteq [n] \setminus \mathcal{E}_{\max}, \mathcal{E}_{\max} \in \mathbb{E}, |\mathcal{E}_{\max}| = n - k\} \ .$$

Proof. It is well-known that the information sets of the dual of a code are exactly given by the complements of the information sets of the code and that an erasure pattern is correctable if and only if the surviving positions contain an information set (see Proposition 3.1). Hence, the set of patterns correctable by the dual code is given by

$$\begin{aligned}
&\{\mathcal{E} \mid \mathcal{E} \subseteq [n] \setminus \mathcal{I}, \mathcal{I} \text{ is an information set of } \mathcal{C}^{\perp}\} \\
&\qquad = \{\mathcal{E} \mid \mathcal{E} \subseteq \mathcal{I}, \mathcal{I} \text{ is an information set of } \mathcal{C}\} \\
&\qquad = \{\mathcal{E} \mid \mathcal{E} \subseteq \mathcal{I}, [n] \setminus \mathcal{I} \in \mathbb{E}, |\mathcal{I}| = k\} \\
&\qquad = \{\mathcal{E} \mid \mathcal{E} \subseteq [n] \setminus \mathcal{E}_{\max}, \mathcal{E}_{\max} \in \mathbb{E}, |\mathcal{E}_{\max}| = n - k\} \ .
\end{aligned}$$

□

It follows that there is a deterministic relation between the sets of erasure patterns correctable in a code and its dual.

Corollary 3.1. *The set of maximal erasure patterns correctable in a code $\mathcal{C}$ uniquely determines the set of erasure patterns correctable in the dual code $\mathcal{C}^{\perp}$.*

For many classes of codes this result has limited practical value, as the set of all maximal erasure patterns is commonly unknown. While the minimum distance of a code provides a guarantee on the erasure patterns correctable in this code, this only leads to a superset of the patterns correctable in the dual code. The interesting point about applying this results to duals of MR codes is that for some topologies the set of maximal patterns can be completely determined (see Table 3.1), leading to an exact characterization of the patterns correctable in the dual code.

Theorem 3.3. *Let $\mathcal{C}_{\mathsf{col}}$ be an $[n_1, b_1]$ code and $\mathcal{C}_{\mathsf{row}}$ be an $[n_2, b_2]$ code. The set of maximal erasure patterns correctable by the TP code $\mathsf{TP}(\mathcal{C}_{\mathsf{col}}, \mathcal{C}_{\mathsf{row}})$ is a subset of*

$$\{([n_1] \times [n_2]) \setminus \mathcal{E} \mid \mathcal{E} \in \mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, 0)\} \ .$$

Moreover, if $\mathsf{TP}(\mathcal{C}_{\mathsf{col}}, \mathcal{C}_{\mathsf{row}})^{\perp} \in \mathbb{C}^{\mathsf{MR}}_{n_1 \times n_2}(b_1, b_2, 0)$ then this is exactly *the set of erasure patterns correctable in $\mathsf{TP}(\mathcal{C}_{\mathsf{col}}, \mathcal{C}_{\mathsf{row}})$.*

Applying Theorem 3.3 to the cases where $\mathbb{E}_{n_1 \times n_2}(b_1, b_2, 0)$ is known (see Table 3.1) establishes the correctable erasure patterns in the corresponding TP codes.

3.5 Global Redundancy $s > 0$

For some topologies with $s = 0$ the set of correctable erasure patterns is fully characterized (as the regular patterns) and, for some cases, explicit constructions are known (see Tables 3.1 and 3.2). For $s > 0$ only the special cases of $b_1 = 0$ (corresponding to PMDS codes, see Section 2.3.1) and $b_1 = b_2 = s = 1$ are known (see Table 3.1). In this section, we characterize the set of correctable erasure patterns $\mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, s)$ as a function of $\mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, 0)$, for any n_1, n_2, b_1, b_2, and s. A similar result for the extension of codes defined by a binary parity-check matrix, i.e., where $\mathbf{H}_{\mathsf{local}}$ is in $\mathbb{F}_2$, has been derived in [GHJY14, Section V.A].

3.5.1 Construction

We generalize the Gabidulin-based code construction of [RKSV13] for $b_1 = 0$ to any number of column parities. These codes have been shown to be MR for the topology $T_{n_1 \times n_2}(0, b_2, s)$, i.e., to be PMDS codes, in [CK16]. We show that by a similar two-stage encoding procedure we can "add global redundancy" to codes for any grid-like topology. Specifically, we give a general construction that, given a code $\mathcal{C}_{\mathsf{out}} \in \mathbb{C}^{\mathsf{MR}}_{n_1 \times n_2}(b_1, b_2, 0)$, returns a code $\mathcal{C} \in \mathbb{C}^{\mathsf{MR}}_{n_1 \times n_2}(b_1, b_2, s)$.

Lemma 3.8. *Let $\mathbb{I}_{\mathsf{out}}$ be the set of information sets of $\mathcal{C}_{\mathsf{out}}[n_{\mathsf{out}}, k_{\mathsf{out}}]_q$ and $\mathcal{C}_{\mathsf{in}}$ be a $\mathsf{Gab}(k_{\mathsf{out}}, s, \boldsymbol{\beta})_{q^{k_{\mathsf{out}}}}$ code as in Definition 2.6. Then the code $\langle \mathbf{G}_{\mathsf{in}} \cdot \mathbf{G}_{\mathsf{out}} \rangle |_{\mathcal{I}}$ is a $\mathsf{Gab}(k_{\mathsf{out}}, s, \boldsymbol{\beta}')_{q^{k_{\mathsf{out}}}}$ code for any $\mathcal{I} \in \mathbb{I}_{\mathsf{out}}$.*

Proof. By definition of an information set, the matrix $\mathbf{G}_{\mathsf{out}}|_{\mathcal{I}}$ is a full-rank matrix over $\mathbb{F}_q$. The statement follows from observing that $\langle \mathbf{G}_{\mathsf{in}} \cdot \mathbf{G}_{\mathsf{out}} \rangle |_{\mathcal{I}} = \langle \mathbf{G}_{\mathsf{in}} \cdot (\mathbf{G}_{\mathsf{out}}|_{\mathcal{I}}) \rangle$ and applying Lemma 2.2. □

We now provide a characterization of the set of erasure patterns correctable by this construction when the code $\mathcal{C}_{\mathsf{out}}$ is an MR code for a grid-like topology.

Theorem 3.4. *Let $\mathcal{C}_{\mathsf{out}} \in \mathbb{C}^{\mathsf{MR}}_{n_1 \times n_2}(b_1, b_2, 0)$ be an $[n_{\mathsf{out}} = n_1 n_2, k_{\mathsf{out}} = (n_1 - b_1)(n_2 - b_2)]$ code and*

$$\mathcal{C}_{\mathsf{in}} := \mathsf{Gab}\Big((n_1 - b_1)(n_2 - b_2), s, \boldsymbol{\beta}\Big)_{q^s} ,$$

with $s \geq (n_1 - b_1)(n_2 - b_2)$. Then the code $\langle \mathbf{G}_{\mathsf{in}} \cdot \mathbf{G}_{\mathsf{out}} \rangle$ corrects all erasure patterns in

$$\{\mathcal{E}' \cup \mathcal{I} \ \mid \ \mathcal{E}' \in \mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, 0), \mathcal{I} \subset ([n_1] \times [n_2] \setminus \mathcal{E}'), |\mathcal{I}| = s\} \ .$$

Proof. Let $\mathcal{E}$ be an erasure pattern of the set above. Then, there is an erasure pattern $\mathcal{E}'$ of the set $\mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, 0)$ with $\mathcal{E} = \mathcal{E}' \cup \mathcal{I}$. The complement of $\mathcal{E}'$ is, by definition, an information set of the outer code $\mathcal{C}_{\mathsf{out}}$. By restricting the overall code to this information set, we thus obtain a Gabidulin code with parameters $[(n_1 - b_1)(n_2 - b_2), (n_1 - b_1)(n_2 - b_2) - s]$ by Lemma 3.8. Note that there are exactly s remaining erasures (given by $\mathcal{I}$) in the remaining positions. Since any Gabidulin code is MDS, the restricted code can correct exactly s erasures, which concludes the proof. □

As noted above, this construction is a generalization of the PMDS code construction given in [RKSV13]. The following example captures this special case.

Example 3.3. *Let $\mathcal{C}_{\mathsf{out}}[n_1 n_2, n_1(n_2 - b_2)]$ be the code spanned by*

$$\mathrm{diag}(\underbrace{\mathbf{G}, \mathbf{G}, \ldots, \mathbf{G}}_{n_1 \ \textit{times}}) \ ,$$

where $\mathbf{G}$ is the generator matrix of an arbitrary $[n_2, n_2 - b_2]_q$ MDS code. Observe that $\mathcal{C}_{\mathsf{out}} \in \mathbb{C}^{\mathsf{MR}}_{n_1 \times n_2}(0, b_2, 0)$ and the set of its information sets is given by

$$\mathbb{I}_{\mathsf{out}} = \{\text{From each block pick arbitrary } n_2 - b_2 \text{ positions}\} \ .$$

Choose $\mathcal{C}_{\mathsf{in}}$ to be an $[n_1(n_2 - b_2), n_1(n_2 - b_2) - s]_{q^{n_1(n_2-b_2)}}$ Gabidulin code to obtain the PMDS code construction of [RKSV13]. Observe that $\mathbb{I}_{\mathsf{out}}$ are exactly the subsets of positions that must give an MDS code in a PMDS code according to the second property of Definition 2.8.

3.5.2 Correctable Erasure Patterns

To show that the code of Theorem 3.4 is indeed MR, it remains to characterize the patterns correctable in $T_{n_1 \times n_2}(b_1, b_2, s)$ given the set of patterns correctable in $T_{n_1 \times n_2}(b_1, b_2, 0)$.

Theorem 3.5. *We have*

$$\mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, s) = \{\mathcal{E}' \cup \mathcal{I} \ \mid \ \mathcal{E}' \in \mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, 0), \mathcal{I} \subset (n_1 \times n_2 \setminus \mathcal{E}'), |\mathcal{I}| = s\} \ ,$$

i.e., any $\mathcal{E} \in \mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, s)$ can be obtained by adding s erasures to some $\mathcal{E}' \in \mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, 0)$.

Proof. "$\supseteq$" follows by the construction of Theorem 3.4.

The other direction is implied by the following argument. Let $\mathcal{E} \in \mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, s)$. Denote by $\mathbf{G}$ the generator matrix of a code $\mathcal{C} \in \mathbb{C}_{n_1 \times n_2}(b_1, b_2, s)$ that corrects $\mathcal{E}$ and let $\mathbf{H}_{\mathsf{local}}, \mathbf{H}_{\mathsf{global}}$ be as in Definition 2.9. Denote by $\mathcal{C}' \in \mathbb{C}_{n_1 \times n_2}(b_1, b_2, 0)$ the code obtained by setting $\mathbf{H}_{\mathsf{global}} = \mathbf{0}$. Then there exists a generator matrix of $\mathcal{C}'$ of the form

$$\mathbf{G}' = \begin{pmatrix} \mathbf{G} \\ \mathbf{G}_{\mathsf{global}} \end{pmatrix} \ ,$$

for some $\mathbf{G}_{\mathsf{global}} \in \mathbb{F}^{s \times n_1 n_2}$. Trivially, we have

$$\mathrm{rank}(\mathbf{G}'|_{[n_1 n_2] \setminus \mathcal{E}}) \geq \mathrm{rank}(\mathbf{G}|_{[n_1 n_2] \setminus \mathcal{E}}) = \dim(\mathcal{C}) \ ,$$

where the last equality holds because $\mathcal{E}$ is correctable in $\mathcal{C}$ by definition. By basic linear algebra arguments there exists a subset $\mathcal{I} \subset \mathcal{E}$ with $|\mathcal{I}| = \dim(\mathcal{C}) - \dim(\mathcal{C}') \leq s$ such that

$$\mathrm{rank}(\mathbf{G}'|_{([n_1 n_2] \setminus \mathcal{E}) \cup \mathcal{I}}) = \mathrm{rank}(\mathbf{G}') = \dim(\mathcal{C}') \ .$$

It follows that $\mathcal{E} \setminus \mathcal{I}$ is correctable in $T_{n_1 \times n_2}(b_1, b_2, 0)$. This concludes the proof. □

As an immediate consequence, it follows that, similar to the definition of PMDS codes, the surviving positions after puncturing the "local redundancy" must form an MDS code of distance $s+1$.

Corollary 3.2. *Denote $n = n_1 n_2$. Then, for any $\mathcal{E}' \in \mathbb{E}^{\max}_{n_1 \times n_2}(b_1, b_2, 0)$ and any $[n, k]$ code $\mathcal{C} \in \mathbb{C}^{\mathsf{MR}}_{n_1 \times n_2}(b_1, b_2, s)$ the code $\mathcal{C}|_{n_1 \times n_2 \setminus \mathcal{E}'}$ must be an $[n - |\mathcal{E}'|, k, s+1]$ MDS code.*

Proof. It follows trivially from the dimension of $\mathcal{C}$ that the code $\mathcal{C}|_{[n_2] \setminus \mathcal{E}'}$ can never correct more than s erasures. The existence of a code for which this restriction is an MDS code follows from Corollary 3.3. □

As Gabidulin codes of length n can be defined over any extension field $\mathbb{F}_{q^M}$ as long $n \geq M$, we arrive at the following statement.

Corollary 3.3. *Let* $\mathcal{C}_{\mathsf{out}}[n,k]_q \in \mathbb{C}^{\mathsf{MR}}_{n_1\times n_2}(b_1,b_2,0)$*. Then there exists an* $[n,k-s]_{q^k}$ *code* $\mathcal{C} \in \mathbb{C}^{\mathsf{MR}}_{n_1\times n_2}(b_1,b_2,s)$*.*

Proof. Follows immediately from Theorems 3.4 and 3.5. □

3.6 Summary and Open Problems

This chapter considered the class of maximally recoverable codes for grid-like topologies. First, it was shown that a conjecture on a characterization of the erasure patterns correctable in such a topology is false. Then, codes for grid-like topologies and the erasure patterns correctable therein were related to the class of Tensor-Product codes. Finally, a generic method for adding global redundancy to codes for grid-like topologies was proposed, along with a characterization of the correctable erasure patterns as a function of the patterns correctable without the global redundancy.

The main open problem, which is further motivated by the disproving of Conjecture 3.1, is the search for a characterization of the erasure patterns correctable in grid-like topologies. Further, explicit constructions are only known for very special parameter settings. Such constructions, preferably of small field size, are a major open problem which needs to be solved to increase the practicability of these codes.

4 Partial MDS Codes with Regeneration

Abstract

Partial MDS (PMDS) and sector-disk (SD) codes are classes of erasure correcting codes that combine locality with strong erasure correction capabilities. This chapter introduces the first known construction of PMDS codes with global regeneration, which allows for efficient repair for patterns of node failures that exceed the local erasure correction capability of the code and thereby invoke repair across different local repair sets. Further, new constructions of PMDS and SD codes with local regeneration are presented, where each local code is a bandwidth-optimal regenerating MDS code. In the event of a node failure, these codes reduce both, the number of servers that have to be contacted as well as the amount of network traffic required for the repair process. The constructions require significantly smaller field size than the only other construction known in literature.

This chapter is based on the work [HPYWZ21] published in the IEEE Transactions on Information Theory*. In part, the results on PMDS codes with local regeneration have been published in the proceedings of the* 2020 IEEE International Symposium on Information Theory (ISIT) *[HPYW20] (nominated for the* 2021 NVMW Memorable Paper Award*).*

4.1 Introduction

The previous chapter dealt with MR codes for grid-like topologies. Such codes and, in general, codes with locality are of broad interest because of the small number of positions required for node recovery, which is an important characteristic of DSSs. Regenerating codes address the other major concern in node recovery, namely the required amount of data transmitted between the nodes. To lower this *repair bandwidth*, regenerating codes allow for a larger number of nodes to be involved in the repair process. The naive approach to node repair in an MDS code of length n and dimension k requires exactly k nodes to be involved. By accessing $d > k$ nodes, but only retrieving a function of the data stored on each node, regenerating codes significantly decrease the repair bandwidth. Lower bounds on the required bandwidth for repair have been derived in [DGW$^+$10; CJM$^+$13] which lead to two extremal code classes, namely *minimum bandwidth regenerating* (MBR) and *minimum storage regenerating* (MSR) *codes*. MBR codes, as constructed in [RSK11; KSP$^+$13; LC14; KK16b], offer the lowest possible repair traffic, but at the cost of increased storage overhead compared to MDS codes. This chapter considers d-MSR codes, as constructed in [SR10; RSK11; SRKR11; TWB12; CJM$^+$13; GFV17; YB17a; LTT18], which require more network traffic for repair than MBR codes, but are optimal in terms of storage overhead, i.e., they are MDS.

As locality and regeneration are two properties that allow for more efficient node repair, it is a natural question whether these approaches can be combined. One common motivation for the use of locality is the physical limitation of the interconnect between nodes from different local repair sets, e.g., because nodes are in different racks or even data centers. In this case, it is of particular importance that the communication required in case of global recovery is low, despite it being less likely than failure events that incur local recovery. In the first part of this chapter, we consider PMDS codes with such global regeneration properties that offer nontrivial repair schemes for the case where local recovery is not possible. Specifically, we introduce a PMDS code construction that becomes an MSR code, when the local redundancy is removed.

Another approach addressing this problem is clustered storage [GPV13; PYGP13; SCM18; SCYM18] and, specifically, *rack-aware regenerating codes* (RRCs) [HLSH19; HLH20]. In this setting, the nodes are partitioned into a smaller number of racks, similar to the partitioning of nodes for codes with locality. When a node fails within a rack, it is regenerated by transmitting from each rack a function of the content of its nodes. The distinction to regenerating codes is that the repair traffic is given by the amount of data transmitted *between* the racks, while communication within each rack is ignored. Aside from this definition of the repair bandwidth, there are two important differences to the model we consider: 1) RRCs require a node that collects the data from the nodes within the rack and computes a function of it that is to be transmitted and 2) RRCs generally *do not have locality*, i.e., no repair is possible within each rack. *Double regenerating codes* [HLZ16] refine this model by considering two levels of regeneration, a local one, i.e., within the racks, and a global one, i.e., across the racks. A sightly different model has been considered in [PAM18], in which repair is conducted by downloading a number of symbols from helper racks (also called clusters) and additionally a number of symbols from a set of nodes of the same rack, where, unlike for RRCs, both contribute to the overall repair bandwidth. Similar to RRCs and double regenerating codes, the codes under this model do not have locality.

A rack-aware setting that also considers local recovery from node failures is given by *multi-rack distributed storage* [TCS14; QLZ$^+$18]. There, a small number of nodes can be repaired/regenerated locally and failure patterns for which this is not possible are repaired by contacting other racks in addition to the surviving local nodes. Similar to RRCs, it is assumed that the contacted helper racks can process the data of the nodes within the rack and that the communication between racks is more costly than within a rack. Along with an information-theoretic bound, [QLZ$^+$18] presents a construction for the case of an efficient local repair of a single node failure. The minimization of the cross-rack repair bandwidth is stated as an open problem. In [TCS14] the authors consider a more general setting which improves both, the repair bandwidth within a rack in case of a small number of failure and across racks for failure patterns that cannot be repaired locally. Similar to RRCs, this model differs from the one considered in this chapter in that racks are able to process the data from their nodes prior to sending it to other racks. Additionally, we consider a stronger notion of locality in requiring the storage code to be PMDS.

The work with closest relation to the model of global regeneration in codes with locality that we consider is [GKJS17], which introduces local redundancy by splitting parity-check equations of HashTag codes [KGØ16; KGJØ17; LTT18]. While it is shown that the codes are distance-optimal LRCs, they are generally not PMDS codes and possess only information locality, i.e., the recovery from a small subset of positions is only guaranteed for a set of systematic positions.

As noted above, the probability of such a global regeneration event to be induced is low compared to the event of local recovery, hence, the relative importance of a bandwidth-efficient solution for each case depends on various system parameters. The second part of this chapter therefore considers PMDS code with local regeneration. While locality limits the number of nodes involved in this process, the local recovery of nodes still induces a large amount of network traffic, as the entire content of the helper nodes needs to be downloaded when applying straight-forward recovery algorithms. To circumvent this bottleneck, several locally regenerating codes [DGW$^+$10] have been proposed [KPLK14; RKSV13; KNK18; GKJS17; Hol14; LLL16]. In [CK16] it was shown that the LRC construction of [RKSV13] is in fact a PMDS code, implicitly giving the first construction of PMDS codes with local regeneration[1]. However, these PMDS codes require a field size exponential in the length of the code and the subpacketization of the local regenerating code (which may itself be exponential in the length of the local code).

4.1.1 Contributions and Outline

In Section 4.3, we propose the first known construction of globally MSR PMDS codes, where the MDS code obtained from puncturing b positions in each local repair set is an MSR code. This allows for a significant reduction in the repair bandwidth in case a global repair event is triggered. For that purpose, we introduce a new MSR

[1] The construction in [RKSV13] consists of two encoding stages, where in the second stage an arbitrary linear MDS code can be used to obtain the local codes. In [CK16] it was shown that the construction in fact gives a PMDS code, independent of the explicit choice of the MDS code in the second encoding stage. It follows that using a regenerating MDS code in the second encoding stage results in a PMDS code with local regeneration.

code construction based on [YB17a] which utilizes Gabidulin instead of Reed–Solomon codes and prove that it is in fact an MSR code. This allows for building PMDS codes with regenerating properties in a similar fashion as the Gabidulin-code-based PMDS code construction (without regeneration) in [RKSV13]. The involved part for retaining the MSR property for any puncturing of b positions in each local repair set is the choice of evaluation points of these Gabidulin codes. An explicit choice based on pairwise trivially intersecting subspaces is presented and proved to fulfill the required property for any such puncturing pattern. The resulting code has a field size in $O(n_l^{\mu(n_l+s)})$ and subpacketization in $O((8n_l)^{\mu n_l(n_l+s)})$.

In Section 4.4.1, we construct a new PMDS code with two global parities ($s = 2$), where each local code is a d-MSR code. The construction is a nontrivial combination of the PMDS codes in [BPSY16] with the MSR codes in [YB17a], and has field size in the order of

$$O(\mu b^2 n_l) \ .$$

In Section 4.4.2, we present a new general construction of locally MSR PMDS codes for any number of global parities. The construction combines an arbitrary family of universal PMDS codes, a class of PMDS codes that allows for the local codes to be chosen almost arbitrarily, and an MSR code whose rows are all MDS codes. This immediately leads to several new explicit locally MSR PMDS codes using known universal PMDS code families and the MSR codes in [YB17a]. The PMDS codes in [RKSV13] result in a field size in the order of

$$O\Big((bn_l)^{\mu(n_l-b)}\Big)$$

and the ones in [MPK19] give a field size in

$$O\Big(\max\{bn_l, \mu+1\})^{n_l-b}\Big) \ .$$

We also slightly generalize the PMDS code family in [GYBS18] such that it becomes universal. The resulting field size of the corresponding locally MSR PMDS code is in

$$O(n_l b(2n_l\mu)^{s(b+1)-1}) \ .$$

All new locally MSR PMDS codes have the same subpacketization as the underlying MSR code from [YB17a].

In Section 4.5, we analyze the field size of the new constructions of locally MSR PMDS codes. For the two-global-parities construction based on [BPSY16] and the universal construction with the PMDS codes in [MPK19] and [GYBS18], there is a reasonable parameter range in which the respective construction has lowest field size among all known constructions. Moreover, for all parameters, there is a new construction that has a smaller field size than the only known construction of [RKSV13].

Table 4.1: An overview of the notation used in this work compared to the notation used in [BPSY16; GYBS18; MPK19; RKSV13; YB17a]. The largest benefit from this comparison is in Sections 4.4.1, 4.4.2 and 4.5, where we construct and discuss PMDS codes with local MSR codes. Therefore, the length and number of parities in the MSR code construction of [YB17a] are matched with the parameters of the local codes in our work. Note that, in our notation, the length of the MSR code in Section 4.3, where we consider PMDS codes with global repair properties, is $\mu(n_l - b)$ and the number of parities is s.

Description	[BPSY16]	[GYBS18]	[MPK19]	[RKSV13]	[YB17a]	This work
Number of local repair sets	r	m	g	g	-	μ
Length of local MSR code	n	n	$r+\delta-1$	$r+\delta-1$	n	n_l
# of local parity symbols	m	r	$\delta-1$	$\delta-1$	r	b
# of global parity symbols	s	s	h	$D-1$	-	s
Code length	rn	mn	n	n	-	μn_l
Subpacketization	-	-	-	α	l	ℓ
# of nodes needed for repair	-	-	-	d	d	d

4.2 Regenerating Codes

This work is largely based on the constructions of PMDS codes by Rawat et al. [RKSV13], Blaum et al. [BPSY16], Gabrys et al. [GYBS18], and Martínez-Peñas and Kschischang [MPK19] as well as the construction of MSR codes by Ye and Barg in [YB17a]. Since the notations in these works are conflicting, i.e., the same symbols are used for different parameters of the codes, Table 4.1 provides an overview of the notation used in this work compared to these works.

In accordance with [DGW$^+$10; CJM$^+$13], we formally define the class of MSR codes.

Definition 4.1 (Regenerating code [DGW$^+$10; CJM$^+$13])**.** *Let $\mathcal{F}, \mathcal{R} \subset [n_l]$ be two disjoint subsets. Let $\mathcal{C}$ be an $[n_l, n_l - b; \ell]_q$ MDS array code. Define $M(\mathcal{C}, \mathcal{F}, \mathcal{R})$ as the smallest number of symbols of $\mathbb{F}_q$ one needs to download from the surviving nodes indexed by $\mathcal{R}$ to recover the erased nodes indexed by $\mathcal{F}$. Then*

$$M(\mathcal{C}, \mathcal{F}, \mathcal{R}) \geq \frac{|\mathcal{F}||\mathcal{R}|\ell}{|\mathcal{F}| + |\mathcal{R}| - n_l + b} \ . \tag{4.1}$$

*For two integers h, d, with $1 \leq h \leq b$ and $n_l - b \leq d \leq n_l - h$, we say that the code $\mathcal{C}$ is an (h,d)-*MSR *code if*

$$\max_{\substack{|\mathcal{F}|=h, |\mathcal{R}|=d \\ \mathcal{F} \cap \mathcal{R} = \emptyset}} M(\mathcal{C}, \mathcal{F}, \mathcal{R}) = \frac{hd\ell}{h + d - n_l + b} \ .$$

If $h = 1$ we say that the code is a d-MSR code. If in addition, $d = n_l - 1$, we simply say that the code is an MSR code.

Informally, in a regenerating array code, as in Definition 4.1, every codeword is required to be recoverable from an arbitrary subset of $n_l - b$ columns. We now define a slightly stronger property, which imposes a similar requirement on every row of a codeword.

Definition 4.2. *Let $\mathcal{C}$ be an $[n_l, n_l - b; \ell]$ regenerating code as in Definition 4.1. We say that the code $\mathcal{C}$ is a* row-wise MDS *regenerating code if for any $i \in [\ell]$ the set $\{\mathbf{C}[i, :] \mid \mathbf{C} \in \mathcal{C}\}$ is an $[n_l, n_l - b]$ MDS code.*

4.2.1 Regenerating Codes with Locality

With these notions established, we combine Definitions 2.8 and 4.1 to formally define the class of globally MSR PMDS codes.

Definition 4.3 (Globally (h, d)-MSR PMDS array code)**.** *Let $\mathcal{C}$ be a $\mathsf{PMDS}(\mu, n_l, b, s, \mathbb{W}; \ell)$ code and d, h be chosen such that $1 \leq h \leq s$ and $\mu(n_l - b) - s \leq d \leq \mu(n_l - b) - h$. The code $\mathcal{C}$ is* globally (h, d)-MSR *if the restriction $\mathcal{C}|_{[\mu n_l] \setminus \cup_{i=1}^{\mu} \mathcal{E}_i}$ is a $[\mu(n_l - b), \mu n_l - b\mu - s, s + 1; \ell]$ (h, d)-MSR code for any $\mathcal{E}_i \subset \mathcal{W}_i$ with $|\mathcal{E}_i| = b$ for all $i \in [\mu]$.*

If $h = 1$ we say the code is a globally d-MSR PMDS code. If in addition, $d = \mu(n_l - b) - 1$, we simply say that the code is a globally MSR PMDS code.

The class of locally MSR PMDS codes, which we construct and analyze in Sections 4.4.1, 4.4.2 and 4.5, is defined similarly, except that the MSR property is required for every *local* MDS code.

Definition 4.4 (Locally (h, d)-MSR PMDS array code)**.** *Let $\mathcal{C}$ be a $\mathsf{PMDS}(\mu, n_l, b, s, \mathbb{W}; \ell)$ code and d, h be chosen such that $1 \leq h$ and $n_l - b \leq d \leq n_l - h$. The code $\mathcal{C}$ is locally (h, d)-MSR if $\mathcal{C}|_{\mathcal{W}_i}$ is an (h, d)-MSR code for all $i \in [\mu]$.*

If $h = 1$ we say the code is a locally d-MSR PMDS code. If in addition, $d = n_l - 1$, we simply say that the code is an locally MSR PMDS code.

Fig. 4.1 shows an illustration of a locally MSR PMDS array code. Assuming the code to be an $(b = 2, s = 2)$-PMDS code, the erasures in the first local code can be corrected locally, but without taking advantage of the regenerating property, as the number of available helper nodes is only $n_l - b$. The erasure in the second local code can be corrected from the remaining $n_l - b + 1$ nodes in the local repair set using the locally regenerating property, and the erasures in the third local code can be recovered by accessing nodes of the other local repair sets and could therefore benefit from the global MSR property.

Throughout the paper, we consider in all constructions $h = 1$. This is the most interesting case since in a storage system it is more likely that one node needs to be

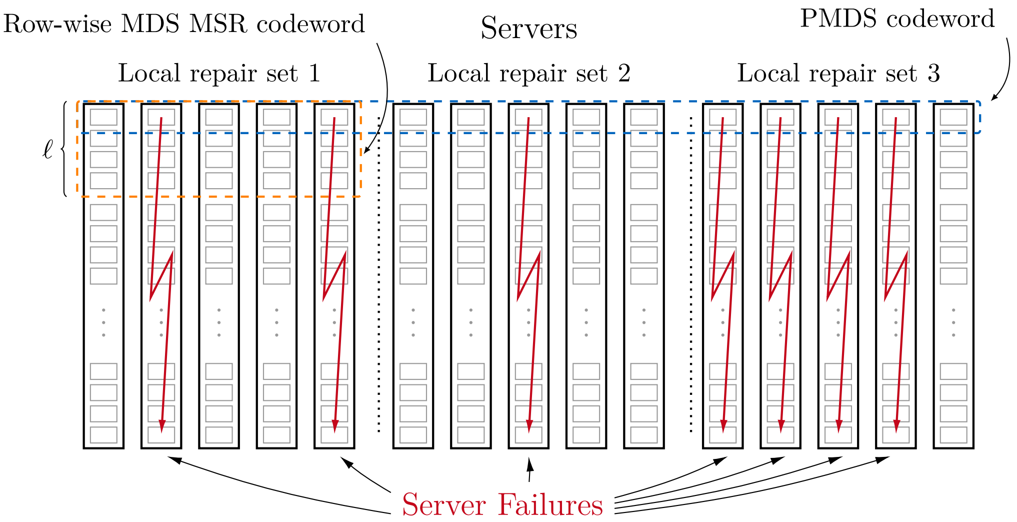


Figure 4.1: Illustration of a DSS encoded with a locally MSR PMDS array code with $n_l = 5$, $\mu = 3$, and each symbol of the code alphabet represented by a small rectangle. The shown erasure pattern can be corrected by an $(b = 2, s = 2)$-PMDS code.

regenerated than multiple nodes. In the globally MSR case, we further fix d to be maximal, i.e., $d = \mu(n_l - b) - 1$. It can be seen from the bound in Definition 4.1 that the repair bandwidth decreases in d, i.e., it is minimal for this choice of d. See Section 4.6 for a discussion on how the results can be generalized.

4.2.2 Ye-Barg Regenerating Codes

We provide a formal definition of the MSR codes of [YB17a, Construction 2] in the notation used in this work.

Definition 4.5 (Ye-Barg (YB) d-MSR codes [YB17a, Construction 2])**.** *Let $z = d + 1 - (n_l - b)$ and $\{\beta_{i,j}\}_{i \in [z], j \in [n_l]}$ be a set of zn_l distinct elements of $\mathbb{F}_q$, where $q \geq zn_l$. The $[n_l, n_l - b; \ell]$ array code $\mathcal{C}$ over $\mathbb{F}_q$ is defined to be the set of codeword arrays with $\ell = z^{n_l}$ rows and n_l columns, where the a-th row fulfills the parity check equations*

$$\mathbf{H}^{(a)} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \beta_{a_1,1} & \beta_{a_2,2} & \dots & \beta_{a_{n_l},n_l} \\ \vdots & \vdots & & \vdots \\ \beta_{a_1,1}^{b-1} & \beta_{a_2,2}^{b-1} & \dots & \beta_{a_{n_l},n_l}^{b-1} \end{pmatrix},$$

for $a \in [\ell]$ and $a - 1 = \sum_{i=1}^{n_l} (a_i - 1) z^{i-1}$ with $a_i \in [z]$.

As each row a fulfills the parity-check equations of an GRS code (see Definition 2.3), it is easy to see that a Ye-Barg code as in Definition 4.5 is in fact *row-wise MDS d*-MSR code.

Remark 4.1. *All constructions presented in the following can also be applied to obtain globally/locally (h, d)-MSR PMDS codes, where the corresponding restriction of the code is (the skew analog of) an (h, d)-MSR code as in [YB17a, Construction 3], which is very similar in structure to [YB17a, Construction 2] given in Definition 4.5. However, as the required subpacketization is larger for the former and the case of d-MSR codes has the highest practical relevance, we focus on this class in this work.*

Remark 4.2. *In Definition 4.5 we define each row of the array code by a set of parity check equations independent of the other $\ell - 1$ rows of the array. Note that this is not possible for array codes in general. However, for the existence of such a description it is sufficient that the matrices A_i, as defined in [YB17a], are diagonal matrices. This allows for the simplified notation of Definition 4.5 for the cases considered in this work, which highlights the fact that each row is indeed an $[n_l, n_l - b]$ RS code, and thus MDS.*

4.3 PMDS Codes with Nontrivial Global Regeneration

By definition, a PMDS code punctured in arbitrary b positions of each local repair set is an MDS code of distance $s + 1$. In the following we construct PMDS codes where each of these MDS codes is an MSR code. For the sake of simplicity, we focus on the case of highest practical interest: MSR codes with that repair one position ($h = 1$) from all $d = \mu(n_l - b) - 1$ remaining positions of the MDS code.

The construction is based on two main observations. First, the principle used in the MSR codes of [YB17a] can also be applied using Gabidulin codes (recall Definition 2.6) instead of RS codes. Second, as already used in the construction of codes for grid-like topologies with global redundancy in Section 3.5, performing linearly independent linear combinations of the symbols of a Gabidulin code yields another Gabidulin code with different code locators. Using these observations and carefully choosing the code locators for each row in an array of Gabidulin codewords, we assure that the code obtained from puncturing b positions in each local repair set is MSR. The studied construction works as follows.

Construction 4.1 (Globally MSR PMDS array codes)**.** *Let μ, n_l, b, s be valid PMDS parameters and $\mathbf{B} \in \mathbb{F}_{q^M}^{\ell \times \mu(n_l - b)}$ be a matrix with entries $\beta_{i,j}, i \in [\ell], j \in [\mu(n_l - b)]$. We define the $[\mu n_l, \mu(n_l - b) - s; \ell]_{q^M}$ array code $\mathcal{C}(\mu, n_l, b, s, \mathbf{B}; \ell)_q$ as*

$$\left\{ \mathbf{C} \in \mathbb{F}_q^{\ell \times \mu n_l} \mid \mathbf{C}[a, :] = \mathbf{u}^{(a)} \cdot \mathbf{G}_{\mathbf{B}}^{(a)} \cdot \mathrm{diag}(\mathbf{G}_{\mathsf{MDS}}, \mathbf{G}_{\mathsf{MDS}}, \ldots) \,\forall\, \mathbf{u}^{(a)} \in \mathbb{F}_{q^m}^{\mu(n_l - b) - s}, a \in [\ell] \right\},$$

where $\mathbf{G}_{\mathbf{B}}^{(a)}$ is a generator matrix of the code $\mathsf{Gab}(\mu(n_l - b), \mu(n_l - b) - s, \mathbf{B}[a, :])$ as in Definition 2.6 and $\mathbf{G}_{\mathsf{MDS}}$ is a generator matrix of an $[n_l, n_l - b]_q$ MDS code.

It is easy to see that if the rows of the matrix $\mathbf{B}$ in Construction 4.1 contain linearly independent elements, then each row of the code is a PMDS code of the code family constructed in [RKSV13] (see also Example 3.3). In the remainder of this section, we prove that if the matrix $\mathbf{B}$ is chosen in a suitable way, then the MDS array codes obtained from erasing b positions in each local repair set are MSR codes of the following type, which can be seen as a Gabidulin-analog of Ye–Barg codes.

Definition 4.6 (Skew Ye–Barg d-MSR codes). *Let μ, n_l, b, s be valid PMDS parameters, $\ell \in \mathbb{Z}_{>0}$, and $\mathbf{B} \in \mathbb{F}_{q^m}^{\ell \times \mu(n_l-b)}$ be a matrix with entries $\mathbf{B}[i,j] = \beta_{i,j}$. Define $\mathcal{C}(\mu, n_l, b, s, \mathbf{B}) \subset \mathbb{F}_q^{\ell \times \mu(n_l-b)}$ to be an $[\mu(n_l-b), \mu(n_l-b)-s; \ell]$ array code over $\mathbb{F}_{q^m}$, where each codeword is a matrix with ℓ rows and $\mu(n_l-b)$ columns, such that for any $a \in [\ell]$ the a-th row is a codeword of a code with parity-check matrix*

$$\mathbf{H}_{\mathbf{B}}^{(a)} = \begin{pmatrix} \beta_{a,1} & \beta_{a,2} & \cdots & \beta_{a,\mu(n_l-b)} \\ \beta_{a,1}^{q^1} & \beta_{a,2}^{q^1} & \cdots & \beta_{a,\mu(n_l-b)}^{q^1} \\ \vdots & \vdots & & \vdots \\ \beta_{a,1}^{q^{s-1}} & \beta_{a,2}^{q^{s-1}} & \cdots & \beta_{a,\mu(n_l-b)}^{q^{s-1}} \end{pmatrix}.$$

Denote by $\mathbf{G}_{\mathbf{B}}^{(a)}$ a generator matrix corresponding to $\mathbf{H}_{\mathbf{B}}^{(a)}$.

Remark 4.3. *Definition 4.6 is essentially the same as Definition 4.5, except that it relies on Gabidulin codes. Note that there is also a difference in presentation: the locators are not given as a set of elements, but instead given explicitly as an input for each row. For Definition 4.5 the corresponding matrix $\mathbf{B}$ is easily obtained from a set $\mathcal{B} = \{\beta_{i,j}\}_{i\in[z], j\in[n_l]}$ of distinct elements of $\mathbb{F}_{q^M}$ by assigning $\mathbf{B}[a,j] = \beta_{a_j,j}$ for $a \in [\ell]$ and $a-1 = \sum_{i=1}^{n_l}(a_i-1)z^{i-1}$ with $a_i \in [z]$, i.e., assigning the code locators of $\mathbf{H}^{(a)}$ to the a-th row of $\mathbf{B}$.*

For the node repair algorithm of Ye–Barg codes [YB17a], it is essential that the rows of a codeword can be partitioned into subsets for which there exist parity checks that differ exactly in position i , i.e., for which all entries are the same except for those at position i, which are all distinct. This is due to the close relation of Ye–Barg to Reed–Solomon codes. In Lemma 4.1 below, we analogously prove that Skew Ye–Barg codes are MSR codes if the matrix of code locators $\mathbf{B}$ has the following property, which is due to their relation to Gabidulin codes.

Definition 4.7 (YB-Grouping Property). *Let μ, n_l, b, s be valid PMDS parameters and $\mathbf{B} \in \mathbb{F}_{q^m}^{\ell \times \mu(n_l-b)}$. We say that the matrix $\mathbf{B}$ has the* YB-grouping property w.r.t. *s if for each position $i \in [\mu(n_l-b)]$ the rows of the matrix can be partitioned into ℓ/s subsets $\mathcal{Z}_1, \mathcal{Z}_2, \dots, \mathcal{Z}_{\frac{\ell}{s}}$ of $|\mathcal{Z}_a| = s$ rows such that for each $a \in [\ell/s]$ the elements $\{\mathbf{B}[z,i] \mid z \in \mathcal{Z}_a\}$ are linearly independent and the elements $\{\mathbf{B}[z,j] \mid z \in \mathcal{Z}_a\}$ are the same for all other positions $j \in [\mu(n_l-b)] \setminus \{i\}$.*

This Ye-Barg grouping property is the key to the Ye-Barg MSR[2] codes of [YB17a]. We modify the presented repair algorithm for our skew Ye-Barg codes to show their MSR property.

Lemma 4.1. *Let μ, n_l, b, s be valid PMDS parameters and $\mathbf{B} \in \mathbb{F}_{q^M}^{\ell \times \mu(n_l-b)}$ be a matrix such that for any $a \in [\ell]$ the elements of its a-th row $\mathbf{B}^{(a)}$ are linearly independent over $\mathbb{F}_q$. Further, let $\mathbf{B}$ have the Ye-Barg grouping property w.r.t. s as in Definition 4.7. Then the code $\mathcal{C}(\mu, n_l, b, s, \mathbf{B})$ as in Definition 4.6 is an MSR code.*

Proof. The MDS property follows directly as each row is a codeword of a Gabidulin code, which are well-known to be MDS. It is easy to check that the recovery algorithm of [YB17a, Theorem 1] also applies to the code of Definition 4.6. For completeness we include a short proof here. Assume node i failed, i.e., we need to recover the set $\{\mathbf{C}[a,i] \ \forall \ a \in [\ell]\}$ from the helper nodes with indices $[\mu(n_l-b)] \setminus \{i\}$. Denote $\mathbf{B}[i,j] = \beta_{i,j}$ and let $\{\mathcal{Z}_{i,1}, \mathcal{Z}_{i,2}, \dots, \mathcal{Z}_{i,\frac{\ell}{s}}\}$ be the partition of $[\ell]$ into the subsets $\mathcal{Z}_{i,z}$ of s row indices for which the parity check equations differ exactly in position i and the entries in position i are linearly independent. Note that such a partition exists for every $i \in [\mu(n_l-b)]$ by definition of the Ye-Barg grouping property. The a-th row of a codeword $\mathbf{C} \in \mathcal{C}$ is determined by the s parity checks

$$0 = \sum_{j=1}^{\mu(n_l-b)} \beta_{a,j}^{q^\xi} \mathbf{C}[a,j] = \beta_{a,i}^{q^\xi} \mathbf{C}[a,i] + \sum_{\substack{j=1 \\ j \neq i}}^{\mu(n_l-b)} \beta_{a,j}^{q^\xi} \mathbf{C}[a,j]$$

for $\xi \in [0, s-1]$. Observe that $\beta_{a,j} = \beta_{a',j} \ \forall \ a, a' \in \mathcal{Z}_{i,z}, j \neq i$ and by slight abuse of notation we denote $\beta_{\mathcal{Z}_{i,z},j} := \beta_{a,j}, \ a \in \mathcal{Z}_{i,z}$. Summing over all $a \in \mathcal{Z}_{i,z}$ gives

$$\sum_{a \in \mathcal{Z}_{i,z}} \beta_{a,i}^{q^\xi} \mathbf{C}[a,i] = \sum_{a \in \mathcal{Z}_{i,z}} \sum_{\substack{j=1 \\ j \neq i}}^{\mu n_l} \left(\beta_{a,j}^{q^\xi} \mathbf{C}[a,j] \right) = \sum_{\substack{j=1 \\ j \neq i}}^{\mu n_l} \left(\beta_{\mathcal{Z}_{i,z},j}^{q^\xi} \sum_{a \in \mathcal{Z}_{i,z}} \mathbf{C}[a,j] \right) . \qquad (4.2)$$

This is a linear system of equations with s unknowns $\mathbf{C}[a,i], a \in \mathcal{Z}_{i,z}$ and s equations, one for each $\xi \in [0, s-1]$. By the Ye-Barg grouping property of $\mathbf{B}$, the elements $\{\beta_{a,i} \mid a \in \mathcal{Z}_{i,z}\}$ are linearly independent and the equations therefore linearly independent. Hence, the unknowns can be uniquely determined if the right hand side of Eq. (4.2) is known. Therefore, for repair of node i, node j transmits the set of symbols

$$\left\{ \sum_{a \in \mathcal{Z}_{i,z}} \mathbf{C}[a,j] \mid z \in [\ell/s] \right\} .$$

[2] Specifically, as these codes consider RS instead of Gabidulin codes the elements in the i-position only need to be distinct in this case, not necessarily linearly independent.

As the cardinality of this set is ℓ/s, the repair bandwidth is $(\mu(n_l-b)-1)\ell/s$ and thereby fulfills the bound on the minimal repair bandwidth of Definition 4.1 with equality, i.e., the code is an MSR code. □

As each row in a codeword of a skew Ye-Barg codes is a codeword of a Gabidulin code, Construction 4.1 can be applied to this code by multiplying it from the right by the $\mu(n_l-b) \times \mu n_l$ matrix $\mathrm{diag}(\mathbf{G}_{\mathsf{MDS}}, \mathbf{G}_{\mathsf{MDS}}, \ldots)$. When puncturing arbitrary b positions in each local repair set, which corresponds to removing columns of this block diagonal matrix, we do not obtain the original skew Ye–Barg code. However, we do get the original code multiplied by an invertible matrix over $\mathbb{F}_q$ from the right. By Lemma 2.2 the rows of the resulting code are again codewords of a Gabidulin code. In the following theorem we give a sufficient condition on the matrix $\mathbf{B}$ for this code to again be a skew Ye-Barg code. If this holds for every possible puncturing pattern, the code is globally MSR PMDS code as in Definition 4.3.

Theorem 4.1. *Let μ, n_l, b, s be valid PMDS parameters, $\mathbb{W} = \{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_\mu\}$ be a partition of $[\mu n_l]$ with $|\mathcal{W}_i| = n_l \ \forall \ i \in [\mu]$. Then, the code $\mathcal{C}(\mu, n_l, b, s, \mathbf{B}; \ell)_{q^M}$ as in Construction 4.1 is a globally MSR PMDS code if the matrix*

$$\mathbf{B} \cdot (\mathrm{diag}(\mathbf{G}_{\mathsf{MDS}}, \mathbf{G}_{\mathsf{MDS}}, \ldots)|_{[\mu n_l]\setminus \cup_{i=1}^{\mu} \mathcal{E}_i})^{-1}$$

has the YB grouping property *(as in Definition 4.7) for any $\mathcal{E}_i \subset \mathcal{W}_i$ with $|\mathcal{E}_i| = b$.*

Proof. Without loss of generality assume that $\mathcal{W}_i = [(i-1)n_l + 1, in_l]$. Denote $\mathcal{I} = [\mu n_l] \setminus \cup_{i=1}^{\mu} \mathcal{E}_i$, where $\mathcal{E}_i \subset \mathcal{W}_i$ with $|\mathcal{E}_i| = b$ for all $i \in [\mu]$, and $\bar{\mathcal{E}_i} = [n_l] \setminus \mathcal{E}_i$. The restriction of the code $\mathcal{C}$ to the positions indexed by $\mathcal{I}$ is the code

$$\mathcal{C}_{\mathcal{I}} = \left\langle \left(\mathbf{G}^{(a)} \cdot \mathrm{diag}(\mathbf{G}_{\mathsf{MDS}}, \mathbf{G}_{\mathsf{MDS}}, \ldots)\right)\Big|_{\mathcal{I}} \right\rangle = \left\langle \mathbf{G}^{(a)} \cdot (\mathrm{diag}(\mathbf{G}_{\mathsf{MDS}}, \mathbf{G}_{\mathsf{MDS}}, \ldots)|_{\mathcal{I}}) \right\rangle \ .$$

As $\mathbf{G}_{\mathsf{MDS}}$ is the generator matrix of an MDS code, the matrix $\mathrm{diag}(\mathbf{G}_{\mathsf{MDS}}|_{\bar{\mathcal{E}}_1}, \mathbf{G}_{\mathsf{MDS}}|_{\bar{\mathcal{E}}_2}, \ldots)$ is a full-rank $\mathbb{F}_q^{\mu(n_l-b)\times\mu(n_l-b)}$ matrix. By Lemma 2.2 it follows that code $\mathcal{C}_{\mathcal{I}}^{(a)}$, consisting of the a-th row of every codeword of $\mathcal{C}_{\mathcal{I}}$, is a $\mathsf{Gab}(\mu(n_l-b), \mu(n_l-b)-s, \boldsymbol{\beta})$ code with

$$\boldsymbol{\beta} = \mathbf{B}[a,:] \cdot (\mathrm{diag}(\mathbf{G}_{\mathsf{MDS}}, \mathbf{G}_{\mathsf{MDS}}, \ldots)|_{\mathcal{I}})^{-1} \ .$$

It follows directly from Lemma 4.1 that the code is MSR if the matrix

$$\mathbf{B} \cdot (\mathrm{diag}(\mathbf{G}_{\mathsf{MDS}}, \mathbf{G}_{\mathsf{MDS}}, \ldots)|_{\mathcal{I}})^{-1}$$

has the Ye-Barg grouping property. □

It remains to construct a matrix $\mathbf{B}$ that fulfills the property of Theorem 4.1. We use the following slightly stronger property to simplify the analysis.

Definition 4.8. *We say that the matrix* $\mathbf{B} \in \mathbb{F}_{q^M}^{\ell \times (\mu(n_l-b))}$ *has the* scrambled YB grouping property *if* $\mathbf{B} \cdot \operatorname{diag}(\mathbf{G}_1, \dots, \mathbf{G}_\mu)$ *has the YB grouping property for all invertible matrices* $\mathbf{G}_i \in \mathbb{F}_q^{(n_l-b)\times(n_l-b)}$.

The following theorem gives a construction of a matrix $\mathbf{B}$ that has the scrambled YB grouping property.

Theorem 4.2. *Let* $M = \mu(n_l - b + s - 1)$ *and choose* μ *subspaces*

$$\mathcal{B}^{(1)}, \dots, \mathcal{B}^{(\mu)} \in \operatorname{Gr}(\mathbb{F}_q^M, n_l - b + s - 1) \ ,$$

i.e., $n_l - b + s - 1$*-dimensional subspaces of* $\mathbb{F}_q^M$*, that span the space* $\mathbb{F}_q^M$.

For $i = [\mu]$*, consider the sets*

$$\mathcal{S}^{(i)} := \{(\beta_1, \dots, \beta_{n_l-b}) \,|\, \langle \beta_1, \dots, \beta_{n_l-b}\rangle_{\mathbb{F}_q} \text{is an } (n_l-b)\text{-dimensional subspace of } \mathcal{B}^{(i)}\}$$

and

$$\mathcal{S} := \left\{(\boldsymbol{\beta}^{(1)} \mid \cdots \mid \boldsymbol{\beta}^{(\mu)}) \mid \boldsymbol{\beta}^{(i)} \in \mathcal{S}^{(i)}\right\}.$$

Then, the cardinality of $\mathcal{S}$ *is*

$$\begin{aligned}\ell := |\mathcal{S}| &= \left(\begin{bmatrix} n_l - b + s - 1 \\ n_l - b \end{bmatrix}_q \prod_{i=0}^{n_l-b-1} \left(q^{n_l-b} - q^i\right)\right)^\mu \\ &\leq 4^\mu q^{\mu(n_l-b)(n_l-b+s-1)}.\end{aligned}$$

Let $\mathbf{B} \in \mathbb{F}_{q^M}^{\ell \times (n_l-b)\mu}$ *be a matrix whose rows are exactly the entries of* $\mathcal{S}$*. Then,* $\mathbf{B}$ *has the scrambled YB grouping property as in Definition 4.8.*

Proof. The cardinality of $\mathcal{S}^{(i)}$ is the number of $(n_l - b)$-dimensional subspaces of an $(n_l - b + s - 1)$-dimensional vector space over $\mathbb{F}_q$, times the number of bases of such a subspace. The latter equals the number of invertible $(n_l - b) \times (n_l - b)$ matrices over $\mathbb{F}_q$. Hence, we have

$$|\mathcal{S}_i| = \begin{bmatrix} n_l - b + s - 1 \\ n_l - b \end{bmatrix}_q \prod_{i=0}^{n_l-b-1} \left(q^{n_l-b} - q^i\right) \leq 4q^{(s-1)(n_l-b)} q^{(n_l-b)^2} = 4q^{(n_l-b)(n_l-b+s-1)} \ ,$$

where the inequality holds by [Ove07, Lemma 3.13]. Overall, we get

$$\begin{aligned}\ell = |\mathcal{S}| &= \prod_{i=1}^{\mu} |\mathcal{S}^{(i)}| \\ &= \left(\begin{bmatrix} n_l - b + s - 1 \\ n_l - b \end{bmatrix}_q \prod_{i=0}^{n_l-b-1} \left(q^{n_l-b} - q^i \right) \right)^{\mu} \\ &\leq 4^{\mu} q^{\mu(n_l-b)(n_l-b+s-1)}.\end{aligned}$$

The matrix containing the elements of $\mathcal{S}$ as rows has the YB grouping property:

- Every element of $\mathcal{S}$ is a vector consisting of linearly independent entries. This is obvious since the $\boldsymbol{\beta}^{(i)}$ are linearly independent for each i, and the entries of the $\boldsymbol{\beta}^{(i)}$ are contained in trivially intersecting subspaces $\mathcal{B}^{(i)}$.

- For a position $j \in [n_l - b]$ in the i-th block and an element $\boldsymbol{\beta} \in \mathcal{S}$, there are the following s elements in $\mathcal{S}$: Choose $s-1$ elements $\gamma_2, \ldots, \gamma_s$ that expand the basis $\beta_1^{(i)}, \ldots, \beta_{n_l-b}^{(i)}$ (which spans an (n_l-b)-dimensional subspace) to the $(n_l - b + s - 1)$-dimensional subspace $\mathcal{B}^{(i)}$. Then, the s vectors

$$\begin{aligned}\boldsymbol{\beta}^{(i)} =: \boldsymbol{\beta}_{(1)}^{(i)} &= \begin{pmatrix} \beta_1^{(i)} & \ldots & \beta_{j-1}^{(i)} & \beta_j^{(i)} & \beta_{j+1}^{(i)} & \beta_{n_l-b}^{(i)} \end{pmatrix} \\ \boldsymbol{\beta}_{(2)}^{(i)} &= \begin{pmatrix} \beta_1^{(i)} & \ldots & \beta_{j-1}^{(i)} & \gamma_2 & \beta_{j+1}^{(i)} & \beta_{n_l-b}^{(i)} \end{pmatrix} \\ &\vdots \\ \boldsymbol{\beta}_{(s)}^{(i)} &= \begin{pmatrix} \beta_1^{(i)} & \ldots & \beta_{j-1}^{(i)} & \gamma_s & \beta_{j+1}^{(i)} & \beta_{n_l-b}^{(i)} \end{pmatrix}\end{aligned}$$

  are all in $\mathcal{S}^{(i)}$. Hence, the vectors

$$\begin{gathered}\begin{pmatrix} \boldsymbol{\beta}^{(1)} \mid \ldots \mid \boldsymbol{\beta}^{(i-1)} \mid \boldsymbol{\beta}_{(1)}^{(i)} \mid \boldsymbol{\beta}^{(i+1)} \mid \ldots \mid \boldsymbol{\beta}^{(\mu)} \end{pmatrix} \\ \begin{pmatrix} \boldsymbol{\beta}^{(1)} \mid \ldots \mid \boldsymbol{\beta}^{(i-1)} \mid \boldsymbol{\beta}_{(2)}^{(i)} \mid \boldsymbol{\beta}^{(i+1)} \mid \ldots \mid \boldsymbol{\beta}^{(\mu)} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \boldsymbol{\beta}^{(1)} \mid \ldots \mid \boldsymbol{\beta}^{(i-1)} \mid \boldsymbol{\beta}_{(s)}^{(i)} \mid \boldsymbol{\beta}^{(i+1)} \mid \ldots \mid \boldsymbol{\beta}^{(\mu)} \end{pmatrix}\end{gathered}$$

  are all in $\mathcal{S}$ and differ only in position j in the i-th block. The entries $\beta_j^{(i)}, \gamma_2, \ldots, \gamma_s$ in the j-th position in the i-th block are linearly independent over $\mathbb{F}_q$ by construction.

Furthermore, we have $\mathcal{S} = \mathcal{S} \cdot \operatorname{diag}(\mathbf{G}_1, \ldots, \mathbf{G}_\mu) \coloneqq \{\boldsymbol{\beta} \cdot \operatorname{diag}(\mathbf{G}_1, \ldots, \mathbf{G}_\mu) \mid \boldsymbol{\beta} \in \mathcal{S}\}$

for all invertible matrices $\mathbf{G}_i \in \mathbb{F}_q^{(n_l-b)\times(n_l-b)}$. To see this, consider the following: multiplying a subblock $\boldsymbol{\beta}^{(i)}$ with an invertible matrix $\mathbf{G}_i$ from the right gives another basis of the same subspace—hence $\boldsymbol{\beta}^{(i)}\mathbf{G}_i \in \mathcal{S}^{(i)}$ and $\boldsymbol{\beta} \cdot \mathrm{diag}(\mathbf{G}_1, \dots, \mathbf{G}_\mu) \in \mathcal{S}$ for all $\boldsymbol{\beta} \in \mathcal{S}$. Since the $\mathbf{G}_i$ are invertible, the mapping $\boldsymbol{\beta} \mapsto \boldsymbol{\beta} \cdot \mathrm{diag}(\mathbf{G}_1, \dots, \mathbf{G}_\mu)$ is bijective.

These two observations imply that a matrix with the elements of $\mathcal{S}$ as rows has the scrambled YB grouping property as in Definition 4.8. □

By combining Theorems 4.1 and 4.2, we get the following existence result for a globally MSR PMDS code.

Corollary 4.1. *Let μ, n_l, b, s be valid PMDS parameters. There is a globally MSR PMDS code with field size*

$$(n_l - 1)^{\mu(n_l-b+s-1)} \leq q^M < [2(n_l-1)]^{\mu(n_l-b+s-1)}$$

and subpacketization

$$\ell = \left(\begin{bmatrix} n_l - b + s - 1 \\ n_l - b \end{bmatrix}_q \prod_{i=0}^{n_l-b-1} \left(q^{n_l-b} - q^i \right) \right)^\mu \leq 4^\mu q^{\mu(n_l-b)(n_l-b+s-1)}.$$

Proof. The corollary follows directly from using the matrix $\mathbf{B}$ constructed in Theorem 4.2 in Construction 4.1 (see Theorem 4.1). Choosing q as the smallest prime power $\geq n_l - 1$ ensures that there is an $[n_l, n_l - b]_q$ MDS code as required in Construction 4.1. The statement follows from observing that, there is a prime power[3] q with $n_l - 1 \leq q < 2(n_l - 1)$. □

Remark 4.4. *There are no globally MSR codes in the literature that we can compare the new construction with. Therefore, we only compare the field size and subpacketization to a PMDS code without the globally MSR property, as well as the subpacketization of an MSR code with the same parameters after puncturing b positions in each local repair set. In other words, we determine how much we "pay" in terms of field size and subpacketization if we go from a purely PMDS or MSR code to a globally MSR PMDS code.*

Construction 4.1 is an adaption of the Gabidulin-based PMDS code construction in [RKSV13] (without local or global regeneration), which has field size $q^M < [2(n_l - 1)]^{\mu(n_l-b)}$. Compared to such a PMDS code, the exponent in the field size in Corollary 4.1 is larger by a factor $1 + \frac{s-1}{n_l-b}$. This difference is significant if the number of global parities is large (recall that $1 \leq s \leq \mu(n_l - b)$). Hence, we pay more in field size for the globally MSR property if there are many global parities. It appears possible to adapt other PMDS constructions, such as [MPK19] or [GYBS18], to have the globally MSR property as well. Such a construction may reduce the field size significantly.

[3] Trivially, there is a power of two in this range. Further, by Bertrand's postulate, there is even a prime number within this range.

Compared to a Ye–Barg MSR code with parameters $[\mu(n_l-b), \mu(n_l-b)-s;\ell]$ *(which are the code parameters after puncturing b positions in each group of a PMDS code) with subpacketization* $[(d+1-n_l+b)\mu(n_l-b)]^{\mu(n_l-b)}$*, the subpacketization of the globally MSR PMDS code in Corollary 4.1 is larger by roughly a factor* $(n_l-b+s-1)$ *in the exponent. Hence, the exponent of the subpacketization is in* $O(\mu n_l)$ *without the PMDS property and in* $O(\mu n_l(n_l+s))$ *for a globally MSR PMDS code.*

Remark 4.5 (Global Regeneration in Grid-Like Topologies)**.** *While the presented construction is specific to PMDS codes, the same approach can be applied to codes for grid-like topologies*[4]*. The PMDS code construction of [RKSV13] can be viewed as a special case of the construction of codes for grid-like topologies presented in Section 3.5 (see Example 3.3). As shown in Theorem 3.5, this construction results in a code* $\mathcal{C} \in \mathbb{C}^{\mathsf{MR}}_{n_1\times n_2}(b_1,b_2,s)$ *that, when punctured in the positions of an erasure pattern* $\mathcal{E}' \in \mathbb{E}^{\max}_{n_1\times n_2}(b_1,b_2,0)$*, again results in a Gabidulin code. To obtain a code where these surviving positions are also a codeword of an MSR code, Construction 4.1 can equivalently be applied by replacing the diagonal matrix* $\mathrm{diag}(\mathbf{G}_{\mathsf{MDS}},\mathbf{G}_{\mathsf{MDS}},\ldots)$*, where* $\mathbf{G}_{\mathsf{MDS}}$ *is a generator matrix of an* $[n_l, n_l-b]$ *MDS codes*[5]*, with a generator matrix of an arbitrary code* $\mathcal{C} \in \mathbb{C}_{n_1\times n_2}(b_1,b_2,0)$ *over* $\mathbb{F}_q$*.*

Let $\mathbf{B} \in \mathbb{F}_{q^M}^{\ell\times\mu(n_l-b)}$ *be a matrix where each row contains elements that are linearly independent over* $\mathbb{F}_q$*. Consider the* $[n_1n_2,(n_1-b_1)(n_2-b_2)-s;\ell]_{q^M}$ *code*

$$\mathcal{C} = \left\{\mathbf{C} \in \mathbb{F}_q^{\ell\times n_1n_2} \mid \mathbf{C}[a,:] = \mathbf{u}^{(a)}\cdot\mathbf{G}_{\mathbf{B}}^{(a)}\cdot\mathbf{G}_0 \,\forall\, \mathbf{u}^{(a)} \in \mathbb{F}_{q^m}^{(n_1-b_1)(n_2-b_2)-s}, a \in [\ell]\right\},$$

where $\mathbf{G}_{\mathbf{B}}^{(a)}$ *is a generator matrix of the code* $\mathsf{Gab}((n_1-b_1)(n_2-b_2),s,\mathbf{B}[a,:])$ *as in Definition 2.6 and* $\mathbf{G}_0$ *is a generator matrix of a code in* $\mathbb{C}^{\mathsf{MR}}_{n_1\times n_2}(b_1,b_2,0)$ *over* $\mathbb{F}_q$*. It is easy to see that each row of the codewords in* $\mathcal{C}$*, and thereby also the entire array code* $\mathcal{C}$*, is in* $\mathbb{C}^{\mathsf{MR}}_{n_1\times n_2}(b_1,b_2,s)$ *(see Theorem 3.5 on Page 46). To obtain a code with global regeneration, similar to Definition 4.3, the difficulty again lies in guaranteeing the YB-grouping property for every possible puncturing pattern. Specifically, we require that the matrix* $\mathbf{B}\cdot\mathbf{G}_0[:,\bar{\mathcal{E}}']$ *has the YB-grouping property as in Definition 4.7, for the complement* $\bar{\mathcal{E}}' := [n_1n_2]\setminus\mathcal{E}'$ *of any erasure pattern* $\mathcal{E}' \in \mathbb{E}^{\max}_{n_1\times n_2}(b_1,b_2,0)$*. However, while for PMDS codes these patterns are easily characterized as those where exactly* n_l-b *positions survive in each local repair set (see Definition 2.8), their characterization is significantly more difficult for grid-like topologies and in many cases unknown (see Table 3.1 on Page 31). Consequently, in the grid-like setting, we are not able to exploit the structure of* $\mathbf{G}_0[:,\bar{\mathcal{E}}']$ *as done in Theorem 4.2 and have to resort to a more generic approach: Choose the rows of* $\mathbf{B}$ *as all bases of all* $(n_1-b_1)(n_2-b_2)-s$ *dimensional subspaces of* $\mathbb{F}_q^{(n_1-b_1)(n_2-b_2)}$*. By the same arguments as in Theorem 4.2*

[4] This remark is given to highlight the connection to codes for grid-like topologies, as considered Chapter 3, and was not included in [HPYWZ21].

[5] Note that this block diagonal matrix spans a code of $\mathbb{C}^{\mathsf{MR}}_{n_1\times n_2}(0,b_2,0)$ with $n_1=\mu$, $n_2=n_l$, and $b_2=b$.

the multiplication by any full-rank matrix over $\mathbb{F}_q$ preserves the YB-grouping property. However, while this method applies to the more general setting of grid-like topologies, it results in a significantly larger subpacketization and is therefore not discussed in more detail here.

4.4 PMDS Codes with Local Regeneration

For the remainder of this chapter we consider PMDS codes with local regeneration, specifically, PMDS codes where each local code is a d-MSR code. We introduce several new constructions, each obtained as a combination of the Ye-Barg MSR code (see Definition 4.5) and a different underlying PMDS code.

4.4.1 Locally Regenerating PMDS and Sector-Disk Codes with Two Global Parities

We begin by constructing array codes from the PMDS codes of [BPSY16] using the ideas of [YB17a] to obtain locally d-MSR PMDS codes. Since the PMDS code construction in [BPSY16] can be easily turned into an SD code (see Remark 2.3), we also include the respective construction of SD codes with local d-MSR codes in this section. While SD fulfill a weaker definition of locality than PMDS codes, the construction is favorable in terms of required field size.

To apply the ideas of [YB17a] when constructing locally MSR PMDS and SD codes, we need the local codes to be RS codes with specific code locators. The construction of PMDS codes given in [BPSY16] has the property that the local codes are RS codes, but the code locators are fixed to be the first n_l powers of some element β of sufficient order. We generalize this construction to allow for different choices of code locators for the local codes.

Let q be a power of 2 and $\beta \in \mathbb{F}_q$ be an element with $\mathrm{order}(\beta) \geq \mu N$. The $[\mu n_l, \mu(n_l - b) - 2]$ code $\mathcal{C}(\mu, n_l, b, 2, \mathcal{L}, N)$ is given by the $(b\mu + 2) \times \mu n_l$ parity-check matrix

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_0 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_0 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{H}_0 \\ \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_\mu \end{pmatrix}, \tag{4.3}$$

where

$$\mathbf{H}_0 = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \beta^{i_1} & \beta^{i_2} & \dots & \beta^{i_{n_l}} \\ \beta^{2i_1} & \beta^{2i_2} & \dots & \beta^{2i_{n_l}} \\ \vdots & \vdots & \ddots & \vdots \\ \beta^{(b-1)i_1} & \beta^{(b-1)i_2} & \dots & \beta^{(b-1)i_{n_l}} \end{pmatrix}$$

for $\mathcal{L} = \{i_1, i_2, \dots, i_{n_l}\}$ and, for $0 \leq j \leq \mu - 1$,

$$\mathbf{H}_{j+1} = \begin{pmatrix} \beta^{bi_1} & \beta^{bi_2} & \dots & \beta^{bi_{n_l}} \\ \beta^{-jN-i_1} & \beta^{-jN-i_2} & \dots & \beta^{-jN-i_{n_l}} \end{pmatrix} .$$

Note that this generalization includes both [BPSY16, Construction A] and [BPSY16, Construction B] as special cases:

$$\mathcal{C}_A = \mathcal{C}(\mu, n_l, b, 2, [0, n_l - 1], n_l)$$

and

$$\mathcal{C}_B = \mathcal{C}(\mu, n_l, b, 2, [0, n_l - 1], N_B)$$

for $N_B = (b+1)(n_l - 1 - b) + 1$.

We now derive a general, sufficient condition on N, based on the set $\mathcal{L}$, such that the code is a PMDS code.

Lemma 4.2. *Let μ, n_l, b and $s = 2$ be valid PMDS parameters and $\mathcal{L}$ be a set of nonnegative integers with $|\mathcal{L}| = n_l$. Then, the code $\mathcal{C}(\mu, n_l, b, 2, \mathcal{L}, N)$ is a PMDS code for any $N \geq (b+1)(\max_{i \in \mathcal{L}} i - b) + 1$.*

Proof. We follow the proofs of [BPSY16, Theorem 5] and [BPSY16, Theorem 7]. The difference to the construction above is that in [BPSY16], the powers $\mathcal{L} = \{i_1, \dots, i_{n_l}\}$ are consecutive, i.e., $i_j = j - 1$. This results in a slightly more technical proof.

Assume b positions in each local repair set have been erased and in addition there are 2 erasures in arbitrary positions. If the two erasures occur in the same local repair set z, all local repair sets except for this one will be corrected by the local codes. Assume the erasures in local repair set z occurred in positions $\mathcal{E}_z \subset [n_l]$. Since all points in $\mathcal{L}$ are distinct, by the same argument as in [BPSY16], the erased positions can be recovered uniquely if the matrix

$$\hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_z \end{pmatrix}$$

restricted to the erased positions $\mathcal{E}$ is of full rank. Assume that the erased positions are those corresponding to the powers $\{j_1, j_2, \dots, j_{b+2}\} \subset \mathcal{L}$. It is easy to see that this matrix $\hat{\mathbf{H}}_{\mathcal{E}}$ can be transformed into a Vandermonde matrix by multiplying the last row by $\beta^{(z-1)N}$ and column corresponding to j_ξ by β^{j_ξ} for all $\xi \in [b+2]$. Therefore, it is of full rank and the erasures can be corrected.

Now consider the case of two local repair sets with $b+1$ erasures each. Assume, without loss of generality, that the positions corresponding to the powers $\{j_1, \dots, j_{b+1}\} \subset \mathcal{L}$ are erased in local repair set 1 and those corresponding to $\{j'_1, \dots, j'_{b+1}\} \subset \mathcal{L}$ in local repair set $z+1$ with $1 \leq z \leq \mu - 1$. Define the matrix

$$\mathbf{F}(j_1, \dots, j_{b+1}; j'_1, \dots, j'_{b+1}; b; N; z) = \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ \alpha^{j_1} & \dots & \alpha^{j_{b+1}} & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{(b-1)j_1} & \dots & \alpha^{(b-1)j_{b+1}} & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 \\ 0 & \dots & 0 & \alpha^{j'_1} & \dots & \alpha^{j'_{b+1}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \alpha^{(b-1)j'_1} & \dots & \alpha^{(b-1)j'_{b+1}} \\ \alpha^{bj_1} & \dots & \alpha^{bj_{b+1}} & \alpha^{bj'_1} & \dots & \alpha^{bj'_{b+1}} \\ \alpha^{-j_1} & \dots & \alpha^{-j_{b+1}} & \alpha^{-Nz-j'_1} & \dots & \alpha^{-Nz-j'_{b+1}} \end{pmatrix} .$$

To show that the erased positions can be recovered, we need to show that this matrix is invertible. By [BPSY16, Lemma 3] this is true if

$$Nz + \sum_{u=1}^{b+1} j'_u - \sum_{u=1}^{b+1} j_u \neq 0 \mod \operatorname{order}(\beta) . \tag{4.4}$$

Note that [BPSY16, Lemma 3] shows this relation only for $0 \leq j_1 < j_2 < \dots < j_{b+1} \leq n_l - 1$ and $0 \leq j'_1 < j'_2 < \dots < j'_{b+1} \leq n_l - 1$. However, it is easy to check that the result is independent of the specific values and only depends on the sums $\sum_{u=1}^{b+1} j_u$ and $\sum_{u=1}^{b+1} j'_u$. Since, by definition, the powers $j_1, j_2, \dots, j_{b+1}$ are distinct integers, their sum is lower bounded by

$$\frac{b(b+1)}{2} = \sum_{u=0}^{b} u \leq \sum_{u=1}^{b+1} j_u \tag{4.5}$$

and upper bound by

$$\begin{aligned} \sum_{u=1}^{b+1} j_u &\leq \sum_{u=0}^{b} (\max_{j \in \mathcal{L}} j - b) + u \\ &= (b+1)(\max_{j \in \mathcal{L}} j - b) + \sum_{u=0}^{b} u = N - 1 + \frac{b(b+1)}{2} . \end{aligned} \tag{4.6}$$

The same bounds hold for the powers $j'_1, j'_2, \ldots, j'_{b+1}$. Combining Eqs. (4.5) and (4.6) we get

$$-(N-1) \leq \sum_{u=1}^{b+1} j'_u - \sum_{u=1}^{b+1} j_u \leq N-1.$$

Applying these bounds to Eq. (4.4) (recall that $1 \leq z \leq \mu - 1$) gives

$$1 = N - (N-1) \leq Nz + \sum_{u=1}^{b+1} j'_u - \sum_{u=1}^{b+1} j_u \leq N(\mu-1) + (N-1) = N\mu - 1 < \mathrm{order}(\beta) \ ,$$

where the final inequality holds by definition of β. It follows that Eq. (4.4) is fulfilled and the lemma statement holds. □

By similar arguments we also give a general, sufficient condition on N for the code to be an SD code.

Lemma 4.3. *Let μ, n_l, b and $s = 2$ be valid PMDS parameters and $\mathcal{L}$ be any set of nonnegative integers with $|\mathcal{L}| = n_l$. Then, the code $\mathcal{C}(\mu, n_l, b, 2, \mathcal{L}, N)$ is an SD code for any $N \geq \max_{j\in\mathcal{L}} j + 1$.*

Proof. The case of $b+2$ erasures in the same local repair set (horizontal code) is the same as in Lemma 4.2 and [BPSY16, Theorem 5]. Now consider the case of b column erasures in positions $\{j_1, \ldots, j_b\} \subset \mathcal{L}$ and an additional erasure in each of the local repair sets $z+1$ and $z'+1$, with $0 \leq z < z' \leq \mu - 1$, in positions $j, j' \in \mathcal{L} \setminus \{j_1, \ldots, j_b\}$. By the same argument as in [BPSY16, Theorem 5] we need to show that $\beta^{-j} + \beta^{-N(z-z')-j'}$ is invertible. With $1 \leq z, z' \leq \mu$ and $0 \leq j, j' \leq N-1$ we get

$$N(z'-z) + j' - j \geq N + j' - j \geq N - (N-1) > 0$$

and

$$N(z'-z) + j' - j \leq N(\mu-1) + N - 1 = N\mu - 1 < \mathrm{order}(\beta) \ .$$

Combining these we get $1 \leq N(z'-z) + j' - j \leq N\mu - 1$, so

$$N(z'-z) + j' - j \neq 0 \mod \mathrm{order}(\beta)$$

and it follows that $\beta^{-j} + \beta^{-N(z-z')-j'}$ is invertible. □

With these generalizations of [BPSY16, Construction A/B] we are now ready to construct PMDS and SD codes, where each local code is a d-MSR code.

Construction 4.2 (Locally d-MSR PMDS/SD array codes)**.** *Let $s = 2$ and $q, \mu, n_l, b, d, N \in \mathbb{Z}_{>0}$ be positive integers with*

- *$b \leq n_l$*
- *q a power of 2*
- *$q \geq \max\{\mu N, z n_l\} + 1$, where $z = d + 1 - (n_l - b)$*
- *$\ell = z^{n_l}$*

For an element $\beta \in \mathbb{F}_q$ with $\operatorname{order}(\beta) \geq \max\{\mu N, z n_l\}$ denote $\beta_{i,j} = \beta^{(i-1)n_l + j - 1}$ for $i \in [z], j \in [n_l]$.

We define the $[\mu n_l, \mu(n_l - b) - 2; \ell]_{q^M}$ array code $\mathcal{C}(\mu, n_l, b, 2, N, d; \ell)_q$ as

$$\left\{ \mathbf{C} \in \mathbb{F}_q^{\ell \times \mu n_l} \mid \mathbf{H}^{(a)} \cdot (\mathbf{C}[a,:])^\top = \mathbf{0} \, \forall \, a \in [\ell] \right\} .$$

The matrix $\mathbf{H}^{(a)}$ is defined as

$$\mathbf{H}^{(a)} = \begin{pmatrix} \mathbf{H}_0^{(a)} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_0^{(a)} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{H}_0^{(a)} \\ \mathbf{H}_1^{(a)} & \mathbf{H}_2^{(a)} & \dots & \mathbf{H}_\mu^{(a)} \end{pmatrix} \in \mathbb{F}_q^{b\mu + 2 \times \mu n_l} ,$$

where

$$\mathbf{H}_0^{(a)} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \beta_{a_1,1} & \beta_{a_2,2} & \dots & \beta_{a_{n_l},n_l} \\ \vdots & \vdots & & \vdots \\ \beta_{a_1,1}^{b-1} & \beta_{a_2,2}^{b-1} & \dots & \beta_{a_{n_l},n_l}^{b-1} \end{pmatrix} \in \mathbb{F}_q^{b \times n_l} , \tag{4.7}$$

with $a \in [\ell]$ and $a - 1 = \sum_{i=1}^{n_l} (a_i - 1) z^{i-1}$ with $a_i \in [z]$. For $0 \leq j \leq \mu - 1$ let

$$\mathbf{H}_{j+1}^{(a)} = \begin{pmatrix} \beta_{a_1,1}^b & \beta_{a_2,2}^b & \dots & \beta_{a_{n_l},n_l}^b \\ \beta^{-jN} \beta_{a_1,1}^{-1} & \beta^{-jN} \beta_{a_2,2}^{-1} & \dots & \beta^{-jN} \beta_{a_{n_l},n_l}^{-1} \end{pmatrix} \in \mathbb{F}_q^{2 \times n_l} .$$

It remains to show that the local codes are MSR codes and the conditions under which the code is a PMDS or SD code.

Theorem 4.3. *Let μ, n_l, b and $s = 2$ be valid PMDS parameters, d be an integer with $n_l - b \leq d \leq n_l - 1$, and $q > \max\{\mu N, z n_l\}$, where $z = d + 1 - (n_l - b)$ and*

$$N = (b+1)(b n_l - 1 - b) + 1 .$$

Then the array code $\mathcal{C}(\mu, n_l, b, 2, N, d; \ell)_q$, as in Construction 4.2, is a locally d-MSR

$\mathsf{PMDS}(\mu, n_l, b, 2, \mathbb{W}; b^{n_l})$ *code over* $\mathbb{F}_q$*, as in Definition 4.4, for* $\mathbb{W} = \{\mathcal{W}_1, \ldots, \mathcal{W}_\mu\}$ *with* $\mathcal{W}_i = [(i-1)n_l + 1, in_l]$*.*

Proof. First, note that the $\beta_{i,j}$ in Construction 4.2 are given by $\beta^0, \beta^1, \ldots, \beta^{bn_l-1}$. As $\mathrm{order}(\beta) \geq zn_l$ (which implies $q > zn_l$) these β^i are distinct. Now consider the j-th local repair set. The a-th row fulfills the parity check equations given in Eq. (4.7) and since all elements $\beta_{i,j}$ are distinct, it is immediate that the local repair set is an $[n_l, n_l - b; b^{n_l}]$ Ye-Barg code as in Definition 4.5.

For the PMDS property, observe that the a-th row, i.e., the row fulfilling the parity-check equations $\mathbf{H}^{(a)}$, is a code $\mathcal{C}(\mu, n_l, b, 2, \mathcal{L}^{(a)}, N)$ as in Lemma 4.2, where $\mathcal{L}^{(a)} = \{(a_i - 1)n_l + i - 1 \mid i \in [n_l]\}$ by definition of the $\beta_{i,j}$. For any a it holds that

$$\max_{i \in \mathcal{L}^{(a)}} i \leq \max_{\substack{i \in \mathcal{L}^{(a)} \\ a \in [\ell]}} i = bn_l - 1 \ .$$

By Lemma 4.2 the code is PMDS if $N > (b+1)(\max_{i\in\mathcal{L}} i - b)$ and the lemma statement follows. □

Corollary 4.2. *Let* μ, n_l, b *and* $s = 2$ *be valid PMDS parameters,* q *be a power of* 2*,* d *be an integer with* $n_l - b \leq d \leq n_l - 1$*, and* $z = d + 1 - (n_l - b)$*. Then, there is a* d*-MSR PMDS code over* $\mathbb{F}_q$ *of field size*

$$\mu b(bn_l - b + n_l - 2) + 1 \leq q \leq 2\mu b(bn_l - b + n_l - 2)$$

and subpacketization $\ell = [d + 1 - (n_l - b)]^{n_l}$*.*

Proof. We use Theorem 4.3 and derive bounds on the smallest field size q satisfying the bound $q > \max\{\mu N, zn_l\}$ with $N = (b+1)(bn_l - 1 - b) + 1 = b(bn_l - b + n_l - 2)$.

First note that $1 \leq z = d + 1 - (n_l - b) \leq b$ for the valid choices of d. Furthermore, note that $b \geq 1$ and $n_l \geq b + 1 \geq 2$. Thus, we have

$$\mu N = \mu b(bn_l - b + n_l - 2) = \mu\Big[bn_l + \underbrace{b^2(n_l - 1) - 2b}_{\geq -1}\Big] \geq \mu(bn_l - 1) \geq bn_l \geq zn_l.$$

Hence, we in fact only require $q > \mu N$. Trivially, there is a power of two between $\mu N + 1$ and $2\mu N$, which proves the claim. □

For completeness, we provide a similar statement for d-MSR SD codes.

Theorem 4.4. *Let* μ, n_l, b *and* $s = 2$ *be valid PMDS parameters and* q *be a power of* 2 *with* $q > \max\{bn_l\mu, zn_l\}$*. Then the array code* $\mathcal{C}(\mu, n_l, b, 2, bn_l, d; \ell)_q$ *as in Construction 4.2 is a locally* d*-MSR* $\mathsf{SD}(\mu, n_l, b, s, \mathbb{W}; z^{n_l})$ *code over* $\mathbb{F}_q$*, for* $\mathbb{W} = \{\mathcal{W}_1, \ldots, \mathcal{W}_\mu\}$ *with* $\mathcal{W}_i = [(i-1)n_l + 1, in_l]$*.*

Proof. The proof follows immediately from the proof of Theorem 4.3 and Lemma 4.3. □

Remark 4.6. *It is easy to check that by removing the last row of the parity-check matrix as in Eq.* (4.3) *of the PMDS codes in [BPSY16], we obtain a PMDS code with one global parity (*$s = 1$*). By the same operation on all the parity-check matrices for the rows of the d-MSR PMDS code in Construction 4.2, we obtain a locally d-MSR PMDS codes with one global parity. We do not discuss this case in detail since the resulting codes have the same field size as the ones with two global parities.*

4.4.2 Universal PMDS Codes with Local Row-Wise MDS MSR Codes

In this section, we present a general technique for constructing PMDS codes with MSR local codes, by combining an arbitrary row-wise MDS MSR code (see Definition 4.2) with a universal PMDS code family. The latter notion was first defined in [MPK19], and we formalize it below in Definition 4.9. Roughly speaking, a universal PMDS code family arises from a PMDS construction in which the local code can be chosen arbitrarily as the $\mathbb{F}_{q^M}$-span of an $\mathbb{F}_q$-linear MDS code. Although the universality requirement seems to be strong, there are several PMDS constructions in the literature that fulfill this property, for instance [RKSV13; MPK19] (cf. the overview in [MPK19]). For the construction of [GYBS18], we show its universality in Section 4.4.2. Hence, some of the PMDS constructions with the smallest field sizes in the literature have this property, which enables the new general construction to achieve rather small field sizes as well. Note that the PMDS construction with local regeneration in Section 4.4.1 is not of the type presented here, since the PMDS family in [BPSY16] is not universal due to strong dependencies between the choice of the local and global parities.

A General Code Construction

The following definition formalizes the notion of a universal PMDS code family, which was introduced in [MPK19].

Definition 4.9 (Universal Partial MDS code family)**.** *Let* μ, n_l, b, s *be valid PMDS parameters. A family of codes is a universal PMDS code family* $\mathcal{F}_{\mathsf{PMDS}}(\mu, n_l, b, s)$ *over* $\mathbb{F}_{q^M}$ *if there is a partition* $\mathbb{W} = \{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_\mu\}$ *(fixed for the entire family) such that*

- *every code* $\mathcal{C} \in \mathcal{F}_{\mathsf{PMDS}}(\mu, n_l, b, s)$ *is a* $\mathsf{PMDS}(\mu, n_l, b, s, \mathbb{W}; 1)$ *code over* $\mathbb{F}_{q^M}$ *and*
- *for any MDS code* $\mathcal{C}_{\mathsf{local}}[n_l, n_l - b, b + 1]$ *over* $\mathbb{F}_q$*, there is exactly one* $\mathcal{C} \in \mathcal{F}_{\mathsf{PMDS}}(\mu, n_l, b, s)$ *such that* $\mathcal{C}|_{\mathcal{W}_i} = \langle \mathcal{C}_{\mathsf{local}} \rangle_{\mathbb{F}_{q^M}} \simeq \mathcal{C}_{\mathsf{local}}^{\times M}$ *for all* $i = 1, \ldots, \mu$*. We denote this unique code by* $\mathcal{F}(\mathcal{C}_{\mathsf{local}}) \coloneqq \mathcal{C}$*, i.e.,* $\mathcal{F}(\cdot)$ *can be seen as an injective mapping between the set of MDS codes over* $\mathbb{F}_q$ *and the family* $\mathcal{F}$*.*

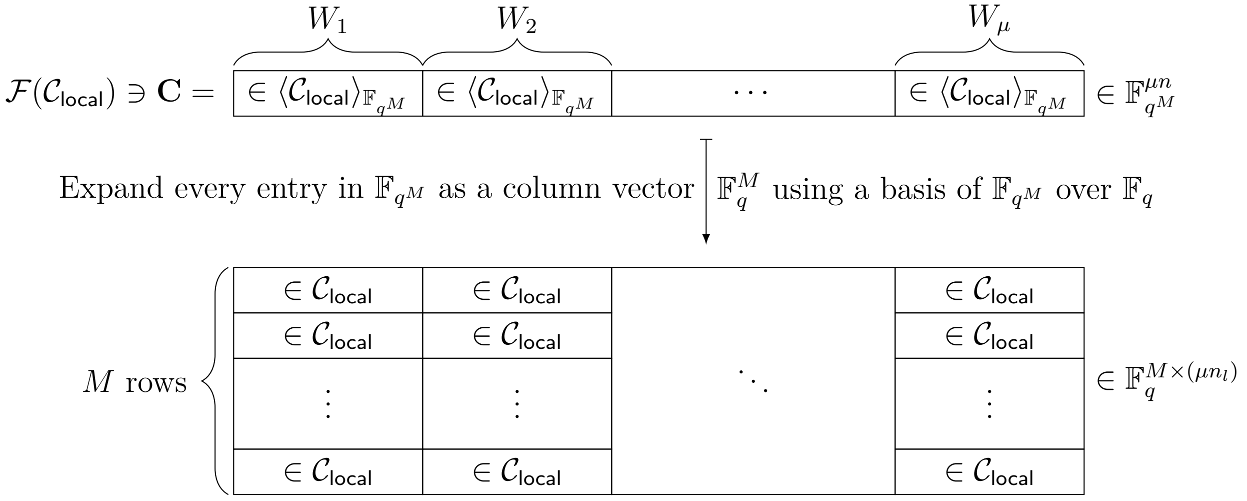

W_1
W_2
W_μ
$\mathcal{F}(\mathcal{C}_{\mathsf{local}}) \ni \mathbf{C} =$
$\in \langle \mathcal{C}_{\mathsf{local}} \rangle_{\mathbb{F}_{q^M}}$
$\cdots$
$\in \mathbb{F}_{q^M}^{\mu n}$
Expand every entry in $\mathbb{F}_{q^M}$ as a column vector $\mathbb{F}_q^M$ using a basis of $\mathbb{F}_{q^M}$ over $\mathbb{F}_q$
M rows
$\in \mathcal{C}_{\mathsf{local}}$
$\vdots$
$\ddots$
$\in \mathbb{F}_q^{M \times (\mu n_l)}$

Figure 4.2: Illustration of the codeword structure of the PMDS code $\mathcal{F}(\mathcal{C}_{\mathsf{local}})$ in Definition 4.9.

Note that that the $\mathbb{F}_{q^M}$-span of the $\mathbb{F}_q$ local code can be viewed as a homogeneous interleaved code (see Definition 2.2) with the same parameters, as observed in Corollary 2.1. An illustration of the expansion of a PDMS codeword as in Definition 4.9 is given in Fig. 4.2.

With these definitions established, we are now ready to present a construction of locally MSR PMDS codes by combining a family of universal PMDS codes with a row-wise MDS MSR code.

Construction 4.3. *Let μ, n_l, b, s be valid PMDS parameters and $\mathcal{F}_{\mathsf{PMDS}}(\mu, n_l, b, s)$ be a universal PMDS code family. Let $\mathcal{C}_{\mathsf{MSR}}[n_l, n_l - b; \ell]$ be a row-wise MDS (h,d)-MSR code and denote by $\mathcal{C}_{\mathsf{MSR}}^{(a)}$ the MDS code in its a-th row for $a \in [\ell]$. We define the code*

$$\mathcal{F}_{\mathsf{PMDS}}(\mathcal{C}_{\mathsf{MSR}}) := \left\{ \mathbf{C} \in \mathbb{F}_{q^M}^{\ell \times \mu n_l} \;\middle|\; \mathbf{C}[a,:] \in \mathcal{F}(\mathcal{C}_{\mathsf{MSR}}^{(a)}) \,\forall\, a \in [\ell] \right\} \ .$$

Note that the code is well-defined, as by Definition 4.9 the family of PMDS codes contains a PMDS code $\mathcal{F}(C_{\mathsf{MSR}}^{(a)})$ for *any* MDS code $C_{\mathsf{MSR}}^{(a)}$ over $\mathbb{F}_q$.

Theorem 4.5. *The code $\mathcal{F}_{\mathsf{PMDS}}(\mathcal{C}_{\mathsf{MSR}})$ in Construction 4.3 is a locally (h,d)-MSR $\mathsf{PMDS}(\mu, n_l, b, s, \mathbb{W}; \ell)$ code over $\mathbb{F}_{q^M}$, for a partition $\mathbb{W} = \{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_\mu\}$ of $[\mu n_l]$ with $|\mathcal{W}_i| = n_l \ \forall\ i \in [\mu]$.*

Proof. By construction, the codewords of $\mathcal{F}_{\mathsf{PMDS}}(\mathcal{C}_{\mathsf{MSR}})$ are matrices whose rows are contained in a PMDS code of the family $\mathcal{F}_{\mathsf{PMDS}}$. In particular, the PMDS code in the a-th row has the MDS code $C_{\mathsf{MSR}}^{(a)}$ as its local code. If we puncture all rows in all positions but $\mathcal{W}_i$ for some $i \in [\mu]$, we obtain in the a-th row the code $C_{\mathsf{MSR}}^{(a)}$. Hence, for any $i \in [\mu]$ we have

$$\mathcal{F}_{\mathsf{PMDS}}(\mathcal{C}_{\mathsf{MSR}})|_{\mathcal{W}_i} = \left\{ \begin{pmatrix} \mathbf{C}^{(1)}[1,:] \\ \vdots \\ \mathbf{C}^{(1)}[M,:] \\ \mathbf{C}^{(2)}[1,:] \\ \vdots \\ \mathbf{C}^{(\ell)}[M,:] \end{pmatrix} \in \mathbb{F}_q^{M\ell \times n_l} \;\middle|\; \mathbf{C}^{(a)}[j,:] \in C_{\mathsf{MSR}}^{(a)} \, \forall a \in [\ell],\, j \in [M] \right\}$$
$$\simeq \underbrace{\mathcal{C}_{\mathsf{MSR}} \times \cdots \times \mathcal{C}_{\mathsf{MSR}}}_{M \text{ times}}$$

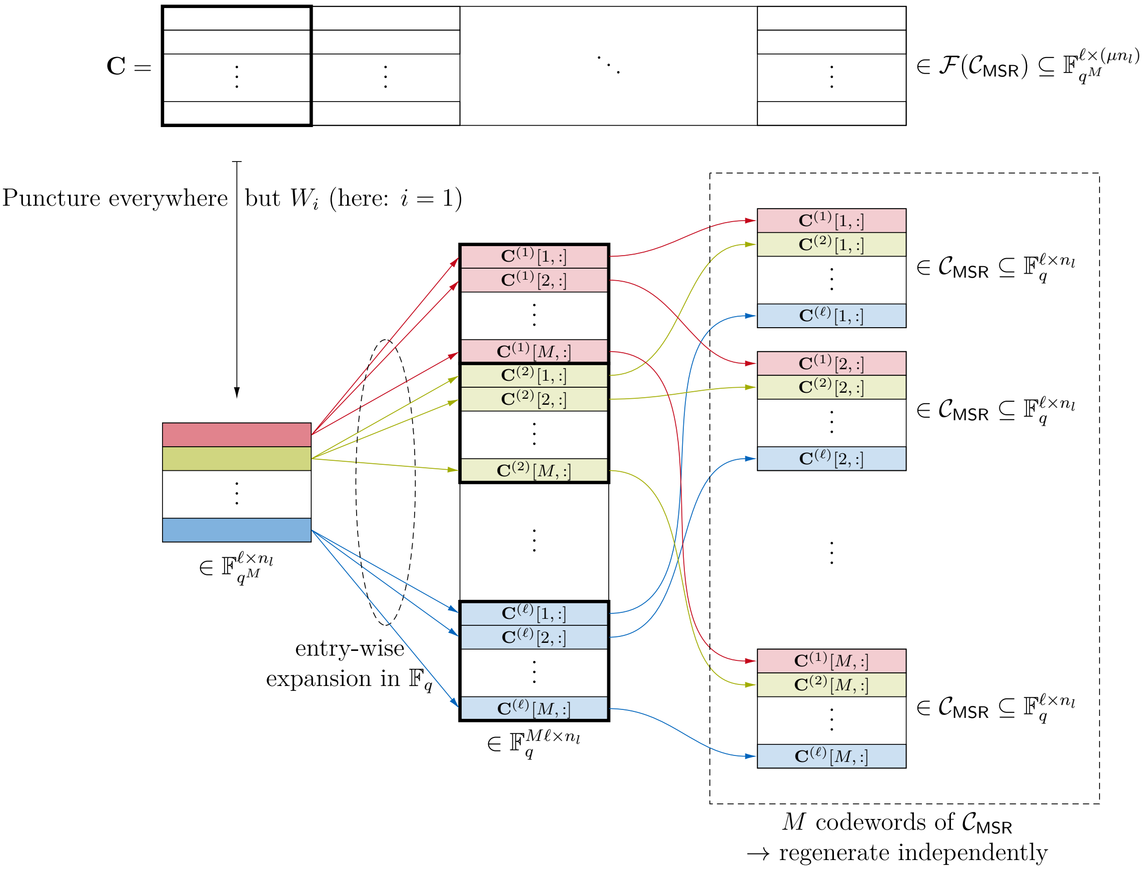
where the last step follows by re-arranging the rows of each codeword (see Fig. 4.3 for an illustration). Hence, each local code is a Cartesian product of M (h, d)-MSR codes and therefore (h, d)-MSR itself. The claim follows by the definition of (h, d)-MSR PMDS codes. □

The remaining difficulty in Construction 4.3 is to find suitable constructions of universal PMDS code families. Some families in the literature already have this property: the Gabidulin-code-based construction of PMDS codes in [RKSV13] and the PMDS code family constructed from linearized RS codes in [MPK19] are both universal. In the following, we show that the construction of [GYBS18] can be turned into a universal PMDS code family. This allows for applying Construction 4.3 to three different classes of universal PMDS codes.

Construction 4.3 using the Gabidulin-Code-Based PMDS Code Family

The code construction in [RKSV13] is based on Gabidulin codes (see Definition 2.6), where the fact that the codes have maximal minimum rank distance is used in [CK16] to show that the constructed codes are indeed PMDS. The construction is a special case of the construction for grid-like topologies given in Section 3.5 (see also Example 3.3), for completeness we briefly recall it here:

- Choose an arbitrary $[n_l, n_l - b, b + 1]_q$ MDS code $\mathcal{C}_{\mathsf{local}}$ and a generator matrix $\mathbf{G}_{\mathsf{local}}$ thereof.
- Choose a Gabidulin code $\mathsf{Gab}(\mu(n_l - b), s)$ (see Definition 2.6) over $\mathbb{F}_{q^M}$. This requires $M \geq \mu(n_l - b)$.
- Encode a message in $\mathbb{F}_{q^M}^{\mu(n_l-b)}$ with this Gabidulin code, which gives a vector $\mathbf{x} \in \mathbb{F}_{q^M}^{\mu(n_l-b)}$.
- Split the vector $\mathbf{x}$ into μ subblocks $\mathbf{x}^{(i)}$ of size $(n_l - b)$, i.e., $\mathbf{x} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(\mu)})$.



$\mathbf{C} =$
$\in \mathcal{F}(\mathcal{C}_{\mathsf{MSR}}) \subseteq \mathbb{F}_{q^M}^{\ell \times (\mu n_l)}$
Puncture everywhere but W_i (here: $i = 1$)
$\in \mathbb{F}_{q^M}^{\ell \times n_l}$
entry-wise expansion in $\mathbb{F}_q$
$\mathbf{C}^{(1)}[1,:]$
$\mathbf{C}^{(1)}[2,:]$
$\mathbf{C}^{(1)}[M,:]$
$\mathbf{C}^{(2)}[1,:]$
$\mathbf{C}^{(2)}[2,:]$
$\mathbf{C}^{(2)}[M,:]$
$\mathbf{C}^{(\ell)}[1,:]$
$\mathbf{C}^{(\ell)}[2,:]$
$\mathbf{C}^{(\ell)}[M,:]$
$\in \mathbb{F}_q^{M\ell \times n_l}$
$\in \mathcal{C}_{\mathsf{MSR}} \subseteq \mathbb{F}_q^{\ell \times n_l}$
M codewords of $\mathcal{C}_{\mathsf{MSR}}$
→ regenerate independently

Figure 4.3: Illustration of the local regeneration procedure implied by the proof of Theorem 4.5.

- Encode each subblock with the generator matrix $\mathbf{G}_{\mathsf{local}}$ to obtain the final codeword $\mathbf{c}$, i.e.,

$$\mathbf{c} = \mathbf{x} \cdot \operatorname{diag}(\mathbf{G}_{\mathsf{local}}, \mathbf{G}_{\mathsf{local}}, \dots, \mathbf{G}_{\mathsf{local}}) = \left(\mathbf{x}^{(1)}\mathbf{G}_{\mathsf{local}}, \dots, \mathbf{x}^{(\mu)}\mathbf{G}_{\mathsf{local}}\right) \in \mathbb{F}_{q^M}^{\mu n_l}.$$

As $\mathcal{C}_{\mathsf{local}}$ is an arbitrary MDS code, we obtain a universal PMDS code family by fixing a Gabidulin code $\mathcal{C}_{\mathcal{G}}$ and varying the local code. For fixed PMDS code parameters, the construction requires only $M \geq \mu(n_l - b)$ (due to the Gabidulin code) and no further restriction on q. Combining this family with the (row-wise MDS) Ye–Barg MSR codes, Theorem 4.5 implies the following statement.

Corollary 4.3. *For all valid PMDS parameters μ, n_l, b, s, integer d with $n_l - b \leq d \leq n_l - 1$, and a partition $\mathbb{W} = \{\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_\mu\}$ of $[\mu n_l]$ with $|\mathcal{W}_i| = n_l \ \forall \ i \in [\mu]$, there exists a d-MSR* $\mathsf{PMDS}(\mu, n_l, b, s, \mathbb{W}; \ell)$ *code over $\mathbb{F}_{q^M}$, if the field size and* *subpacketization satisfy*

$$M \geq \mu(n_l - b), \quad q \geq zn_l, \quad \textit{and} \quad \ell = z^{n_l},$$

where $z = d{+}1{-}n_l{+}b$*. In particular, such a code exists if* $q^M = [(d+1-n_l+b)n]^{\mu(n_l-b)}$.

Construction 4.3 using the Linearized-RS-Codes-Based PMDS Code Family

The PMDS code construction in [MPK19] is similar to that of [RKSV13], however, instead of Gabidulin codes it employs linearized Reed–Solomon codes. These sum-rank-metric codes were first introduced in [MP18] and can be seen as a combination of RS and Gabidulin codes. We do not formally define these codes here, but briefly summarize some of their key properties. Let $\mu < q$, $n_l' \leq M$, and $k' \leq n_l'\mu$. Consider a $[\mu n_l', k']_{q^M}$ linearized RS code. These code are designed for the sum-rank metric w.r.t. the parameter μ, in which codewords are subdivided into μ blocks of size n_l' and the distance of two codewords is the sum of the rank distances of the μ blocks. The distance of linearized RS codes w.r.t. this metric is $\mu n_l' - k' + 1$. Again, this property is essential for the codes of [MPK19] to be PMDS. The construction works as follows:

- Choose an arbitrary MDS code $\mathcal{C}_{\mathsf{local}}[n_l, n_l - b, b + 1]$ over $\mathbb{F}_q$ and a generator matrix $\mathbf{G}_{\mathsf{local}}$ thereof.
- Choose a linearized Reed–Solomon code $\mathcal{C}_{\mathsf{LRS}}$ (cf. [MP18; MPK19]) of parameters $[\mu(n_l - b), \mu(n_l - b) - s]$ over $\mathbb{F}_{q^M}$. This requires $M \geq n_l - b$ and $q > \mu$.
- Encode a message in $\mathbb{F}_{q^M}^{\mu(n_l-b)-s}$ with the linearized Reed–Solomon code $\mathcal{C}_{\mathsf{LRS}}$, which gives a vector $\mathbf{x} \in \mathbb{F}_{q^M}^{\mu(n_l-b)}$.
- Split the vector $\mathbf{x}$ into μ subblocks $\mathbf{x}^{(i)}$ of size $(n_l - b)$, i.e., $\mathbf{x} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(\mu)})$.
- Encode each subblock with the generator matrix $\mathbf{G}_{\mathsf{local}}$ to obtain the final codeword $\mathbf{c}$, i.e.,

$$\mathbf{c} = \mathbf{x} \cdot \operatorname{diag}(\mathbf{G}_{\mathsf{local}}, \mathbf{G}_{\mathsf{local}}, \ldots, \mathbf{G}_{\mathsf{local}}) = \left(\mathbf{x}^{(1)}\mathbf{G}_{\mathsf{local}}, \ldots, \mathbf{x}^{(\mu)}\mathbf{G}_{\mathsf{local}}\right) \in \mathbb{F}_{q^M}^{\mu n_l}.$$

As $\mathcal{C}_{\mathsf{local}}$ is an arbitrary MDS code, we obtain a universal PMDS code family by fixing a linearized Reed–Solomon code $\mathcal{C}_{\mathsf{LRS}}$ and varying the local code. For fixed PMDS code parameters, the construction requires only $M \geq n_l - b$ and $q > \mu$. Note that, compared to the Gabidulin-based PMDS construction above, the restriction on M is much weaker, but we require an additional condition on q. Combining this family with the (row-wise MDS) Ye–Barg MSR codes, Theorem 4.5 implies the following statement.

Corollary 4.4. *For all valid PMDS parameters μ, n_l, b, s, integer d with $n_l - b \leq d \leq n_l - 1$, and a partition $\mathbb{W} = \{\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_\mu\}$ of $[\mu n_l]$ with $|\mathcal{W}_i| = n_l \ \forall \ i \in [\mu]$, there is a d-MSR* $\mathsf{PMDS}(\mu, n_l, b, s, \mathbb{W}; \ell)$ *code over $\mathbb{F}_{q^M}$, if the field size and subpacketization satisfy*

$$M \geq n_l - b, \quad q \geq \max\{z n_l, \mu + 1\}, \quad \textit{and} \quad \ell = z^{n_l},$$

where $z = d + 1 - n_l + b$. In particular, such a code exists for a field of size

$$q^M = \max\{(d+1-n_l+b)n_l, \mu+1\}^{n_l - b} \ .$$

Construction 4.3 using the PMDS Code Family by Gabrys et al.

The local codes of the PMDS code construction in [GYBS18, Section IV.A] are specific RS codes. In order to apply Construction 4.3, we need to show that these PMDS codes are in fact universal. The following theorem presents a slight generalization of the construction and proves its universality. Note that the proof heavily rely on ideas from [GYBS18, Lemma 2], [GYBS18, Corollary 5], and [GYBS18, Lemma 7].

Theorem 4.6 (Generalization of the PMDS Construction in [GYBS18]). *Let n_l, μ, b, s be valid PMDS parameters and $\alpha_{1,1}, \alpha_{1,2}, \dots, \alpha_{\mu,n_l} \in \mathbb{F}_{q^M}$ be distinct field elements such that any subset of $(b+1)s$ elements of the $\alpha_{i,j}$ is linearly independent over $\mathbb{F}_q$. Define*

$$\mathbf{H}^{(j)} = \begin{pmatrix} \alpha_{j,1} & \alpha_{j,2} & \dots & \alpha_{j,n_l} \\ \alpha_{j,1}^q & \alpha_{j,2}^q & \dots & \alpha_{j,n_l}^q \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{j,1}^{q^{s-1}} & \alpha_{j,2}^{q^{s-1}} & \dots & \alpha_{j,n_l}^{q^{s-1}} \end{pmatrix} \quad \forall \, 1 \leq j \leq \mu.$$

Then, the $[\mu n_l, \mu(n_l - b) - s]_{\mathbb{F}_{q^M}}$ code with parity-check matrix

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}^{(0)} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^{(0)} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{H}^{(0)} \\ \mathbf{H}^{(1)} & \mathbf{H}^{(2)} & \dots & \mathbf{H}^{(\mu)} \end{pmatrix} \in \mathbb{F}_{q^M}^{(\mu b + s) \times \mu n_l}$$

is a PMDS code, where $\mathbf{H}^{(0)} \in \mathbb{F}_q^{b \times n_l}$ is a parity-check matrix of an arbitrary $[n_l, n_l - b]_q$ MDS code.

Proof. Let $\mathbf{c} = \left(\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(\mu)}\right)$ be a codeword of the PMDS code, where the μ blocks $\mathbf{c}^{(i)} \in \mathbb{F}_{q^M}^{n_l}$ denote the local codewords. By definition, for all $i = 1, \dots, \mu$, we have

$$\mathbf{H}^{(0)} \mathbf{c}^{(i)\top} = \mathbf{0} \tag{4.8}$$

Furthermore, with $\boldsymbol{\alpha}_i \coloneqq (\alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n_l})$, we have

$$\sum_{i=1}^{\mu} \boldsymbol{\alpha}_i^{q^j} \mathbf{c}^{(i)\top} = 0, \tag{4.9}$$

for all $j = 0, \dots, s-1$. Define the sets $\mathcal{V}_1, \dots, \mathcal{V}_\mu \subset [n_l]$ with $\mathcal{V}_i = \{v_{i,1}, v_{i,2}, \dots, v_{i,b}\}$. Denote by $\bar{\mathcal{V}}_i \coloneqq [n_l] \setminus \mathcal{V}_i$ the respective complement. The positions $\mathcal{V}_1, \dots, \mathcal{V}_\mu$ correspond to the puncturing patterns $\mathcal{E}_i$ in the definition of PMDS codes (see Definition 2.8), i.e., to those subsets of positions of the local codes that, when restricted to, need to result in an $[\mu n_l - \mu b, \mu n_l - \mu b - s]$ MDS code.

By definition, the matrix $\mathbf{H}^{(0)}$ is a parity-check matrix of an $[n_l, n_l - b]_q$ MDS code. Hence, any subset of b columns of $\mathbf{H}^{(0)}$ is invertible. For any $i \in [\mu]$ the vector $\mathbf{c}^{(i)}$ is a codeword of this local code and we have

$$\begin{aligned} \mathbf{0} = \mathbf{H}^{(0)} \cdot (\mathbf{c}^{(i)})^\top &= \mathbf{H}^{(0)}|_{\mathcal{V}_i} \cdot (\mathbf{c}^{(i)}|_{\mathcal{V}_i})^\top + \mathbf{H}^{(0)}|_{\bar{\mathcal{V}}_i} \cdot (\mathbf{c}^{(i)}|_{\bar{\mathcal{V}}_i})^\top \\ \Rightarrow \quad (\mathbf{c}^{(i)}|_{\mathcal{V}_i})^\top &= (\mathbf{H}^{(0)}|_{\mathcal{V}_i})^{-1} \cdot \mathbf{H}^{(0)}|_{\bar{\mathcal{V}}_i} \cdot (\mathbf{c}^{(i)}|_{\bar{\mathcal{V}}_i})^\top \end{aligned}$$

Since all entries of $\mathbf{H}^{(0)}$ are from $\mathbb{F}_q$ by definition we have $\mathbf{H}^{(0)} = (\mathbf{H}^{(0)})^{q^j}$ for the component-wise power. It follows that

$$\begin{aligned} 0 &= \sum_{i=1}^{\mu} \boldsymbol{\alpha}_i^{q^j} (\mathbf{c}^{(i)})^\top \\ &= \sum_{i=1}^{\mu} (\boldsymbol{\alpha}_i|_{\mathcal{V}_i})^{q^j} \left(\mathbf{c}^{(i)}|_{\mathcal{V}_i}\right)^\top + (\boldsymbol{\alpha}_i|_{\bar{\mathcal{V}}_i})^{q^j} \left(\mathbf{c}^{(i)}|_{\bar{\mathcal{V}}_i}\right)^\top \\ &= \sum_{i=1}^{\mu} \Big[\underbrace{(\boldsymbol{\alpha}_i|_{\mathcal{V}_i}) \cdot \left(\mathbf{H}^{(0)}|_{\mathcal{V}_i}\right)^{-1} \cdot \left(\mathbf{H}^{(0)}|_{\bar{\mathcal{V}}_i}\right) + (\boldsymbol{\alpha}_i|_{\bar{\mathcal{V}}_i})}_{=:\, \gamma_{\mathcal{V}_i}} \Big]^{q^j} \left(\mathbf{c}^{(i)}|_{\bar{\mathcal{V}}_i}\right)^\top . \end{aligned}$$

Thus, the vector $\left(\mathbf{c}^{(1)}|_{\bar{\mathcal{V}}_1}, \mathbf{c}^{(2)}|_{\bar{\mathcal{V}}_2}, \dots, \mathbf{c}^{(\mu)}|_{\bar{\mathcal{V}}_\mu}\right)$ is contained in a code with parity-check matrix

$$\mathbf{H}_\gamma \coloneqq \begin{pmatrix} \gamma_{\mathcal{V}}^{q^0} \\ \gamma_{\mathcal{V}}^{q^1} \\ \vdots \\ \gamma_{\mathcal{V}}^{q^{s-1}} \end{pmatrix},$$

where

$$\gamma_{\mathcal{V}} \coloneqq \left(\gamma_{\mathcal{V}_1}, \gamma_{\mathcal{V}_2}, \dots, \gamma_{\mathcal{V}_\mu}\right) \in \mathbb{F}_{q^M}^{\mu(n_l - b)}$$

and $\boldsymbol{\gamma}_{\mathcal{V}}^{q^i}$ denotes the element-wise power of $\boldsymbol{\gamma}_{\mathcal{V}}$. Recall that

$$\boldsymbol{\gamma}_{\mathcal{V}_i} = (\boldsymbol{\alpha}_i|_{\mathcal{V}_i}) \cdot \left(\mathbf{H}^{(0)}|_{\mathcal{V}_i}\right)^{-1} \cdot \left(\mathbf{H}^{(0)}|_{\bar{\mathcal{V}}_i}\right) + (\boldsymbol{\alpha}_i|_{\bar{\mathcal{V}}_i}) \; .$$

Since $\left(\mathbf{H}^{(0)}|_{\mathcal{V}_i}\right)^{-1} \cdot \left(\mathbf{H}^{(0)}|_{\bar{\mathcal{V}}_i}\right)$ is an $b \times (n_l - b)$ matrix, each entry of $\boldsymbol{\gamma}_{\mathcal{V}_i}$, and thus each entry of $\boldsymbol{\gamma}_S$, is a linear combination of at most $b+1$ of the $\alpha_{i,j}$. Furthermore, each such linear combination contains, nontrivially, one element from $\alpha_{i,j}$ (namely the corresponding entry in $\boldsymbol{\alpha}_i|_{\bar{\mathcal{V}}_i}$) that appears only in this linear combination. Hence, any set of s entries from $\boldsymbol{\gamma}_{\mathcal{V}}$ depends on at most $s(b+1)$ of the $\alpha_{i,j}$, which are linearly independent by assumption. It follows that these s entries from $\boldsymbol{\gamma}_S$ are also linearly independent over $\mathbb{F}_q$. Consequently, any s columns of the parity-check matrix $\mathbf{H}_\gamma$ are linearly independent and $\mathbf{H}_\gamma$ is a parity-check matrix of an $[n_l\mu - b\mu, n_l\mu - b\mu - s]_{q^M}$ MDS code.

It remains to show that the local codes equal the $\mathbb{F}_{q^M}$-span of an $[n_l, n_l - b]_q$ MDS code. By construction it is obvious that each local code is a subcode of such an MDS code, as it fulfills the parity-checks given by $\mathbf{H}^{(0)}$. The fact that each local code is equal to this MDS code, follows directly from the dimension of the code (see also Proposition 2.2). Hence, the overall code is a PMDS code. □

As the MDS code over $\mathbb{F}_q$ can be chosen arbitrarily for fixed $\alpha_{1,1}, \alpha_{1,2}, \dots, \alpha_{\mu,n_l} \in \mathbb{F}_{q^M}$, Theorem 4.6 immediately implies a universal PMDS code family as in Definition 4.9. By Theorem 4.5, we get the following result.

Corollary 4.5. *For all valid PMDS parameters μ, n_l, b, s, integer d with $n_l - b \leq d \leq n_l - 1$, and $\mathbb{W} = \{\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_\mu\}$ a partition of $[\mu n_l]$ with $|\mathcal{W}_i| = n_l \; \forall \; i \in [\mu]$, there is a d-MSR PMDS array code as in Construction 4.3 of field size*

$$n_l\Big[d+1-(n_l-b)\Big](n_l\mu)^{s(b+1)-1} \leq q^M \leq 2n_l\Big[d+1-(n_l-b)\Big](2n_l\mu)^{s(b+1)-1}$$

and subpacketization

$$\ell = \Big[d+1-(n_l-b)\Big]^{n_l}.$$

Proof. We combine the universal PMDS code family in Theorem 4.6 with Ye–Barg codes (cf. Definition 4.5) using Construction 4.3. We choose q and M large enough such that we can ensure that suitable field elements $\alpha_{i,j}$ (of the PMDS code family) and $\beta_{i,j}$ (of the Ye–Barg codes) exist. A sufficient condition for the existence of the $\beta_{i,j}$ is $q \geq n_l(d+1-(n_l-b))$. Thus, we can choose q to be the smallest prime power greater or equal to $n_l(d+1-(n_l-b))$, which is at most $q \leq 2n_l(d+1-(n_l-b))$.

For the $\alpha_{i,j}$, it is a bit more involved. By Theorem 4.6, we need to find $n_l\mu$ elements of $\mathbb{F}_{q^M}$ such that any subset of $s(b+1)$ elements is linearly independent. We use the same idea as in [GYBS18, Lemma 7]. Take the columns of a parity-check matrix of an $[n_l\mu, n_l\mu - M, s(b+1)+1]_q$ code and interpret each column in $\mathbb{F}_q^M$ as an element of $\mathbb{F}_{q^M}$. It is well-known that any subset of $d_{\min} - 1$ columns of the parity-check matrix of a code of distance $d_{\min}$ are linearly independent and it follows that these elements fulfill the required condition.

It remains to determine the extension degree M and base field size q such that a code with these parameters exists. We adapt the result in [Rot06, Problem 8.9] and present it in terms of our notation. For any $n' = q^w - 1$, there exists a code with parameters $[n', n' - M, s(b+1)+1]_q$ for

$$M \leq 1 + \Big(s(b+1) - 1\Big)w.$$

Choose w to be the smallest integer with $n' = q^w - 1 \geq n_l\mu$. Note that there is such an w with $q^w - 1 \leq 2n_l\mu - 1$, i.e., $\log_q(n_l\mu) \leq w \leq \log_q(2n_l\mu)$. Hence, there is an $[n', n' - M, s(b+1)+1]_q$ code with $M \leq 1 + \Big(s(b+1) - 1\Big)\log_q(2n_l\mu)$. Shortening the code in (arbitrary) $n' - n_l\mu$ positions gives an $[n_l\mu, n_l\mu - M, s(b+1)+1]_q$ code with $M \leq 1 + \Big(s(b+1) - 1\Big)\log_q(2n_l\mu)$. □

4.5 Discussion and Comparison of PMDS Code Constructions with Local Regeneration

In the previous sections, we presented multiple constructions of locally MSR PMDS codes, each based on a different PMDS code construction. In this section we compare the parameters of these new constructions among each other and to the only existing construction of locally d-MSR PMDS codes, which was presented in [RKSV13]. Table 4.2 summarizes their respective field sizes and, for easier reference, labels the five constructions by the letters A–E.

The known Construction E (see [RKSV13, Construction 1, case "$(b+\delta-1) \mid n_l$"]) first encodes an information word from $\mathbb{F}_{q^M}^{\ell \times (\mu(n_l-b)-s)}$ with an $[\ell\mu(n_l-b), \ell(\mu(n_l-b)-s)]_{q^M}$ Gabidulin code. The resulting codeword is then subdivided into μ groups, each of length $\ell(n_l - b)$. These subblocks are then independently encoded using a generator matrix of an $[n_l, n_l - b; \ell]_q$ d-MSR code. This gives a d-MSR PMDS array code with subpacketization ℓ and field size q^M, where the only requirements on ℓ and q are the constraints of the MSR code and $M \geq \ell\mu(n_l - b)$ in order for the Gabidulin code to exist. An advantage of this construction over the constructions of this work is that it does not require the MSR code to be row-wise MDS. However, the field size is exponential in the subpacketization, i.e., doubly exponential in n_l for Ye–Barg codes.

In the following theorem we collect some relations between the obtained field sizes. Informally, the observations can be summarized as:

- Construction C always has smaller field size than Constructions B and E.
- For two global parities, Construction A has the smallest field size among all constructions (unless b or μ are very large).

Table 4.2: Comparison of field sizes of locally d-MSR PMDS array code constructions (parameters: d, n_l, μ, b, s such that $b \leq n_l$, $s \leq (n_l - b)\mu$, and $n_l - b \leq d \leq n_l - 1$).

Label	Construction	Restr.	Smallest field size $Q_\star = q^M$
A	Corollary 4.2	$s = 2$	$\mu b(bn_l - b + n_l - 2) + 1 \leq Q_{\mathsf{A}}$ $\leq 2\mu b(bn_l - b + n_l - 2)$
B	Corollary 4.3	–	$Q_{\mathsf{B}} = [(d+1-n_l+b)n_l]^{\mu(n_l-b)}$
C	Corollary 4.4	–	$Q_{\mathsf{C}} = \max\Big\{(d+1-n_l+b)n_l, \mu+1\Big\}^{n_l-b}$
D	Corollary 4.5	–	$n_l\Big[d+1-n_l+b\Big](n_l\mu)^{s(b+1)-1} \leq Q_{\mathsf{D}}$ $\leq 2n_l\Big[d+1-n_l+b\Big](2n_l\mu)^{s(b+1)-1}$
E	[RKSV13] + Ye–Barg	–	$Q_{\mathsf{E}} = [(d+1-n_l+b)n_l]^{(d+1-n_l+b)^{n_l}\mu(n_l-b)}$

- For a large number of global or local parities (and $s > 2$), Construction C has the smallest field size among all constructions.
- For a small number of global (but $s > 2$) and local parities, Construction D has the smallest field size among all constructions.

Theorem 4.7. *For all valid PMDS parameters μ, n_l, b, s and integers d with $n_l - b < d \leq n_l - 1$, denote by $Q_{\mathsf{A}}, Q_{\mathsf{B}}, Q_{\mathsf{C}}, Q_{\mathsf{D}}, Q_{\mathsf{E}}$ the smallest field sizes obtained from the constructions in Table 4.2.*

(i) For all parameters, we have $Q_{\mathsf{C}} < Q_{\mathsf{B}} < Q_{\mathsf{E}}$.

(ii) For $s = 2$, we have $Q_{\mathsf{A}} < Q_{\mathsf{D}}$. If in addition, $b < n_l - 3$, and $\mu \leq n_l^{n_l-b-3}$, then $Q_{\mathsf{A}} < Q_{\mathsf{C}}$.

(iii) For $s(b+1) + 2b - 1 \geq 2n_l$, we have $Q_{\mathsf{C}} < Q_{\mathsf{D}}$.

(iv) For $2s(b+1) + b \leq n_l$, we have $Q_{\mathsf{D}} < Q_{\mathsf{C}}$.

Proof. We use the two properties (a) $a \leq a + 1 < 3^a$ and (b) $a^b \geq ab$ for integers $a, b \geq 1$, which can both be proven easily by induction.

Ad (i): As $d > n_l - b$, we have $d + 1 - n_l - b > 1$, so obviously $Q_{\mathsf{E}} > Q_{\mathsf{B}}$. If $(d + 1 - n_l - b)n_l \geq \mu + 1$, it is clear that $Q_{\mathsf{B}} > Q_{\mathsf{C}}$ (here we use $\mu \geq 2$). In the case $(d + 1 - n_l - b)n_l \geq \mu + 1$, we have

$$Q_{\mathsf{C}} = (\mu+1)^{n_l-b} \overset{(\mathsf{a})}{<} 3^{\mu(n_l-b)} \leq [(d+1-n_l+b)n_l]^{\mu(n_l-b)} \ ,$$

where $(d+1-n_l+b)n_l \geq 3$ holds by assumption.

Ad (ii): We have

$$\begin{aligned} Q_{\mathsf{A}} &\leq 2\mu b(bn_l - b + n_l - 2) \\ &< 2\mu n_l^2[2(b+1)-1] \\ &\overset{\text{(b)}}{\leq} 2n_l(n_l\mu)^{2(b+1)-1} \\ &\leq n_l\Big[d+1-n_l+b\Big](n_l\mu)^{s(b+1)-1} \leq Q_{\mathsf{D}} \ , \end{aligned}$$

where we use $d+1-n_l+b \geq 2$. Furthermore, if also $b < n_l - 3$ and $\mu \leq n_l^{n_l-b-3}$, we have

$$Q_{\mathsf{A}} \leq 2\mu b(bn_l - b + n_l - 2) < 2\mu n_l^3 \leq 2n_l^{n_l-b} \leq [(d+1-n_l+b)n_l]^{n_l-b} \leq Q_{\mathsf{C}} \ .$$

Ad (iii): Denote $z := d+1-n_l+b$ and recall that $2 \leq z \leq b < n_l$. We must show $Q_{\mathsf{D}} > (zn_l)^{n_l-b}$ and $Q_{\mathsf{D}} > (\mu+1)^{n_l-b}$. We start with the first inequality:

$$Q_{\mathsf{D}} \geq n_l z(n_l\mu)^{s(b+1)-1} \geq n_l z n_l^{s(b+1)-1} \geq n_l z\Big(\underbrace{n_l^2}_{>n_l z}\Big)^{\frac{s(b+1)-1}{2}} > \Big(n_l z\Big)^{\frac{s(b+1)+1}{2}} \geq (n_l z)^{n_l-b}.$$

The second inequality holds since

$$Q_{\mathsf{D}} \geq n_l b(n_l\mu)^{s(b+1)-1} > (\mu+1)^{s(b+1)-1} \geq (\mu+1)^{n_l-b} \ .$$

Ad (iv): Define $\xi := \max\{n_l b, \mu\}$ (we use $z := d+1-n_l+b$ with $2 \leq z \leq b < n_l$ as above). It suffices to show $Q_{\mathsf{D}} < \xi^{n_l-b}$ under the given conditions. We have

$$Q_{\mathsf{D}} \leq 2n_l z(\underbrace{2n_l}_{\leq n_l z \leq \xi}\ \mu)^{s(b+1)-1} < \xi^{2s(b+1)} \leq \xi^{n_l-b} \leq Q_{\mathsf{C}}.$$

This concludes the proof. □

Figs. 4.4 to 4.6 plot the field size bounds of Table 4.2 over the number of local parities b for different sets of PMDS code parameters and $d = n_l - 1$. The field size of Construction E (known construction) far exceeds the plot range and is therefore not included in the figures.

4.6 Summary and Open Problems

This chapter considered PMDS codes with global and local regeneration properties. For both settings, we have presented constructions of PMDS array codes with global/local MSR codes.

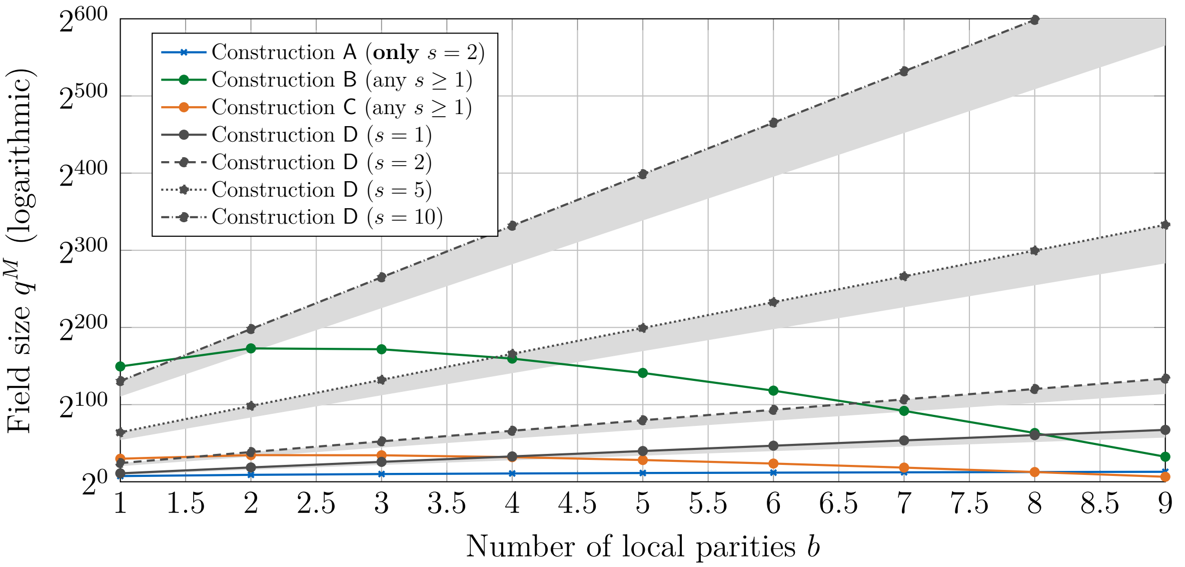


Figure 4.4: Comparison of the field sizes of Constructions A–D for $n_l = 10$, $\mu = 5$, and $d = 9$. Construction E is not shown as it is out of plot range. Lines are upper bounds, shadows indicate lower bounds. The field sizes of Constructions B and C are independent of s.

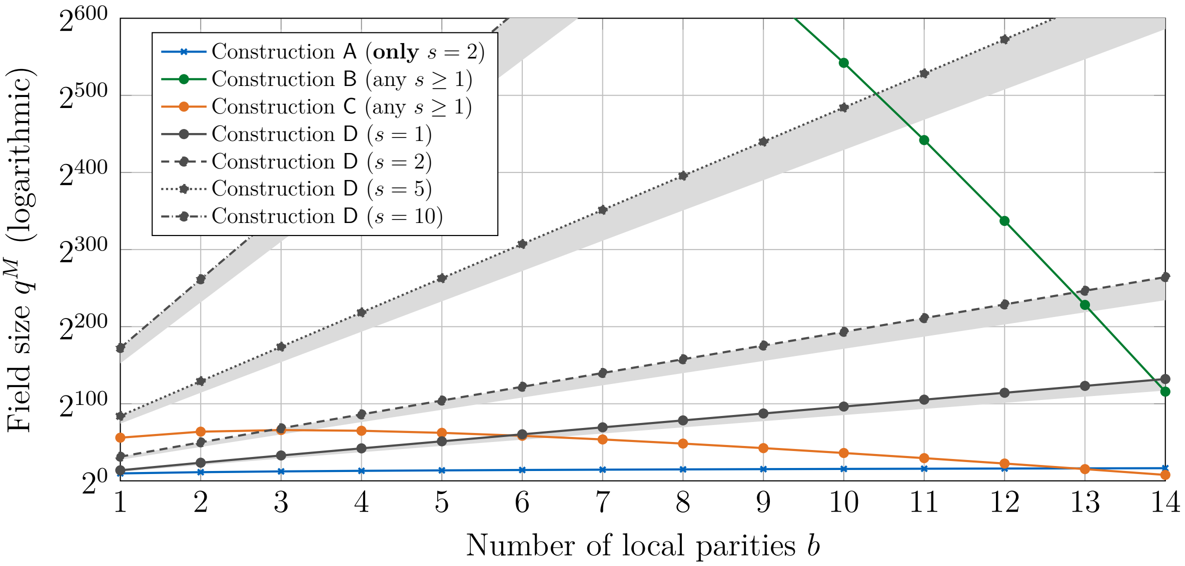


Figure 4.5: Comparison of the field sizes of Constructions A–D for $n_l = 15$, $\mu = 15$, and $d = 14$. Construction E is not shown as it is out of plot range. Lines are upper bounds, shadows indicate lower bounds. The field sizes of Constructions B and C are independent of s.

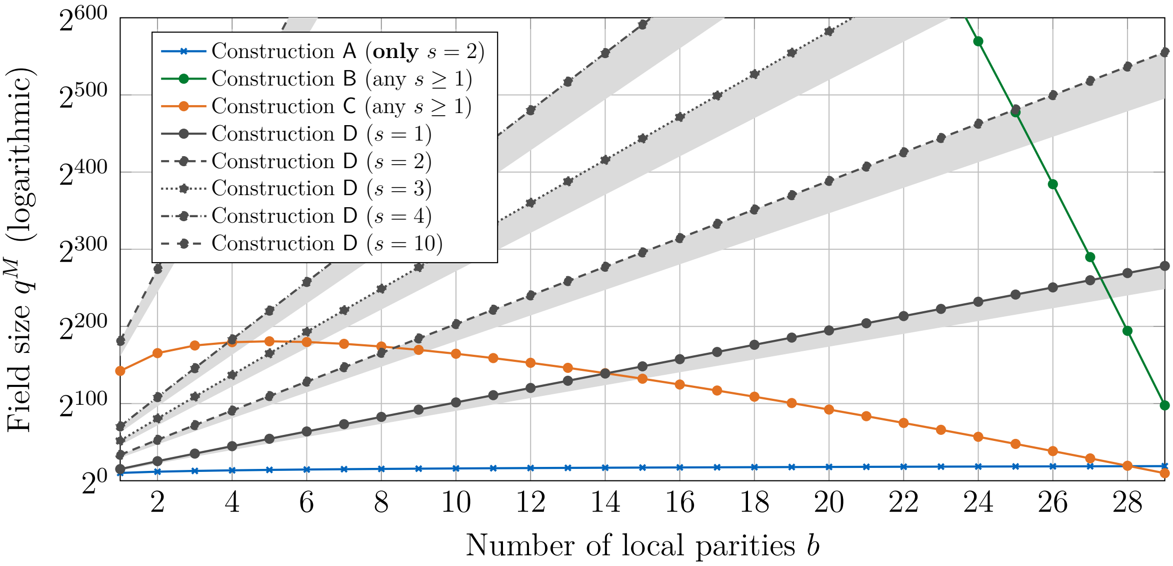


Figure 4.6: Comparison of the field sizes of Constructions A–D for $n_l = 30$, $\mu = 10$, and $d = 29$. Construction E is not shown as it is out of plot range. Lines are upper bounds, shadows indicate lower bounds. The field sizes of Constructions B and C are independent of s.

Specifically, we have presented a construction of a globally MSR PMDS code by introducing a new MSR code, which can be seen as the skew-analog of Ye–Barg codes (similar to the analogy between Gabidulin and Reed–Solomon codes), and combining it with the Gabidulin-code-based PMDS construction of [RKSV13]. This is the first known construction of globally MSR PMDS codes. Compared to the underlying PMDS code, the required field size is increased by a factor in the exponent. The required subpacketization, compared to a Ye–Barg MSR code without the PMDS property, is also increased by a factor in the exponent.

In the second part of this chapter, we have presented a construction for PMDS codes with two global parities based on [BPSY16] where each local code is an MSR code and whose field size is polynomial for a fixed number of local parities. Furthermore, we have introduced a general construction that combines an arbitrary family of universal PMDS codes with a row-wise MDS MSR code. After proving the universality property for the PMDS construction of [GYBS18], we have explicitly stated the resulting field size and subpacketization for three families of locally MSR PMDS codes based on the combination of the universal PMDS code constructions of [RKSV13; MPK19; GYBS18] with Ye–Barg MSR codes [YB17a]. Finally, we have compared the obtained field sizes of the presented constructions. All constructions have a significantly smaller field size than the only existing construction of PMDS codes with local regeneration given in [RKSV13].

Several open problems related to the presented results offer interesting opportunities for further research. Applying the ideas of the presented constructions to the recently proposed PMDS code constructions of [CMST20; GG20; MP20] could reduce the required field size. Both, the constructions of locally and globally MSR PMDS codes require large levels of subpacketization. The former rely on Ye-Barg regenerating codes, which are known to be suboptimal in terms of subpacketization. However, aside from being optimal in terms of repair bandwidth, they are also row-wise MDS, a property that is essential to the presented constructions. A construction that can afford to relax this requirement could improve the required subpacketization by employing different classes of MSR codes as the local MDS codes. Additionally, the construction of globally MSR PMDS codes is based on Gabidulin codes and thereby inherently suffers from a large required field size. This field size could be lowered by instead employing linearized RS codes to achieve similar gains as shown for locally MSR PMDS codes in Section 4.5. Aside from the improvements of the constructions, lower bounds on the required subpacketization and field size would help evaluate the performance of the presented constructions. Finally, for the globally MSR PMDS codes, it remains an open problem to utilize surviving local redundancy nodes, in particular in the extreme case where $r + 1$ nodes in a single local repair set fail while all other nodes survive.

5

Decoding of Lifted Affine-Invariant Codes

Abstract

Lifting of affine invariant codes is a powerful method for constructing codes with strong locality and availability properties, which recently lead to significant advances in the construction of, e.g., batch and PIR codes. This chapter introduces a simple bounded distance decoder for lifted affine-invariant codes that is guaranteed to decode up to half of an asymptotically tight bound on their minimum distance. Further, long q-ary lifted affine-invariant codes are shown to correct almost all error patterns of relative weight ${}^{q-1}\!/_{q} - \epsilon$ for $\epsilon > 0$.

This chapter is based on the work [HP20] published in the proceedings of the 2020 IEEE Information Theory Workshop (ITW)*.*

5.1 Introduction

The previous chapters considered codes that provide one (Chapter 4) or two (Chapter 3) disjoint repair sets for each position, with the primary purpose of correcting erasures. While this notion of locality is well-suited for use in, e.g., distributed storage, locality also has implications in other areas of research and, specifically, in error correction. The class of lifted affine-invariant codes studied in this chapter is naturally suited to provide a large number of repair sets (*availability*) and therefore of interest for applications such as locally decodable and testable codes [GKS13], batch codes [HPPV20], low-degree testing [AS03], and list decoding [GRS00; STV01].

The essential property of a lifted affine-invariant code (see Section 2.3.3) is that the restriction of a codeword to any subspace of a given dimension is a codeword of the code being lifted. A popular example of codes in this class are lifted RS codes [GKS13], which, on a high level, are generalizations of q-ary Reed–Muller codes [DGMW70; KLP68; MCJ73] that provide the same locality and availability properties at a higher rate. These properties naturally lead to local decoding algorithms, i.e., randomized approaches to correctly recover a single symbol with high probability. However, they can also be exploited to design algorithms for the recovery of the entire codeword symbol-by-symbol through aggregation of the local decoding results.

The presented results can be applied to any lifted affine-invariant code, such as lifted RS codes. As this class contains q-ary RM codes as subcodes, the decoder also applies to this well-known class of codes. Due to this close relation and the fact that lifted RS codes are one of the most prominent examples of lifted affine-invariant codes, we shortly recall some of the more recent decoding algorithms for q-ary RM codes here. A q-ary RM code $\mathcal{RM}_q(u, m)$ consists of the evaluation of all m-variate polynomials of degree at most u with coefficients in $\mathbb{F}_q$. These codes have been shown to be subfield subcodes of RS codes over $\mathbb{F}_{q^m}$ [KLP68]. Thus, any decoding algorithm for RS codes can be used to decode RM codes. Randomized list-decoding algorithms for RM codes were proposed in [GRS00; STV01; AS03; GKZ08] and three deterministic list-decoders for Reed–Muller codes running in polynomial time were introduced in [PW04]. Two of the latter view RM codes as subfield subcodes of RS codes and can decode beyond half the minimum distance requiring a polynomial number of field operations in the large field $\mathbb{F}_{q^m}$. An approach for a global decoding algorithm of RM codes based on local decoding has been discussed in [KK16a].

For $u < q$ the code $\mathcal{RM}_q(u, m)$ is a subcode of the corresponding lifted RS code, as introduced in [GKS13]. It is known [GKS13; HPPV20] (see also Section 6.2) that for fixed m and large q, the rate of lifted RS codes approaches one, whereas the rate of non-binary RM codes does not exceed $1/m!$. Surprisingly, similar to RM codes, they can also be seen as subfield subcodes of (low-degree) RS codes [GK16] and thereby (list-)decoded by RS (list-)decoders over $\mathbb{F}_{q^m}$.

5.1.1 Contributions and Outline

Section 5.2 introduces a new deterministic bounded distance (BD) decoding algorithm for lifted affine-invariant codes. As long as the base code admits an efficient unique decoding algorithm, this decoder runs in polynomial time and is guaranteed to decode up to half of the asymptotically tight bound on the minimum distance given in [GKS13, Lemma 5.7] (see Lemma 2.3). Applying the decoder to lifted RS codes requires $n^2 \operatorname{poly}(\log q)$ operations in $\mathbb{F}_q$ (see Theorem 5.1), given a BMD decoder for the (q-ary) RS base code running in $q \operatorname{poly}(\log q)$ (see, e.g., [Gao03]).

Then, in Section 5.3, we analyse a fast randomized decoder for long q-ary codes constructed by lifting a fixed affine-invariant code. A random pattern of errors with relative weight less than $q-1/q - \epsilon$ is shown to be correctable with probability at least $1 - \delta$, where δ can be exponentially small in length, in time $\log \delta^{-1} \operatorname{poly}(\epsilon^{-1})$. This resembles the behaviour of randomized decoders for low-rate binary RM codes shown in [Dum04; Kri70].

5.2 Bounded Distance Decoding

In this section, we introduce a simple bounded distance decoder for lifted affine-invariant codes, solely based on the definition of lifting. Specifically, the decoder is based on principle that the restriction of a lifted affine-invariant code $\mathcal{L}(\mathcal{F})$ to any affine subspace belongs to the code $\mathcal{F}$. Assuming a fixed value at one position, we derive the minimal number of positions in which the evaluations of two functions would have to disagree for the decoding of these restrictions to give the respective result. We then show that the value that results in the lowest number must be the correct value of the function at this position, as long as the number of positions in which the evaluations of the functions disagree (number of errors) is within the decoding radius. The resulting decoder is therefore a BD decoder, i.e., given a number of errors within the defined range, it succeeds with probability 1.

A (partial) partition of $\mathbb{F}_Q^m$, treated as an m-dimensional vector space, is a partition of the vector space into trivially intersecting z-dimensional subspaces $\mathcal{V}_1, \ldots, \mathcal{V}_s \subset \mathbb{F}_Q^m$, i.e., $\mathcal{V}_i \cap \mathcal{V}_j = \{\mathbf{0}\}$ for $i \neq j$. Such a (partial) partition has been shown to exist [Bu80] for any $z \leq {}^{m}/_{2}$ and

$$s = \begin{cases} \frac{Q^m-1}{Q^z-1}, & \text{if } z \mid m, \\ Q^{m-z}, & \text{if } z \nmid m. \end{cases} \tag{5.1}$$

Before introducing the decoding algorithm, we require some preliminary definitions related to the to the distance properties of lifted affine-invariant codes.

Definition 5.1. *For $1 \leq z \leq \frac{m}{2}$, let $\mathrm{ev}_{\mathbb{F}_Q^z}(\mathcal{F})$ with $\mathcal{F} \subseteq \{\mathbb{F}_Q^z \to \mathbb{F}_q\}$ be an affine-invariant code of distance $d_{\mathcal{F}}$. Define $t_{\mathcal{F}} \coloneqq \lfloor \frac{d_{\mathcal{F}}-1}{2} \rfloor$. Consider a function $g \in \{\mathbb{F}_Q^m \to \mathbb{F}_q\}$ and let $g_{\mathbf{a}}^{(\mathcal{V}_i)}$ be as in Eq. (2.9). For any point $\mathbf{a} \in \mathbb{F}_Q^m$, field element $\alpha \in \mathbb{F}_q$, and integer $j \in [0, t_{\mathcal{F}}]$, define $M_{\mathbf{a}}(\alpha, j)$ to be the number of affine subspaces of the form $\mathbf{a} + \mathcal{V}_i$ such that there exists[1] a $\hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)} \in \mathcal{F}$ with $\hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{0}) = \alpha$ and*

$$\begin{aligned} &\bullet\ \mathsf{d_H}\left(\mathrm{ev}_{\mathbb{F}_Q^z \setminus \mathbf{0}}(g_{\mathbf{a}}^{(\mathcal{V}_i)}), \mathrm{ev}_{\mathbb{F}_Q^z \setminus \mathbf{0}}(\hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)})\right) = j, && \text{if } d_{\mathcal{F}} \text{ is even,} \\ &\bullet\ \mathsf{d_H}\left(\mathrm{ev}_{\mathbb{F}_Q^z}(g_{\mathbf{a}}^{(\mathcal{V}_i)}), \mathrm{ev}_{\mathbb{F}_Q^z}(\hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)})\right) = j, && \text{if } d_{\mathcal{F}} \text{ is odd.} \end{aligned} \tag{5.2}$$

Further, define $M_{\mathbf{a}}(\star)$ to be the number of affine subspaces $\mathbf{a}+\mathcal{V}_i$ for which no $\hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)} \in \mathcal{F}$ that satisfies Eq. (5.2) exists for any $\alpha \in \mathbb{F}_q$ and $j \in [0, t_{\mathcal{F}}]$. For $\alpha \in \mathbb{F}_q$, denote

$$\delta_{\mathbf{a}}(\alpha) \coloneqq \begin{cases} \mathbb{1}\{g(\mathbf{a}) \neq \alpha\}, & \text{if } d_{\mathcal{F}} \text{ is odd,} \\ 0, & \text{if } d_{\mathcal{F}} \text{ is even.} \end{cases}$$

[1] Note that by definition of $t_{\mathcal{F}}$ there exists at most one such function in $\mathcal{F}$.

For $\mathbf{a} \in \mathbb{F}_Q^m$ *and* $\alpha \in \mathbb{F}_q$*, define*

$$N_{\mathbf{a}}(\alpha) \coloneqq \mathbb{1}\{g(\mathbf{a}) \neq \alpha\} + \sum_{j=0}^{t_{\mathcal{F}}} (j - \delta_{\mathbf{a}}(\alpha)) M_{\mathbf{a}}(\alpha, j) \\ + \sum_{\beta \neq \alpha} \sum_{j=0}^{t_{\mathcal{F}}} (d_{\mathcal{F}} - 1 - j + \delta_{\mathbf{a}}(\beta)) M_{\mathbf{a}}(\beta, j) + (t_{\mathcal{F}} + 1 - \delta_{\mathbf{a}}(\alpha)) M_{\mathbf{a}}(\star).$$

Despite the complicated formal notation, the underlying idea of this definition is rather simple. In a lifted affine-invariant code, the function $g_{\mathbf{a}}^{(\mathcal{V}_i)}$ corresponds to the word of the affine-invariant code obtained by restricting the function g to the affine subspace $\mathcal{V}_i + \mathbf{a}$ (with the point $\mathbf{a}$ shifted to the origin). When applying a decoder for the affine-invariant code in all affine subspaces containing the point $\mathbf{a}$, each decoding outcome can be viewed as a "vote" for the value of the correct word in this point. Intuitively, a vote is "more reliable" when fewer errors were corrected. The variable $M_{\mathbf{a}}(\alpha, j)$ counts the number of these votes for the value α at position $\mathbf{a}$ from decoders that corrected exactly j errors in the respective restriction to the affine subspace. The number of decoders that failed to return a valid codeword is given by $M_{\mathbf{a}}(\star)$. The different events contributing to $M_{\mathbf{a}}(\beta, j)$ and $M_{\mathbf{a}}(\star)$ are illustrated in Fig. 5.1.

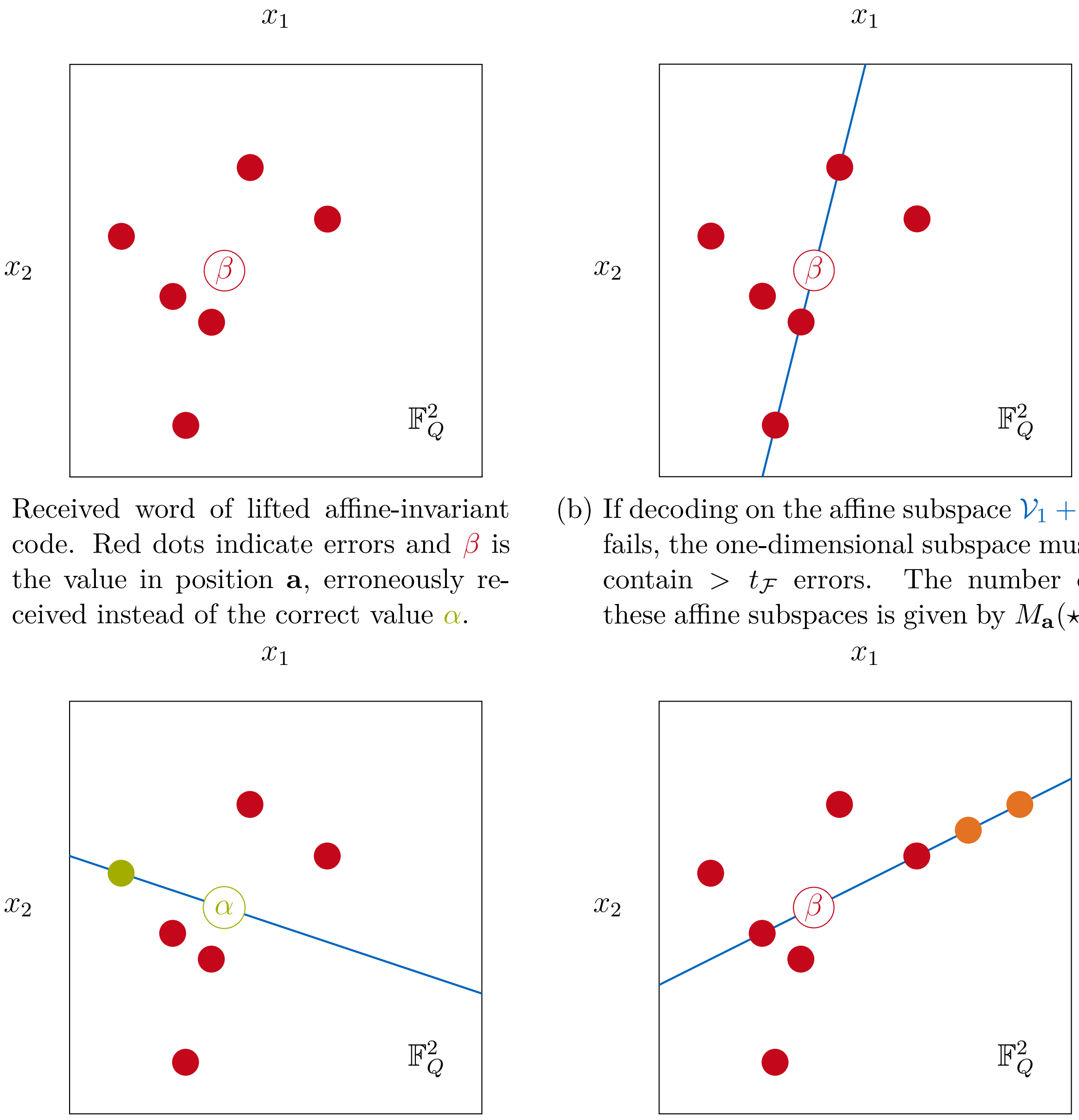
On a high level, the value $N_{\mathbf{a}}(\alpha)$ then aggregates these values to reflect the minimal number of errors that must have occurred for these votes to be possible, when assuming α to be the correct value in position $\mathbf{a}$. The first sum accounts for the subspaces that "voted" for the value α at point $\mathbf{a}$. Given the assumption that α is the correct value in this position, all decoders that "voted" for a different value must have returned an incorrect word, which is accounted for in the second double sum. The last term accounts for the errors required for a local decoder to fail. Finally, the indicator function and the various δ account for the possibility of an error at position $\mathbf{a}$.

Formally, we can apply Definition 5.1 to express the distance between the evaluations of two functions in terms of the distances between the evaluations of their respective restrictions to affine subspaces.

Lemma 5.1. *Let* $\mathcal{F}$*,* g*, and* $N_{\mathbf{a}}(\alpha)$ *be as in Definition 5.1 and* $\mathcal{L}(\mathcal{F}) \subseteq \{\mathbb{F}_Q^m \to \mathbb{F}_q\}$ *be the set of functions of a lifted code as in Definition 2.13. Then for any* $\mathbf{a} \in \mathbb{F}_Q^m$*,* $\alpha \in \mathbb{F}_q$*, and* $f \in \mathcal{L}(\mathcal{F})$ *with* $f(\mathbf{a}) = \alpha$ *it holds that*

$$\mathrm{d}_{\mathsf{H}}\left(\mathrm{ev}_{\mathbb{F}_Q^m}(f), \mathrm{ev}_{\mathbb{F}_Q^m}(g)\right) \geq N_{\mathbf{a}}(\alpha) \ .$$

Proof. Given the point $\mathbf{a} \in \mathbb{F}_Q^m$, we count the number of evaluations that must differ between f and g given the values of $M_{\mathbf{a}}(\beta, j)$, $\forall \beta \in \mathbb{F}_q$, $\forall j \in \{0, 1 \ldots, t_{\mathcal{F}}\}$, and $M_{\mathbf{a}}(\star)$. Recall that $M_{\mathbf{a}}(\beta, j)$ is the number of affine subspaces $\mathbf{a} + \mathcal{V}_i$ for which $\hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{0}) = \beta$ and Eq. (5.2) holds. By definition of a (partial) partition, these affine subspaces intersect *only* in $\mathbf{a}$ and the sum over all $M_{\mathbf{a}}(\beta, j)$ is the number of affine subspaces of the form



(a) Received word of lifted affine-invariant code. Red dots indicate errors and β is the value in position $\mathbf{a}$, erroneously received instead of the correct value α.

(b) If decoding on the affine subspace $\mathcal{V}_1 + \mathbf{a}$ fails, the one-dimensional subspace must contain $> t_{\mathcal{F}}$ errors. The number of these affine subspaces is given by $M_{\mathbf{a}}(\star)$.

(c) The affine subspace $\mathcal{V}_2 + \mathbf{a}$ is decoded correctly, correcting the $j = 2$ errors indicated in green. The number of these affine subspaces is given by $M_{\mathbf{a}}(\alpha, 2)$.

(d) The affine subspace $\mathcal{V}_3 + \mathbf{a}$ is decoded incorrectly to a word at distance $\geq d_{\mathcal{F}}$ of the correct word. The number of these affine subspaces is given by $M_{\mathbf{a}}(\beta, 2)$.

Figure 5.1: Illustration of the different decoding events on affine subspaces of dimension $z = 1$ in a lifted affine-invariant code over $\mathbb{F}_Q^2$ (see also Fig. 2.1 on Page 26) contributing to $M_{\mathbf{a}}(\beta, j)$ and $M_{\mathbf{a}}(\star)$. The distance of the affine-invariant code is $d_{\mathcal{F}} = 5$. Note that each decoding event is independent, e.g., correcting the error in position $\mathbf{a}$ in Fig. 5.1c does not affect the other decoding results.

$\mathbf{a}+\mathcal{V}_i$, i.e.,

$$\sum_{\beta,j} M_{\mathbf{a}}(\beta,j)+M_{\mathbf{a}}(\star)=s, \tag{5.3}$$

for s as in Eq. (5.1). Hence, the distance between f and g is lower bounded by

$$\mathrm{d_H}\left(\mathrm{ev}_{\mathbb{F}_Q^m}(f),\mathrm{ev}_{\mathbb{F}_Q^m}(g)\right) \geq \mathbb{1}\{f(\mathbf{a})\neq g(\mathbf{a})\}+\sum_{i=1}^{s}\mathrm{d_H}\left(\mathrm{ev}_{\mathbb{F}_Q^z\setminus\mathbf{0}}(f_{\mathbf{a}}^{(\mathcal{V}_i)}),\mathrm{ev}_{\mathbb{F}_Q^z\setminus\mathbf{0}}(g_{\mathbf{a}}^{(\mathcal{V}_i)})\right).$$

As $f(\mathbf{a})=\alpha$ by assumption, we have $\mathbb{1}\{f(\mathbf{a})\neq g(\mathbf{a})\}=\mathbb{1}\{g(\mathbf{a})\neq\alpha\}$. For the remaining positions, first consider the case of odd $d_{\mathcal{F}}$. For all affine subspaces contributing to $M(\alpha,j)$ (see Fig. 5.1c), where $j\leq t_{\mathcal{F}}$, we have

$$\mathrm{d_H}\left(\mathrm{ev}_{\mathbb{F}_Q^z}(f_{\mathbf{a}}^{(\mathcal{V}_i)}),\mathrm{ev}_{\mathbb{F}_Q^z}(g_{\mathbf{a}}^{(\mathcal{V}_i)})\right) \geq \begin{cases} j, & \text{if } \hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)}=f_{\mathbf{a}}^{(\mathcal{V}_i)},\\ d_{\mathcal{F}}-j\geq j, & \text{else.}\end{cases} \tag{5.4}$$

Excluding point $\mathbf{a}$ in f and g, i.e., the origin in the restrictions to $\mathbf{a}+\mathcal{V}_i$, we obtain

$$\mathrm{d_H}\left(\mathrm{ev}_{\mathbb{F}_Q^z\setminus\{\mathbf{0}\}}(f_{\mathbf{a}}^{(\mathcal{V}_i)}),\mathrm{ev}_{\mathbb{F}_Q^z\setminus\{\mathbf{0}\}}(g_{\mathbf{a}}^{(\mathcal{V}_i)})\right)\geq j-\mathbb{1}\{g_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{0})\neq\alpha\}\ .$$

Now consider the affine subspaces contributing to $M(\beta,j)$ with $\beta\neq\alpha$ (see Fig. 5.1d). As $f_{\mathbf{a}}^{(\mathcal{V}_i)},\hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)}\in\mathcal{F}$ and $f_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{0})=\alpha\neq\beta=\hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{0})$, we have

$$\mathrm{d_H}\left(\mathrm{ev}_{\mathbb{F}_Q^z}(f_{\mathbf{a}}^{(\mathcal{V}_i)}),\mathrm{ev}_{\mathbb{F}_Q^z}(\hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)})\right)\geq d_{\mathcal{F}}\ .$$

Therefore

$$\mathrm{d_H}\left(\mathrm{ev}_{\mathbb{F}_Q^z\setminus\{\mathbf{0}\}}(f_{\mathbf{a}}^{(\mathcal{V}_i)}),\mathrm{ev}_{\mathbb{F}_Q^z\setminus\{\mathbf{0}\}}(\hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)})\right)\geq d_{\mathcal{F}}-\underbrace{\mathbb{1}\{f_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{0})\neq g_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{0})\}}_{=1}.$$

Further, we have $\mathrm{d_H}\left(\mathrm{ev}_{\mathbb{F}_Q^z}(g_{\mathbf{a}}^{(\mathcal{V}_i)}),\mathrm{ev}_{\mathbb{F}_Q^z}(\hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)})\right)=j$ and

$$\mathrm{d_H}\left(\mathrm{ev}_{\mathbb{F}_Q^z\setminus\{\mathbf{0}\}}(g_{\mathbf{a}}^{(\mathcal{V}_i)}),\mathrm{ev}_{\mathbb{F}_Q^z\setminus\{\mathbf{0}\}}(\hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)})\right)=j-\mathbb{1}\{g_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{0})\neq\beta\}.$$

By the triangle inequality we get

$$\mathrm{d_H}\left(\mathrm{ev}_{\mathbb{F}_Q^z\setminus\{\mathbf{0}\}}(f_{\mathbf{a}}^{(\mathcal{V}_i)}),\mathrm{ev}_{\mathbb{F}_Q^z\setminus\{\mathbf{0}\}}(g_{\mathbf{a}}^{(\mathcal{V}_i)})\right)\geq d_{\mathcal{F}}-1-j+\mathbb{1}\{g_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{0})\neq\beta\}.$$

Finally, as $f_{\mathbf{a}}^{(\mathcal{V}_i)}\in\mathcal{F}$, a necessary condition for an affine subspace to contribute to $M_{\mathbf{a}}(\star)$ (see Fig. 5.1b) is

$$\mathrm{d_H}\left(\mathrm{ev}_{\mathbb{F}_Q^z}(f_{\mathbf{a}}^{(\mathcal{V}_i)}),\mathrm{ev}_{\mathbb{F}_Q^z}(g_{\mathbf{a}}^{(\mathcal{V}_i)})\right)\geq t_{\mathcal{F}}+1\ .$$

Again, excluding position $\mathbf{a}$ gives

$$\mathrm{d}_{\mathsf{H}}\left(\mathrm{ev}_{\mathbb{F}_Q^z\setminus\{\mathbf{0}\}}(f_{\mathbf{a}}^{(\mathcal{V}_i)}), \mathrm{ev}_{\mathbb{F}_Q^z\setminus\{\mathbf{0}\}}(g_{\mathbf{a}}^{(\mathcal{V}_i)})\right) \geq t_{\mathcal{F}} + 1 - \underbrace{\mathbb{1}\{f_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{0}) \neq g_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{0})\}}_{=\mathbb{1}\{g_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{0})\neq\alpha\}}.$$

By the same arguments, we obtain the lower bounds for even $d_{\mathcal{F}}$. The only difference to the case of odd $d_{\mathcal{F}}$ is that the erasure placed in position $\hat{g}_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{0})$ means that none of the j errors can be in position $\mathbf{a}$.

The lemma statement follows from observing that $N_{\mathbf{a}}(\alpha)$ is defined as the weighted sum over these cases. □

Lemma 5.1 provides a lower bound on the distance between the evaluations of two functions solely based on the results of the decoders applied to the restrictions to the subspaces of the (partial) partition. Similarly, bounds on the distance of a lifted affine-invariant code can also be derived as a function of the distance of the affine-invariant code [GKS13, Lemma 5.7] (see Lemma 2.3). The following formalizes the bound on the distance considered in the presented BD decoder.

Definition 5.2. *For $1 \leq z \leq \frac{m}{2}$, let $\mathrm{ev}_{\mathbb{F}_Q^z}(\mathcal{F})$ with $\mathcal{F} \subseteq \{\mathbb{F}_Q^z \to \mathbb{F}_q\}$ be an affine-invariant code of distance $d_{\mathcal{F}}$. Let $\mathcal{L}(\mathcal{F}) \subseteq \{\mathbb{F}_Q^m \to \mathbb{F}_q\}$ be the set of functions of a lifted code as in Definition 2.13. Define*

$$d_{\mathsf{low}} := \begin{cases} (d_{\mathcal{F}} - 1)\frac{Q^m-1}{Q^z-1} + 1, & \text{if } z \mid m, \\ (d_{\mathcal{F}} - 1)Q^{m-z} + 1, & \text{otherwise.} \end{cases}$$

Remark 5.1. *Note that by [GKS13, Lemma 5.7] (see Lemma 2.3) we have $d_{\mathcal{L}(\mathcal{F})} \geq d_{\mathsf{low}}$. Further, as $z \mid m$ implies $Q^z - 1 \mid Q^m - 1$, it is easy to check that d_{low} coincides with the lower bound of [GKS13, Lemma 5.7] in this case. On the other hand, d_{low} is slightly lower if $z \nmid m$, due to the fact that [GKS13, Lemma 5.7] employs arguments based on* all *affine subspaces passing through a point, while d_{low} can be obtained by only considering the evaluation in the points of a partial partition of $\mathbb{F}_Q^m$, as will be shown in the proof of Theorem 5.1.*

It remains to show the existence of a bounded distance decoder based on the distance measure introduced in Lemma 5.1.

Theorem 5.1. *For $1 \leq z \leq \frac{m}{2}$, consider an affine-invariant code $\mathrm{ev}_{\mathbb{F}_Q^z}(\mathcal{F})$ with $\mathcal{F} \subseteq \{\mathbb{F}_Q^z \to \mathbb{F}_q\}$ of distance $d_{\mathcal{F}}$. Denote by $\mathfrak{D}'$ a decoder for this code running in time $T(\mathfrak{D}')$ which corrects $t_{\mathcal{F}} := \left\lfloor \frac{d_{\mathcal{F}}-1}{2} \right\rfloor$ errors and, if $d_{\mathcal{F}}$ is even, one erasure. Let $\mathcal{L}(\mathcal{F}) \subseteq \{\mathbb{F}_Q^m \to \mathbb{F}_q\}$ be the set of functions of the lifted code as in Definition 2.13 and d_{low} be as in Definition 5.2.*

Then, there exists a decoder $\mathfrak{D}$ for the lifted affine-invariant code $\mathrm{ev}_{\mathbb{F}_Q^m}(\mathcal{L}(\mathcal{F}))$ that corrects $t_{\mathsf{low}} := \lfloor \frac{d_{\mathsf{low}}-1}{2} \rfloor$ errors in time $O(Q^{2m-z}T(\mathfrak{D}'))$ and $O(Q^{2m-2z}T(\mathfrak{D}'))$ for even and odd distance $d_{\mathcal{F}}$, respectively.

Proof. For completeness, we include a short proof that $d_{\mathcal{L}(\mathcal{F})} \geq d_{\mathsf{low}}$, as it is closely related to the principle of the presented decoder. Let $f, \tilde{f} \in \mathcal{L}(\mathcal{F})$ and assume $f(\mathbf{a}) \neq \tilde{f}(\mathbf{a})$. As the affine subspaces $\mathbf{a} + \mathcal{V}_1, ..., \mathbf{a} + \mathcal{V}_s$ intersect only in $\mathbf{a}$, we have

$$d(f, g) \geq \mathbb{1}\{f(\mathbf{a}) \neq \tilde{f}(\mathbf{a})\} + \sum_{i=1}^{s} \mathrm{d}_{\mathsf{H}} \left(\mathrm{ev}_{\mathbb{F}_Q^z \setminus \{\mathbf{0}\}}(f_{\mathbf{a}}^{(\mathcal{V}_i)}), \mathrm{ev}_{\mathbb{F}_Q^z \setminus \{\mathbf{0}\}}(\tilde{f}_{\mathbf{a}}^{(\mathcal{V}_i)}) \right).$$

By Definition 2.13 we have $f_{\mathbf{a}}^{(\mathcal{V}_i)}, \tilde{f}_{\mathbf{a}}^{(\mathcal{V}_i)} \in \mathcal{F}$. Further, from $f(\mathbf{a}) \neq \tilde{f}(\mathbf{a})$ it follows that $f_{\mathbf{a}}^{(\mathcal{V}_i)} \neq \tilde{f}_{\mathbf{a}}^{(\mathcal{V}_i)}$, so

$$\mathrm{d}_{\mathsf{H}} \left(\mathrm{ev}_{\mathbb{F}_Q^z \setminus \{\mathbf{0}\}}(f_{\mathbf{a}}^{(\mathcal{V}_i)}), \mathrm{ev}_{\mathbb{F}_Q^z \setminus \{\mathbf{0}\}}(\tilde{f}_{\mathbf{a}}^{(\mathcal{V}_i)}) \right) \geq d_{\mathcal{F}} - 1 \ .$$

The bound of $d_{\mathcal{L}(\mathcal{F})} \geq d_{\mathsf{low}}$ follows from setting s as in Eq. (5.1).

Finally, we now show the existence of a unique decoder for up to t_{low} errors by proving that in this case

$$f(\mathbf{a}) = \operatorname*{arg\,min}_{\alpha \in \mathbb{F}_q} \{N_{\mathbf{a}}(\alpha)\}$$

with $N_{\mathbf{a}}(\alpha)$ as in Definition 5.1.

Suppose that a function $g \in \{\mathbb{F}_Q^m \to \mathbb{F}_q\}$ is close to some function $f \in \mathcal{L}(\mathcal{F})$ such that

$$\mathrm{d}_{\mathsf{H}} \left(\mathrm{ev}_{\mathbb{F}_Q^m}(f), \mathrm{ev}_{\mathbb{F}_Q^m}(g) \right) \leq t_{\mathsf{low}} \ .$$

For the correct value of $\alpha = f(\mathbf{a})$, we have

$$N_{\mathbf{a}}(\alpha) \leq \mathrm{d}_{\mathsf{H}} \left(\mathrm{ev}_{\mathbb{F}_Q^m}(f), \mathrm{ev}_{\mathbb{F}_Q^m}(g) \right) \leq t_{\mathsf{low}}$$

by Lemma 5.1. For a contradiction, assume there exists an $\alpha' \in \mathbb{F}_q$ with $\alpha' \neq \alpha$ and

$N_{\mathbf{a}}(\alpha') \leq N_{\mathbf{a}}(\alpha)$. Then

$$\begin{aligned}
N_{\mathbf{a}}(\alpha) + N_{\mathbf{a}}(\alpha') \geq\ & \underbrace{\mathbb{1}\{g^{(\mathcal{V}_i)}(\mathbf{a}) \neq \alpha\} + \mathbb{1}\{g^{(\mathcal{V}_i)}(\mathbf{a}) \neq \alpha'\}}_{\geq 1} \\
& + \sum_{j=0}^{t_{\mathcal{F}}} \underbrace{(2t_{\mathcal{F}}+1)}_{\geq d_{\mathcal{F}}-1}(M_{\mathbf{a}}(\alpha, j) + M_{\mathbf{a}}(\alpha', j)) \\
& + \sum_{\beta \neq \alpha, \alpha'} \sum_{j=0}^{t_{\mathcal{F}}} \underbrace{2(d_{\mathcal{F}} - 1 - j + \delta_{\mathbf{a}}(\beta))}_{\geq d_{\mathcal{F}}-1} M(\beta, j) \\
& + \underbrace{(2t_{\mathcal{F}} + 2 - \delta_{\mathbf{a}}(\hat{\alpha}) - \delta_{\mathbf{a}}(\hat{\alpha}))}_{\geq d_{\mathcal{F}}-1} M_{\mathbf{a}}(\star) \\
\geq\ & 1 + (d_{\mathcal{F}} - 1)\left(\sum_{\alpha, j} M_{\mathbf{a}}(\alpha, j) + M_{\mathbf{a}}(\star)\right) \overset{Eq.~(5.3)}{=} d_{\mathsf{low}}\ . \qquad (5.5)
\end{aligned}$$

As $N_{\mathbf{a}}(\alpha) \leq t_{\mathsf{low}}$, this implies $N_{\mathbf{a}}(\alpha') \geq d_{\mathsf{low}} - t_{\mathsf{low}} > t_{\mathsf{low}}$ by definition of t_{low}, which contradicts the initial assumption of $N_{\mathbf{a}}(\alpha') \leq N_{\mathbf{a}}(\alpha)$. We conclude that $f(\mathbf{a}) = \arg\min_{\alpha \in \mathbb{F}}\{N_{\mathbf{a}}(\alpha)\}$.

To estimate the running time of the described algorithm first note that the values $N_{\mathbf{a}}(\alpha)\ \forall\ \mathbf{a} \in \mathbb{F}_Q^m, \alpha \in \mathbb{F}_q$ can be obtained from the decoding results

- $\mathfrak{D}'(\mathrm{ev}_{\mathbb{F}_Q^z}(g_{\mathbf{a}}^{(\mathcal{V}_i)}))\ \forall\ \mathbf{a} \in \mathbb{F}_Q^m, i \in [s]$ if $d_{\mathcal{F}}$ is odd,
- $\mathfrak{D}'(\mathrm{ev}^{*}_{\mathbb{F}_Q^z}(g_{\mathbf{a}}^{(\mathcal{V}_i)}))\ \forall\ \mathbf{a} \in \mathbb{F}_Q^m, i \in [s]$ if $d_{\mathcal{F}}$ is even, where $\mathrm{ev}^{*}_{\mathbb{F}_Q^z}$ is equal to $\mathrm{ev}_{\mathbb{F}_Q^z}$, except that an erasure is placed at the origin.

Therefore, the required number of instances of the decoder of $\mathfrak{D}'$ is proportional to the number of points $|\mathbb{F}_Q^m|$, the number of vector spaces s in the (partial) partition, and the running time $T(\mathfrak{D}')$ of the decoder $\mathfrak{D}'$. Hence, it can be estimated by $O(Q^{2m-z}T(\mathfrak{D}'))$. When $d_{\mathcal{F}}$ is odd, we have $g_{\mathbf{a}'}^{(\mathcal{V}_i)}(\mathbf{y}) = g_{\mathbf{a}}^{(\mathcal{V}_i)}(\mathbf{y} + \varphi_{\mathcal{V}_i}(\mathbf{a}' - \mathbf{a}))\ \forall\ \mathbf{a}, \mathbf{a}' \in \mathcal{V}_i$, i.e., the respective evaluations are only a permutation of the positions, and therefore need to run the local decoder $\mathfrak{D}'$ only once per affine subspace, resulting in a running time of $O(Q^{2m-2z}T(\mathfrak{D}'))$. □

Remark 5.2. *We have two additional comments:*

1. *If $\mathcal{L}(\mathcal{F})$ (or the considered subcode) is linear and has low rate, the complexity of its decoding might be reduced. Any linear $[n,k]_q$-code can be represented as a systematic code. Thus, it suffices to reconstruct k information symbols and, if necessary, encode them to get the whole codeword. In this case, the running time is $O(knQ^{-z}T(\mathfrak{D}'))$ (plus $O(nk)$ operations for encoding).*

2. *A randomized local correction algorithm for lifted Reed–Solomon and Reed–Muller codes was proposed in [GK16]. A key idea of that algorithm is similar to the algorithm of Theorem 5.1, namely: assign appropriate weights to the results of local decoding and aggregate them to get a final decision for a symbol.*

5.3 High-Error Randomized Decoding

In this section, we show that for any fixed linear affine-invariant code $\mathrm{ev}_{\mathbb{F}_Q^z}(\mathcal{F})$ with $\mathcal{F} \subseteq \{\mathbb{F}_Q^z \to \mathbb{F}_q\}$, long lifted codes $\mathrm{ev}_{\mathbb{F}_Q^m}(\mathcal{L}(\mathcal{F}))$ can correct almost all patterns of errors of relative weight less than ${}^{q-1}\!/_{q} - \epsilon$ with $\epsilon > 0$. To this end, we consider the q-ary symmetric channel (q-SC) with error probability $p_{q,\epsilon} := {}^{q-1}\!/_{q} - \epsilon$ that takes a q-ary symbol at its input and outputs either the unchanged input symbol, with probability $1 - p_{q,\epsilon}$, or one of the other $q-1$ symbols, each with probability ${}^{p_{q,\epsilon}}\!/_{q-1}$. In the following, we discuss the case when $\mathcal{F}$ is a single parity-check (SPC) code, but the same decoding algorithm also applies for any non-trivial affine-invariant code $\mathrm{ev}_{\mathbb{F}_Q^m}(\mathcal{L}(\mathcal{F}))$ with $\mathcal{F} \subsetneq \{\mathbb{F}_Q^z \to \mathbb{F}_q\}$.

We begin with a general statement on the probability of the outcome of the sum over random variables distributed according to a q-SC.

Lemma 5.2. *Let $\xi_1, \ldots, \xi_k$ be i.i.d. random variables with $\mathrm{supp}(\xi_i) = \mathbb{F}_q$ that take the value 0 with probability $1 - p_{q,\epsilon}$ and any value of $\mathbb{F}_q^\star$ with probability ${}^{p_{q,\epsilon}}\!/_{q-1}$. Then*

$$\Pr\left\{\sum_{i=1}^{k} \xi_i = 0\right\} = \frac{1}{q} + \left(1 - \frac{1}{q}\right)\left(1 - \frac{q p_{q,\epsilon}}{q-1}\right)^k$$

and for any $\alpha \in \mathbb{F}_q^\star$

$$\Pr\left\{\sum_{i=1}^{k} \xi_i = \alpha\right\} = \frac{1}{q} - \frac{1}{q}\left(1 - \frac{q p_{q,\epsilon}}{q-1}\right)^k \quad .$$

Proof. We prove the statement by induction on k. For $k = 1$, the statement holds by definition as it simply reflects the realization probabilities of a single random variable. Suppose that the statement holds for $k-1$. From the independence of the variables

ξ_i and the inductive assumption, it follows that

$$\begin{aligned}
\Pr\left\{\sum_{i=1}^{k}\xi_i = 0\right\} &= \sum_{\beta\in\mathbb{F}_q} \Pr\left\{\sum_{i=1}^{k-1}\xi_i = -\beta \text{ and } \xi_k = \beta\right\}\\
&= \sum_{\beta\in\mathbb{F}_q} \Pr\left\{\sum_{i=1}^{k-1}\xi_i = -\beta\right\}\cdot \Pr\left\{\xi_k = \beta\right\}\\
&= \left(\frac{1}{q} + \left(1-\frac{1}{q}\right)\left(1-\frac{qp_{q,\epsilon}}{q-1}\right)^{k-1}\right)(1-p_{q,\epsilon}) + \left(\frac{1}{q} - \frac{1}{q}\left(1-\frac{qp_{q,\epsilon}}{q-1}\right)^{k-1}\right)p_{q,\epsilon}\\
&= \frac{1}{q} + \left(1-\frac{qp_{q,\epsilon}}{q-1}\right)^{k-1}\left(\frac{(q-1)(1-p_{q,\epsilon})}{q} - \frac{p_{q,\epsilon}}{q}\right)\\
&= \frac{1}{q} + \left(1-\frac{1}{q}\right)\left(1-\frac{qp_{q,\epsilon}}{q-1}\right)^{k}.
\end{aligned}$$

By the symmetry of the channel, $\Pr\left\{\sum_{i=1}^{k}\xi_i = \alpha\right\}$ is the same for any $\alpha \in \mathbb{F}_q^*$. Thus,

$$\begin{aligned}
\Pr\left\{\sum_{i=1}^{k}\xi_i = \alpha\right\} &= \frac{1-\Pr\left\{\sum_{i=1}^{k}\xi_i = 0\right\}}{q-1}\\
&= \frac{1}{q} - \frac{1}{q}\left(1-\frac{qp_{q,\epsilon}}{q-1}\right)^{k}.
\end{aligned}$$

□

Lemma 5.2 provides the success probability of a simple symbol decoder of a linear code of dimension k based on taking the positions of an information set and returning the linear combination corresponding to the desired symbol. In this case, the ξ_i represent the value of the error in each position. Note that by symmetry of the channel, the coefficients of this linear combination do not affect the distribution of the sum.

Before giving the main statement of this section, we derive an upper bound on the dimension of a lifted SPC code.

Lemma 5.3. *Consider an SPC code* $\mathrm{ev}_{\mathbb{F}_Q}(\mathcal{F})$ *with* $\mathcal{F} \subseteq \{\mathbb{F}_Q \to \mathbb{F}_q\}$*, i.e., for any* $f \in \mathcal{F}$ *it holds that* $\sum_{\mathbf{a}\in\mathbb{F}_Q} f(\mathbf{a}) = 0$*. Let* $\mathcal{L}(\mathcal{F}) \subseteq \{\mathbb{F}_Q^m \to \mathbb{F}_q\}$ *be the set of functions of the lifted code as in Definition 2.13. Then the dimension of the code* $\mathrm{ev}_{\mathbb{F}_Q^m}(\mathcal{L}(\mathcal{F}))$ *is* $\Theta_Q(m^{Q-2})$*.*

Proof. First we introduce some useful notation. Denote $Q = p^l$ for a prime integer p (recall that q is a prime power and Q a power of q). For two tuples $\mathbf{u}, \mathbf{v} \in \mathbb{Z}_Q^m$ we say $\mathbf{u} = (u_1, \dots u_m)$ is less than or equal to $\mathbf{v} = (v_1, \dots, v_m)$ by the *p-partial order*, i.e., $\mathbf{u} \leq_p \mathbf{v}$, if for $u_i = \sum_{j=1}^{l} u_i^{(j)} p^{j-1}$ and $v_i = \sum_{j=1}^{l} v_i^{(j)} p^{j-1}$ it holds that $u_i^{(j)} \leq v_i^{(j)}$ for all

$i \in [m]$ and $j \in [l]$. Next, we define a slight modification of the modulo operation[2], denoted $\bmod^* Q$, that takes a non-negative integer and maps it to an element from $\mathbb{Z}_Q$ by

$$a \mod^* Q := \begin{cases} 0, & \text{if } a = 0, \\ b \in [Q-1], & \text{if } a \neq 0,\ a = b \mod Q-1. \end{cases}$$

For a tuple $\mathbf{u}$, we define its *degree* $\deg(\mathbf{u})$ to be $\sum_{i=1}^m u_i$.

In [GKS13], it was proved that $\dim_{\mathbb{F}_q}\left(\mathrm{ev}_{\mathbb{F}_Q^m}(\mathcal{L}(\mathcal{F}))\right)$ can be determined by counting the number of *good* tuples $\mathbf{v} \in \mathbb{Z}_Q^m$ such that there is no $\mathbf{u} \in \mathbb{Z}_Q^m$ with $\mathbf{u} \leq_p \mathbf{v}$ and $\deg(\mathbf{u}) \mod^* Q = Q-1$. Consider a set $\mathcal{S} \subset [m]$ with[3] $|\mathcal{S}| = Q-2$ and a tuple $\mathbf{v} \in \mathbb{Z}_Q^m$ that has the property $v_i = 1$ for $i \in S$ and $v_i = 0$ otherwise. Clearly, all $\binom{m}{Q-2}$ such tuples are good as $\deg(\mathbf{u}) \leq \deg(\mathbf{v}) \leq Q-2$ for any $\mathbf{u} \in \mathbb{Z}_Q^m$ with $\mathbf{u} \leq_p \mathbf{v}$ and therefore $\deg(\mathbf{u}) \mod^* Q \neq Q-1$. Thus, $\dim_{\mathbb{F}_q}\left(\mathrm{ev}_{\mathbb{F}_Q^m}(\mathcal{L}(\mathcal{F}))\right) \geq \Omega_Q(m^{Q-2})$, as first shown in [GKS13].

It remains to prove the upper bound on the dimension. We shall prove that the number of appropriate $\mathbf{v} \in \mathbb{Z}_Q^m$ is at most $\left(1 + \log_p Q^m\right)^{Q-2}$. Toward a contradiction, assume that it is larger than this value. First consider all $\mathbf{v}$ with $\sum_{i=1}^m \sum_{j=1}^{\log_p Q} v_i^{(j)} \leq Q-2$. Each $\mathbf{v}$ is represented by the coefficients $v_i^{(j)}$ for $i \in [m]$ and $j \in [\log_p(Q)]$ or, equivalently, by a multi-set $\mathcal{V}$ of cardinality $|\mathcal{V}| = Q-2$ containing each (i,j) with multiplicity $\delta_{\mathcal{V}}^{(i,j)} = v_i^{(j)}$ and an empty-symbol with multiplicity

$$\delta_{\mathcal{V}}^{\mathsf{empty}} = Q - 2 - \sum_{i=1}^{m} \sum_{j=1}^{\log_p Q} v_i^{(j)} .$$

Each element of this multi-set is either one of the $\log_p(Q^m)$ index pairs or the empty-symbols. Hence, the number of such *unordered* multi-sets of cardinality $Q-2$, and thereby the number of appropriate tuples $\mathbf{v}$, is upper bounded by the number of *ordered* $(Q-2)$-tuples over the same alphabet, i.e.,

$$\left|\left\{\mathbf{v} \in \mathbb{Z}_Q^m \;\middle|\; \sum_{i=1}^{m} \sum_{j=1}^{\log_p(Q)} v_i^{(j)} \leq Q-2\right\}\right| \leq \left(1 + \log_p(Q^m)\right)^{Q-2} .$$

Hence, if the number of appropriate $\mathbf{v} \in \mathbb{Z}_Q^m$ exceeds $\left(1 + \log_p Q^m\right)^{Q-2}$ there exists at least one $\mathbf{v}$ such that $\sum_{i=1}^m \sum_{j=0}^{\log_p Q-1} v_i^{(j)} \geq Q-1$. We will prove that this $\mathbf{v}$ cannot be good, i.e., there exists a $\mathbf{u} \in \mathbb{Z}_Q^m$ with $\mathbf{u} \leq_p \mathbf{v}$ such that $\deg(\mathbf{u}) \mod^* Q = Q-1$.

[2] The operation is the same as taking $\bmod Q-1$, except that any non-zero multiple of $Q-1$ is mapped to $Q-1$ instead of 0.

[3] Note that we consider Q to be fixed, so asymptotically $m > Q-2$.

To this end, consider a sequence of $Q-1$ distinct tuples $\mathbf{u}_1, \ldots, \mathbf{u}_{Q-1} \in \mathbb{Z}_Q^m$ with positive degrees such that $\mathbf{u}_{\iota-1} \leq_p \mathbf{u}_\iota \leq_p \mathbf{v}$ for $\iota \in [Q-1]$. Clearly, if all $\deg(\mathbf{u}_\iota) \mod^* Q$ are different, then there exists a $\iota \in [Q-1]$ such that

$$\deg(\mathbf{u}_\iota) \mod^* Q = Q-1 \ .$$

On the other hand, assume there exist two tuples $\mathbf{u}_\iota, \mathbf{u}_{\iota'}$ with $\iota < \iota'$ and $\deg(\mathbf{u}_\iota) \mod^* Q = \deg(\mathbf{u}_{\iota'}) \mod^* Q$. Then for the tuple $\mathbf{u} := \mathbf{u}_{\iota'} - \mathbf{u}_\iota \neq \mathbf{0}$ it holds that

$$\begin{aligned}\deg(\mathbf{u}) \mod^* Q &= \sum_{i=1}^{m} u_{\iota',i} - u_{\iota,i} \mod^* Q \\ &= \sum_{i=1}^{m} u_{\iota',i} - \sum_{i=1}^{m} u_{\iota,i} \mod^* Q \\ &= \deg(\mathbf{u}_{\iota'}) - \deg(\mathbf{u}_\iota) \mod^* Q = Q-1 \ ,\end{aligned}$$

where the final equality follows from two observations:

1. The partial order $\mathbf{u}_\iota \leq_p \mathbf{u}_{\iota'}$ and the fact that the $\mathbf{u}_\iota$ are distinct implies $\deg(\mathbf{u}_\iota) < \deg(\mathbf{u}_{\iota'})$ and therefore $\deg(\mathbf{u}_{\iota'}) - \deg(\mathbf{u}_\iota) > 0$. Hence, by definition of the operation $\mod^* Q$ we have

$$\deg(\mathbf{u}_{\iota'}) - \deg(\mathbf{u}_\iota) \mod^* Q = b$$

   for some $b \in [Q-1]$ such that $\deg(\mathbf{u}_{\iota'}) - \deg(\mathbf{u}_\iota) \mod Q-1 = b \mod Q-1$.

2. As the $\mathbf{u}_\iota$ are of positive degree, the assumption $\deg(\mathbf{u}_\iota) \mod^* Q = \deg(\mathbf{u}_{\iota'}) \mod^* Q$ implies

$$\begin{aligned}&\deg(\mathbf{u}_{\iota'}) \mod Q-1 = \deg(\mathbf{u}_\iota) \mod Q-1 \\ &\Rightarrow \ \deg(\mathbf{u}_{\iota'}) - \deg(\mathbf{u}_\iota) \mod Q-1 = 0 \ .\end{aligned}$$

It follows that the tuple $\mathbf{u}$ satisfies the two required conditions $\deg(\mathbf{u}) \mod^* Q = Q-1$ and $\mathbf{u} \leq_p \mathbf{u}_{\iota'} \leq_p \mathbf{v}$. Thus, $\mathbf{v}$ is not a *good tuple* and this contradiction completes the proof. □

With these preliminary results established, we are now ready to provide the main statement of this section.

Theorem 5.2. *Consider an SPC code* $\mathrm{ev}_{\mathbb{F}_Q}(\mathcal{F})$ *with* $\mathcal{F} \subseteq \{\mathbb{F}_Q \to \mathbb{F}_q\}$*, i.e., for any* $f \in \mathcal{F}$ *it holds that* $\sum_{\mathbf{a}\in\mathbb{F}_Q} f(\mathbf{a}) = 0$*. Let* $\mathcal{L}(\mathcal{F}) \subseteq \{\mathbb{F}_Q^m \to \mathbb{F}_q\}$ *be a the set of functions of the lifted code as in Definition 2.13.*

1. ***Parameters of the code:*** *The code is of length* $n = Q^m$ *and dimension*

$$\dim_{\mathbb{F}_q}\left(\mathrm{ev}_{\mathbb{F}_Q^m}(\mathcal{L}(\mathcal{F}))\right) = \Theta_Q\left(m^{Q-2}\right) \ .$$

2. ***High-error randomized decoder:*** *Let $f \in \mathcal{L}(\mathcal{F})$ and g be a function for which each value is obtained independently by transmitting the corresponding value of f over the q-SC with error probability $p_{q,\epsilon} := {}^{q-1}/_{q} - \epsilon$. For any $\delta > \exp(-cn)$ with some constant $c = c(Q, \epsilon)$, there exists a decoder $\mathfrak{D}$ running in time $O_Q\left(\frac{\log \frac{1}{\delta} + \log \log n}{\epsilon^{2Q-2}}\right)$ such that the error probability $\Pr\{\mathfrak{D}(g) \neq f\} < \delta$.*

Proof. The lower bound on the dimension of the code was already proved in [GKS13], the upper bound in Lemma 5.3.

Let $f \in \mathcal{L}(\mathcal{F})$ and g be a noisy version of f, where each symbol of f is corrupted by the q-SC with error probability $p_{q,\epsilon} = {}^{q-1}/_{q} - \epsilon$. We fix a partial partition of $\mathbb{F}_Q^m$ into one-dimensional vector spaces $\mathcal{V}_1, \ldots, \mathcal{V}_s$, where $s \leq {}^{Q^m-1}/_{Q-1}$.

For any $i \in [s]$ and $\mathbf{a} \in \mathbb{F}_Q^m$ it holds that $\mathrm{ev}_{\mathbb{F}_Q}(f_{\mathbf{a}}^{(\mathcal{V}_i)}) \in \mathcal{F}$ by Definition 2.13. Since $\mathcal{F}$ is an SPC code, the symbol $f(\mathbf{a})$ can be reconstructed from the evaluations of f in the points $\mathbf{b} \in \mathbf{a} + \mathcal{V}_i \setminus \{\mathbf{a}\}$ as

$$f(\mathbf{a}) = - \sum_{\mathbf{b} \in \mathbf{a} + \mathcal{V}_i \setminus \{\mathbf{a}\}} f(\mathbf{b}) \ .$$

Define the indicator random variables[4]

$$\psi_{\mathbf{a}}^{(i)} := \mathbb{1}\left\{ -f(\mathbf{a}) = \sum_{\mathbf{b} \in \mathbf{a} + \mathcal{V}_i \setminus \{\mathbf{a}\}} g(\mathbf{b}) \right\},$$

$$\psi_{\mathbf{a}}^{(i,\alpha)} := \mathbb{1}\left\{ -f(\mathbf{a}) = \alpha + \sum_{\mathbf{b} \in \mathbf{a} + \mathcal{V}_i \setminus \{\mathbf{a}\}} g(\mathbf{b}) \right\} \quad \text{for } \alpha \in \mathbb{F}_q^* \ .$$

Then, by Lemma 5.2, the expected value of these random variables is

$$\mathbb{E}[\psi_{\mathbf{a}}^{(i)}] = \frac{1}{q} + \frac{q-1}{q} \left(\frac{\epsilon q}{q-1} \right)^{Q-1} =: \hat{p}$$

and for any $\alpha \in \mathbb{F}_q^*$,

$$\mathbb{E}[\psi_{\mathbf{a}}^{(i,\alpha)}] = \frac{1}{q} - \frac{1}{q} \left(\frac{\epsilon q}{q-1} \right)^{Q-1} =: \check{p} \ .$$

Define the random variables

$$\Sigma_{\mathbf{a}} := \sum_{i \in [s]} \psi_{\mathbf{a}}^{(i)}, \quad \Sigma_{\mathbf{a}}^{(\alpha)} := \sum_{i \in [s]} \psi_{\mathbf{a}}^{(i,\alpha)} \quad \text{for } \alpha \in \mathbb{F}_q^\star \ .$$

[4] The first variable can be seen as the indication that the respective line gives a correct decoding result for this position, while the second variable indicates that the difference between the correct result and the returned value is α.

Now, consider the decoder with output $\tilde{f}$ that, for each position $\mathbf{a}$, decides for the value $\tilde{f}(\mathbf{a}) = \beta$ that maximizes the number of affine subspaces $\mathbf{a} + \mathcal{V}_i$ for which

$$\sum_{\mathbf{b} \in \mathbf{a} + \mathcal{V}_i \setminus \{\mathbf{a}\}} g(\mathbf{b}) = -\beta \ .$$

In other words, it decides for the value β in position $\mathbf{a}$ that receives the most "votes" from the decoders of the SPC code on the different lines. Then the result of the decoding is incorrect with probability

$$\begin{aligned}\Pr\left\{\exists \alpha \in \mathbb{F}_q^* \text{ s.t. } \Sigma_{\mathbf{a}} < \Sigma_{\mathbf{a}}^{(\alpha)}\right\} &\leq \sum_{\alpha \in \mathbb{F}_q^*} \Pr\left\{\Sigma_{\mathbf{a}} < \Sigma_{\mathbf{a}}^{(\alpha)}\right\} \\ &= (q-1) \Pr\left\{\Sigma_{\mathbf{a}} < \Sigma_{\mathbf{a}}^{(\alpha)}\right\} \ ,\end{aligned}$$

where the last equality holds because the channel is symmetric. Note that $\Sigma_{\mathbf{a}}$ and $\Sigma_{\mathbf{a}}^{(\alpha)}$ are binomial random variables, which we denote by $B(s, \hat{p})$ and $B(s, \check{p})$. Let $\bar{p} := {}^{\hat{p}+\check{p}}/_2$. By Hoeffding's bound [Hoe94], it holds that

$$\begin{aligned}\Pr\{B(s,\hat{p}) < \bar{p}s\} &= \Pr\{B(s,\hat{p}) < (\hat{p} - \underbrace{(\hat{p} - \bar{p})}_{\hat{p}-\check{p}/2})s\} \\ &\leq \exp\left(-2\left(\frac{\hat{p}-\check{p}}{2}\right)^2 s\right) = \exp\left(-\frac{1}{2}\left(\hat{p}-\check{p}\right)^2 s\right) \\ \Pr\{B(s,\check{p}) \geq \bar{p}s\} &= \Pr\{B(s,\hat{p}) \geq (\check{p} + \underbrace{(\bar{p} - \check{p})}_{\hat{p}-\check{p}/2})s\} = \exp\Big(-\frac{1}{2}\left(\hat{p}-\check{p}\right)^2 s\Big) \ .\end{aligned}$$

Therefore, the probability of an incorrect decoding result is bounded by

$$\begin{aligned}\Pr\left\{\exists \alpha \in \mathbb{F}_q^* \text{ s.t. } \Sigma_{\mathbf{a}} < \Sigma_{\mathbf{a}}^{(\alpha)}\right\} &\leq (q-1) \Pr\left\{\Sigma_{\mathbf{a}} < \Sigma_{\mathbf{a}}^{(\alpha)}\right\} \\ &\leq (q-1)\left(\Pr\left\{B(s,\hat{p}) < \bar{p}s\right\} + \Pr\left\{B(s,\check{p}) \geq \bar{p}s\right\}\right) \\ &\leq 2(q-1)\exp\left(-\frac{1}{2}(\hat{p}-\check{p})^2 s\right) \ .\end{aligned}$$

By Lemma 5.3, the dimension of the lifted code is $\dim_{\mathbb{F}_q}(\mathrm{ev}_{\mathbb{F}_Q^m}(\mathcal{L}(\mathcal{F}))) = \Theta_Q(m^{Q-2})$. To reconstruct the original polynomial f, it suffices to recover the evaluation of f in an information set of the code. Thus, by the union bound, the probability of error in recovering f is given by

$$O_Q\left(m^{Q-2}\exp\left(-\frac{1}{2}(\hat{p}-\check{p})^2 s\right)\right) \ .$$

This is less than δ, if

$$s = \Omega_Q \left(\frac{\log \frac{1}{\delta} + \log m}{\epsilon^{2Q-2}} \right) .$$

□

5.4 Summary and Open Problems

This chapter introduced a simple BD decoder that is applicable to any lifted affine-invariant code. The principle of this decoder is to apply a local decoder on all subspaces containing a given point and aggregate the decoding results to determine an estimate of the codeword at this position. This estimate is shown to be correct as long as the number of errors does not exceed half of an asymptotically tight bound on the minimum distance of lifted affine-invariant codes.

Furthermore, we have shown that lifted affine-invariant codes are able to correct errors of very high weight with arbitrarily high success probability, as the code length approaches infinity. To this end, we considered a decoder that, similar to the introduced BD decoder, decodes the affine-invariant (SPC) code obtained by restricting to one-dimensional subspaces. By bounding the probability of a correct result in each of these local decoding steps, we were able to estimate the probability of a correct decision in a given codeword positions, obtained through majority decision among the corresponding local decoding results.

A promising direction for future research is further exploring probabilistic decoding of lifted affine-invariant codes. While the introduced BD decoder has the advantage of providing a decoding guarantee, the result on the high-error regime shows that these codes possess a structure that is potentially able to decode far beyond the unique decoding radius with only a small probability of error.

6 Other Results on Codes with Locality

This chapter briefly summarizes a selection of other works on codes with locality. For more details the interested reader is referred to the respective publications.

6.1 Error Decoding of Locally Repairable and Partial MDS Codes

This abstract summarizes the results of [HPW21] published in the IEEE Transactions on Information Theory. *In part, the results on list decoding of LRCs have been published in the proceedings of the* 2018 IEEE International Symposium on Information Theory (ISIT) *[HW18] and the results on error decoding of PMDS codes in the proceedings of the* 2019 IEEE Information Theory Workshop (ITW) *[HPW19].*

This work considers the application of two powerful methods for increasing the decoding radius, interleaving and list decoding, to PMDS codes and LRCs. In list decoding, the goal of the decoder is to return all codewords within a given distance, called the decoding radius, of the received word. It is known that all linear codes can be decoded up to the q-ary Johnson radius [Joh62; Bas65] and decodability beyond this radius with a maximal list size polynomial in the code length is only known for a few nontrivial classes of codes [GX12; GR08; PV05]. In this work, the inherent structure of distance-optimal LRCs is used to show that this class of codes can be decoded up to a newly derived radius, which exceeds the Johnson radius as long as the normalized code rate exceeds the normalized rate in each local code. When the number of local repair sets is fixed, the maximal list size is shown to be polynomial in the code length. A general list-decoding algorithm for LRCs that achieves this radius is proposed along with an explicit realization for LRCs that are subcodes of Reed–Solomon codes (such as Tamo–Barg LRCs [TB14a]). Further, a probabilistic algorithm of low complexity for unique decoding of LRCs is given and its success probability analyzed.

The second part of the work considers the decoding of high-order interleaved PMDS codes. It is shown that for a wide range of parameters interleaved decoding can increase their decoding radius beyond the minimum distance, while the probability of successful decoding approaches 1 as the code length goes to infinity. To this end, the decoding algorithm for high-order interleaved codes by Metzner and Kapturowski [MK90] is applied. The inherent advantage of this decoder is that it is applicable to *any* linear code of sufficiently large interleaving order. To analyze its success probability, a new sufficient success condition is derived. The structure of PMDS codes is then used to prove a bound on the probability that this condition is fulfilled. Some families of PMDS code parameters are shown to correct $n-k-1$ errors, i.e., almost up to the Singleton bound, with a success probability approaching 1 as the length goes to infinity.

The two proposed methods apply to any distance-optimal LRC or high-order interleaved PMDS code, respectively, and therefore *do not require a change in the structure of the codes*. Hence, they can be viewed as a worst-case measure that can be employed as a last resort in the case of error events, without any increase in costs, e.g., in terms of storage overhead, for any system employing a code of this class.

6.2 Lifted Reed–Solomon Codes and Lifted Multiplicity Codes

This abstract summarizes the results of [HPP$^+$21] published in the IEEE Transactions on Information Theory. *In part, the results on lifted RS codes have been published in the proceedings of the* 2020 IEEE International Symposium on Information Theory (ISIT) *[HPPV20] (nominated for the* 2020 ISIT Best Student Paper Award*) and the results on lifted multiplicity codes in the proceedings of the* 2020 IEEE Information Theory Workshop (ITW) *[HPP$^+$20].*

Lifted Reed-Solomon [GKS13] and lifted multiplicity codes [KSY14] are classes of lifted affine-invariant codes, constructed from specific sets of m-variate polynomials. These codes allow for the design of high-rate codes that can recover every codeword or information symbol from many disjoint sets. As both classes of codes are based on generalizations of RM codes, it is a natural question whether these techniques can be combined to further improve the parameters. Some progress in the study of these *lifted multiplicity codes* has recently been made in [Wu15; LW19]. In [Wu15], the authors show asymptotic results for any number of variables, while [LW19] is devoted to improving these bounds on the required redundancy in the bi-variate case.

This work continues the study of lifted RS and lifted multiplicity codes. First, new lower bounds on the rate of lifted RS codes for any number of variables m are established, which match the known bounds of [GKS13] for $m=2$ and [PV19] for $m=3$, and improve upon the result of [GKS13] for any $m>2$. Next, these results are used to provide a lower bound on the rate of lifted multiplicity codes obtained from polynomials in an arbitrary number of variables, which matches the known bounds of [LW19] and improves upon those of [Wu15] for any $m \geq 2$. Specifically, a subcode of a lifted multiplicity code is investigated, formed by the linear span of m-variate monomials whose restriction to an arbitrary line in $\mathbb{F}_q^m$ is equivalent to a low-degree univariate polynomial. The tight asymptotic behavior of the fraction of such monomials is determined for the case of a fixed number of variables m and large alphabet size q.

Using these results, new explicit construction of batch codes [IKOS04] utilizing lifted Reed-Solomon codes are introduced. For some parameter regimes, these codes have a better trade-off between parameters than all previously known batch code constructions. Further, it is shown that lifted multiplicity codes have a better trade-off between redundancy and the number of disjoint recovering sets for every codeword symbol than previously known constructions, thereby providing the best known PIR codes [FVY15] for some parameter regimes. Finally, a new local self-correction algorithm for lifted multiplicity codes is presented.

6.3 Secure Codes with Accessibility for Distributed Storage

This abstract summarizes the results of [HKFW21] published in the IEEE Transactions on Information Forensics & Security*. In part, the results have been published in the proceedings of the* 2020 IEEE Global Communications Conference (GLOBECOM) *[HKFW20].*

The problem of *locality*, i.e., the ability of a distributed storage system (DSS) to recover a specified number of node failures from only a small number of surviving nodes, has been studied intensively in recent literature. This increased interest is driven by the desire to avoid large overhead and organizational complexity stemming from involving a large number of nodes in the repair process. By the same reasoning, it is not desirable for a system having to connect to a large number of nodes to recover data requested by a user. When user data is not secret, this is not an issue, as a systematic encoding of the data offers a simple and optimal solution, where a user retrieving data can obtain files from a single server. However, when any number of t nodes should not be able to learn anything about the user data, systematic encoding is no longer possible. This work considers the problem of *local access* in secure DSS, a problem closely related to secret sharing [Sha79].

Specifically, the problem of efficient access to information stored on a DSS is considered in the presence of a passive eavesdropper under different efficiency measures. First, the secure recovery of any file or subset of a given number of files from a limited number of nodes is investigated. For this case, the capacity for the alphabet independent case is established along with an explicit code construction attaining it. A similar problem was considered in [Hua17; HB16; HB17], however, the proposed solution imposes a very specific *access structure*, where accessing any file always requires accessing the same subset of nodes, while other nodes never participate in the retrieval process. To remedy this shortcoming, methods for balancing the access load, i.e., ensuring that any server is involved equally often in serving the user requests, are investigated. Further, three constructions over small fields are proposed, as well as an existence results based on a random coding argument. Then, the proposed framework is generalized to the case of limited repair bandwidth. A bound on repair bandwidth is derived along with an explicit code construction attaining it in the case of large file size.

Part II

Decoding of Interleaved Alternant Codes

7

Decoding of Interleaved Alternant Codes

Abstract

Interleaved Reed–Solomon codes admit efficient decoding algorithms which correct burst errors far beyond half the minimum distance in the random errors regime, e.g., by computing a common solution to the Key Equation for each Reed–Solomon code, as described by Schmidt et al. If this decoder does not succeed, it may either *fail* to return a codeword or *miscorrect* to an incorrect codeword, and good upper bounds on the fraction of error matrices for which these events occur are known.

The decoding algorithm immediately applies to interleaved alternant codes as well, i.e., the subfield subcodes of interleaved Reed–Solomon codes, but the fraction of decodable error matrices differs, since the error is now restricted to a subfield. This chapter presents new general lower and upper bounds on the fraction of error matrices decodable by Schmidt et al.'s decoding algorithm, thereby making it the only decoding algorithm for interleaved alternant codes for which such bounds are known.

This chapter is based on the work[1] *[HLN$^+$21] published in the* IEEE Transactions on Information Theory*. In part, the results have been published in the proceedings of the* 2020 IEEE Information Theory Workshop (ITW) *[HLN$^+$20].*

7.1 Introduction

We now turn away from codes with locality and consider decoding of a class of subcodes of GRS codes, namely, alternant codes[2]. Specifically, we are interested in ℓ-interleaved homogeneous interleaved codes, as in Definition 2.2, where the component code is an

[1] That work also contains a comparatively simple bound that is only applicable to the restrictive case where the interleaving order exceeds the number of errors. As the author of this thesis was not the principle author for that part, it is excluded here.

[2] These concepts are not mutually exclusive, as taking a subfield subcode of an LRC that is a subcode of an GRS code, such as Tamo-Barg LRCs [TB14b], yields an LRC that is a subcode of

alternant codes, as formally introduced in Section 2.2.5. Note that, aside from the other applications discussed in Section 2.2.3 and below, this class of codes is naturally connected DSSs. These systems commonly store a large number of codewords and the corruption of any number of nodes would imply errors in the same positions of all these codewords (see also Section 6.1).

Generalized Reed–Solomon (GRS) codes are among the most-studied classes of constituent codes for interleaved codes. There are several decoders for ℓ-interleaved $[n, k]$ GRS codes [KL97; BKY03; BMS04; SSB09b; Nie13; YL18] that decode up to $t_{\max} := \frac{\ell}{\ell+1}(n-k)$ errors. As this decoding radius exceeds the unique decoding radius of the interleaved RS code[3], all of these decoders necessarily fail for *some* error patterns. For an interleaved GRS code over q^m and errors of weight t, applying the decoder of [FT91; SSB09a] leads to a fraction of approximately $q^{-m(t_{\max}-t)}$ errors that are not correctable. In other words, for uniformly distributed errors of a given weight, the probability of unsuccessful decoding decreases exponentially in the difference between the maximal decoding radius and the actual error weight.

There are also various other decoding algorithms for interleaved GRS codes that decode beyond the radius $t_{\max}$ [CS03; PV04; Par07; SSB07; CH13; WZB14; PR17]. For some of these decoders, simulation results suggest a large fraction of error matrices of weight up to the claimed maximal radius can be successfully decoded, and in some very special cases, it is possible to derive bounds on this fraction. However, in general, only little is known about the fraction of decodable errors for these decoders, which are therefore not considered in this chapter.

Surprisingly, despite the abundant research on the topic of interleaved RS codes, little is known about the decoding of interleaved alternant codes. This family of codes contains some of the best-known and most-often used algebraic codes over small fields, including BCH [Hoc59; BRC60] and Goppa codes [Gop70]. In principle, alternant codes can be used as constituent codes in any of the applications of interleaved codes mentioned in Section 2.2.3. Additionally, we see several concrete reasons to specifically consider alternant codes:

- Alternant codes (especially BCH codes) are some of the most-often used algebraic codes in practice, including applications such as data storage and communications. Any system that already uses these codes and is prone to burst errors may be retroactively upgraded to enable a larger error-correction capability. For instance, in NOR and NAND flash memory, Hamming and BCH codes are considered as the standard error correction approach (see [WDPZ11; LRS06; CLS09]). Traditionally, Hamming codes are used in single-level flash memories to correct single errors, as they have a simple decoding algorithm, which only

an alternant code. There are also more intricate combinations of the concepts, such as the class of cyclic LRCs studied in [TBGC15], which combine the approaches of Tamo-Barg LRCs and BCH codes [Hoc59; BRC60].

[3] Recall that interleaving preserves the minimum distance (see Corollary 2.1)

uses a small circuit area. For multi-level flash memories, however, single-error correction is not sufficient and BCH codes with larger distance are employed. In [SRZ06], the scenario of more than four levels, i.e., storing more than two bits per flash memory cell, was investigated and it was shown that BCH codes of larger correction capability are needed. To address the fact that errors in flash memories might occur over whole bit or word lines, [YEC12] employs product codes with BCH component codes. This motivates the use of *interleaved* alternant and in particular interleaved BCH codes.

- In applications where the cost of *encoding* is dominant, e.g., in storage systems where writing occurs more often than reading an erroneous codeword, encoding in a subfield reduces the complexity. Hence, it might be advantageous to use alternant codes instead of GRS codes in some of the applications of interleaved codes. Note that *decoding* is usually done in the field of the corresponding GRS code, so the reduction in complexity is less significant.

- In some applications, such as code-based cryptography, GRS and algebraic-geometry codes cannot be used due to their vast structure, which can be turned into structural attacks on the cryptosystem. However, their subfield subcodes are in many cases unbroken (see [CMCP17, Conclusion] and [CR20, Section 7.5.3]). In particular, the codes proposed in McEliece's original paper [McE78], binary Goppa codes, have withstood efficient attacks for more than 40 years. In a McEliece-type system, the ciphertext is the sum of a codeword of a public code and a randomly chosen "error" which hides the codeword from the attacker. If multiple codewords are encrypted in parallel, they form an interleaved code and the errors can be aligned in bursts of larger weight. This approach has the potential to increase the designed security parameter, or in turn reduce the key size, and was first studied in [EWZ18; HLPW19]. This comes at the cost of a (hopefully very small) probability of unsuccessful decryption, which corresponds to the probability of unsuccessful decoding of the interleaved decoder.

As alternant codes are subcodes of GRS codes, interleaved alternant codes can be decoded by the decoders of interleaved GRS codes. However, the set of all errors of a given weight differs for interleaved alternant codes, as it only contains matrices over the subfield corresponding to the alternant code, not the field of the GRS code. Therefore, the bounds on the fraction of decodable error matrices for the decoding of interleaved GRS codes do not apply to interleaved alternant codes. Aside from a theoretical interest, it is crucial for all of the above mentioned applications to estimate this fraction, or, equivalently, the probability of successful decoding for errors drawn uniformly at random from this set.

This chapter introduces new lower bounds on the probability of success for decoding interleaved alternant codes with the decoder from [FT91; SSB09b] for uniformly distributed errors of a given weight. Further, for comparison, we also derive upper bounds on the probability of successful decoding. To the best of our knowledge, this is the first work that studies the success probability of decoding interleaved alternant codes for general parameters.

7.1.1 Outline and Contributions

To begin, Section 7.2 recaps the syndrome-based interleaved decoder from [FT91; SSB09b] and formally defines the event of a decoding failure and a miscorrection. We derive a necessary and sufficient condition for the decoder to succeed, which simplifies the subsequent analyses. Section 7.3 establishes some technical preliminary results, which are then used in Sections 7.4 and 7.5 for the derivation of the main results in this chapter:

- Theorem 7.2 provides a framework for lower bounding the probability of decoding success for interleaved alternant codes with the decoder of [FT91; SSB09b], by relating it to properties of the set of all alternant codes obtained from the generalization of specific RS codes. Based on this framework, Theorem 7.3 presents a lower bound on the probability of successful decoding by applying the technical results established in Section 7.3.

- Theorem 7.4 gives an upper bound on the probability of success for decoding interleaved alternant codes with the considered decoder. This result allows us to evaluate the performance of the lower bound presented in Theorem 7.3.

In Section 7.7 we present numerical evaluations of the new bounds for different code parameters and discuss their implications. Finally, we conclude the chapter and discuss some open problems in Section 7.8.

7.2 Decoding Algorithms for Interleaved Alternant Codes

For completeness, we begin with a brief recap of the decoding algorithm for interleaved RS codes presented in [FT91; SSB09b]. Let $\mathcal{C}^{\times \ell}$ be an ℓ-interleaved alternant code with $\mathcal{C} \in \mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$, as in Definition 2.5, and $\mathbf{H}$ be the parity-check matrix of the corresponding $\mathsf{GRS}(n, d_{\min}, \boldsymbol{\beta}, \boldsymbol{\nu})_{q^m} \in \mathbb{G}(n, d_{\min}, \boldsymbol{\beta})_{q^m}$ of $\mathcal{C}$ as in Definition 2.3.

Remark 7.1. *GRS codes allow for $\beta = 0$ to be an element of the code locators $\boldsymbol{\beta}$. However, as this complicates the decoding process described in the following (see Eq.* (7.2)*) and for consistency with [SSB09b], we restrict the code locators to be $\beta_i \neq 0$ $\forall i \in [n]$ for the remainder of this chapter.*

Consider a channel where burst errors of column weight t occur. A codeword $\mathbf{C} \in \mathcal{C}^{\times \ell}$ of an ℓ-interleaved alternant code is transmitted and the channel returns the received word

$$\mathbf{R} = \mathbf{C} + \widetilde{\mathbf{E}} \in \mathbb{F}_q^{\ell \times n} ,$$

where each row of $\mathbf{C} \in \mathbb{F}_q^{\ell \times n}$ is a codeword of $\mathcal{C}$ and $\widetilde{\mathbf{E}} \in \mathbb{F}_q^{\ell \times n}$ with $|\operatorname{colsupp}(\widetilde{\mathbf{E}})| = t$, i.e., $\widetilde{\mathbf{E}}$ has exactly t nonzero columns. Since $\mathcal{C} \subset \mathsf{GRS}$, the interleaved alternant code $\mathcal{C}^{\times \ell}$ is a subcode of an interleaved GRS code and the received word $\mathbf{R}$ can be decoded by any decoding algorithm for interleaved GRS codes and, in particular, *syndrome-based collaborative decoding algorithms*. Such algorithms, to name a few, can be found in [FT91] for BCH codes and [KL97; BKY03; SSB09b] for interleaved GRS code (for a more extensive overview see Section 2.2.3). We briefly recapitulate the decoding method below and summarize a naive version[4] of [SSB09b, Algorithm 2] in Algorithm 1.

The syndromes of each row of $\mathbf{R}$ are given by

$$\mathbf{S} := \mathbf{R} \cdot \mathbf{H}^\top = \widetilde{\mathbf{E}} \cdot \mathbf{H}^\top \in \mathbb{F}_{q^m}^{\ell \times (d_{\min}-1)} . \tag{7.1}$$

Define the *error locator polynomial* by[5]

$$\Lambda(x) := \prod_{i=1}^{t} (1 - \beta_{j_i}^{-1} x) = 1 + \Lambda_1 x + \dots + \Lambda_t x^t , \tag{7.2}$$

where the t roots $\{\beta_{j_1}, \dots, \beta_{j_t}\}$ of $\Lambda(x)$ are the code locators corresponding to the error positions. Obviously, as the decoder does not know the error positions, it is unable to directly set up this polynomial. However, as shown in [Pet60], for any $i \in [\ell]$, the vector of coefficients of $\Lambda(x)$, denoted by $\mathbf{\Lambda} = (\Lambda_1, \dots, \Lambda_t)$, fulfills the linear equations

$$\underbrace{\begin{pmatrix} \mathbf{S}[i,1] & \mathbf{S}[i,2] & \dots & \mathbf{S}[i,t] \\ \mathbf{S}[i,2] & \mathbf{S}[i,3] & \dots & \mathbf{S}[i,t+1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}[i,d_{\min}-1-t] & \mathbf{S}[i,d_{\min}-t] & \dots & \mathbf{S}[i,d_{\min}-2] \end{pmatrix}}_{\mathbf{S}^{(i)}(t)} \begin{pmatrix} \Lambda_t \\ \Lambda_{t-1} \\ \vdots \\ \Lambda_1 \end{pmatrix} = \underbrace{\begin{pmatrix} -\mathbf{S}[i,t+1] \\ -\mathbf{S}[i,t+2] \\ \vdots \\ -\mathbf{S}[i,d_{\min}-1] \end{pmatrix}}_{\mathbf{T}^{(i)}(t)} .$$

Thus, determining the error positions $\operatorname{colsupp}(\widetilde{\mathbf{E}})$ is equivalent to solving the linear

[4] This is to mean that we do not consider improvements regarding performance here, for more details, see Remark 7.2.

[5] Since we restrict the code locators to be $\beta_i \neq 0 \ \forall i \in [n]$ (see Remark 7.1) the error locator polynomial is well-defined.

system of equations $\mathfrak{S}(t)$ in t unknowns given by

$$\underbrace{\begin{pmatrix} \mathbf{S}^{(1)}(t) \\ \mathbf{S}^{(2)}(t) \\ \vdots \\ \mathbf{S}^{(\ell)}(t) \end{pmatrix}}_{\mathbf{S}(t)} \underbrace{\begin{pmatrix} \Lambda_t \\ \Lambda_{t-1} \\ \vdots \\ \Lambda_1 \end{pmatrix}}_{\mathbf{\Lambda}} = \underbrace{\begin{pmatrix} \mathbf{T}^{(1)}(t) \\ \mathbf{T}^{(2)}(t) \\ \vdots \\ \mathbf{T}^{(\ell)}(t) \end{pmatrix}}_{\mathbf{T}(t)} . \tag{7.3}$$

After determining $\mathbf{\Lambda}$ from Eq. (7.3), the roots of the error locator polynomial as in Eq. (7.2) uniquely determine the error positions and we may use a standard method for error evaluation such as *Forney's algorithm* [For65] (see [Rot06, Section 6.6]) to calculate the error values in $\hat{\mathbf{E}}$. Then, by subtracting the calculated error $\hat{\mathbf{E}}$ from $\mathbf{R}$, we obtain the estimated codeword $\hat{\mathbf{C}} = \mathbf{R} - \hat{\mathbf{E}}$. Alternatively, the positions can be declared erasures and corrected by an arbitrary erasure decoder to obtain $\hat{\mathbf{C}}$.

Algorithm 1: Syndrome-based Collaborative Decoding Algorithm

```
Input: received word $\mathbf{R}$
Output: $\hat{\mathbf{C}}$ or decoding failure
Calculate the syndrome matrix $\mathbf{S}$                                  // See Eq. (7.1)
if $\mathbf{S}[i,:] = \mathbf{0}$ for all $i$ then return $\hat{\mathbf{C}} = \mathbf{R}$
Find minimal $t^\star$ s.t. $\mathbf{S}(t^\star) \cdot \mathbf{\Lambda}^\star = \mathbf{T}(t^\star)$ has a solution $\mathbf{\Lambda}^\star$     // See Eq. (7.3)
if the solution $\mathbf{\Lambda}^\star$ is not unique then output decoding failure and stop
if $\Lambda^\star(x)$ has $t^\star$ distinct roots in $\mathbb{F}_{q^m}$ then
    Evaluate the errors $\hat{\mathbf{E}}$ by Forney's algorithm [For65]
    Calculate $\hat{\mathbf{C}} = \mathbf{R} - \hat{\mathbf{E}}$
else
    Output decoding failure
end
```

For a channel adding errors with some distribution, the collaborative decoding algorithm given in Algorithm 1 may yield three different results:

- The algorithm returns the correct result, i.e., $\hat{\mathbf{C}} = \mathbf{C}$, with *success probability* P_{suc}.
- The algorithm returns an erroneous result, i.e., $\hat{\mathbf{C}} \neq \mathbf{C}$, with *miscorrection probability* P_{misc}.
- The algorithm returns a `decoding failure`, with *failure probability* P_{fail}.

Note that from the perspective of the decoder a successful decoding event cannot be distinguished from a miscorrection.

Remark 7.2 (Practical Implementations)**.** *Algorithm 1 is a naive approach useful for proving the success probability and does not reflect an efficient implementation. For practical implementations, one can use some fast algorithm for solving the linear system of equations of Line 3, for instance, [SS11, Algorithm 3] with the complexity of $O(\ell {d_{\min}}^2)$ operations in $\mathbb{F}_{q^m}$, or the currently fastest algorithm [RS21] with complexity $O^{\sim}(\ell^{\omega-1} d_{\min})$ where $O^{\sim}$ omits the* log*-factors in $d_{\min}$ and ω is the matrix multiplication exponent, for which the best algorithms known give $\omega < 2.38$ [CW90; LG14].*

Algorithm 1 yields a *bounded distance* decoder which can decode beyond half of the minimum distance $\left\lfloor \frac{d_{\min}-1}{2} \right\rfloor$ with high probability. Clearly, the solution $\mathbf{\Lambda}^{\star}$ cannot be unique if the number of equations in Eq. (7.3) is less than the number of unknowns, which implies the maximum decoding radius of Algorithm 1 given in the following theorem.

Theorem 7.1 (Maximum Decoding Radius [SSB09b, Theorem 3])**.** *Let $\mathcal{C}^{\times \ell}$ be an ℓ-interleaved alternant code with $\mathcal{C} \in \mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$. For a received word $\mathbf{R} = \mathbf{C} + \widetilde{\mathbf{E}}$, where $\mathbf{C} \in \mathcal{C}^{\times \ell}$ and $|\operatorname{colsupp}(\widetilde{\mathbf{E}})| = t$, Algorithm 1 may only succeed, i.e., return $\hat{\mathbf{C}} = \mathbf{C}$, if t satisfies*

$$t \leq t_{\max} := \frac{\ell}{\ell+1}(d_{\min} - 1) \ . \tag{7.4}$$

Proof. There are t unknowns and $\ell(d_{\min} - 1 - t)$ equations in the linear system of equations of Eq. (7.3). Trivially, this cannot result in a unique solution for the t unknowns $\Lambda_1, \dots, \Lambda_t$ if the number of unknowns is larger than the number of equations, i.e., we may only obtain a unique solution from Eq. (7.3) if

$$t \leq \ell(d_{\min} - 1 - t) \ .$$

The statement follows from solving the inequality for t. □

By the nature of a bounded distance decoder, where correction spheres inevitably overlap for some error patterns of weight t larger than half of the minimum distance $\left\lfloor \frac{d_{\min}-1}{2} \right\rfloor$, Algorithm 1 is unsuccessful with some probability when $t > \left\lfloor \frac{d_{\min}-1}{2} \right\rfloor$. The focus of this work is to bound this success probability, assuming a uniform distribution of errors of given weight t. The techniques we use are based on analyzing Eq. (7.3) and overbounding the number of cases where $\operatorname{rank}(\mathbf{S}(t)) < t$ when t errors occurs. To bound the success probability of Algorithm 1 based on this analysis, we first show that $\operatorname{rank}(\mathbf{S}(t)) < t$ is a *necessary and sufficient* condition for Algorithm 1 to be unsuccessful. In other words, as $\operatorname{rank}(\mathbf{S}(t)) \leq t$ by design, the decoder succeeds exactly when $\mathbf{S}(t)$ is of full rank t. The arguments are an extension of the proof of [SSB09b, Lemma 2].

Lemma 7.1 (Condition for unsuccessful decoding)**.** *Let $\mathcal{C}^{\times\ell}$ be an ℓ-interleaved alternant code with $\mathcal{C} \in \mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$. For a received word $\mathbf{R} = \mathbf{C} + \widetilde{\mathbf{E}}$, where $\mathbf{C} \in \mathcal{C}^{\times\ell}$ and $|\operatorname{colsupp}(\widetilde{\mathbf{E}})| = t$, Algorithm 1 is* not successful*, i.e., returns $\hat{\mathbf{C}} \neq \mathbf{C}$ or a* `decoding failure`*, if and only if* $\operatorname{rank}(\mathbf{S}(t)) < t$*.*

Proof. Denote by $\Lambda(x)$ the *true* error locator polynomial corresponding to the t error positions (indices of nonzero columns) $|\operatorname{colsupp}(\widetilde{\mathbf{E}})|$. Then $\Lambda(x)$ has t distinct roots in $\mathbb{F}_{q^m}$ and $\boldsymbol{\Lambda}$ is a solution of $\mathfrak{S}(t)$ as in Eq. (7.3). Trivially, if $t = 0$ we have $\mathbf{S} = \mathbf{0}$ and the algorithm always returns the correct word $\mathbf{C}$ in Line 2. Now assume $t > 0$.

Necessary condition: We show that unsuccessful decoding implies $\operatorname{rank}(\mathbf{S}(t)) < t$.

The algorithm can fail only on Lines 4 and 9. Line 3 determines the *minimal* $t^\star$ such that $\mathfrak{S}(t^\star)$ has at least one solution $\boldsymbol{\Lambda}^\star$, hence $t^\star \leq t$. Note that $\boldsymbol{\Lambda}^\star$ is also a solution to $\mathfrak{S}(t)$ since $t \geq t^\star$ (see [SSB09b, Lemma 2]). If the algorithm fails on Line 4, the system $\mathfrak{S}(t^\star)$ has many solutions, hence $\mathfrak{S}(t)$ also has many solutions and $\operatorname{rank}(\mathbf{S}(t)) < t$. A failure on Line 9 occurs if $\Lambda^\star(x)$ does not have $t^\star$ different roots, which implies $\Lambda^\star(x) \neq \Lambda(x)$. Again, the system $\mathfrak{S}(t)$ has at least two solutions $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}^\star$ and $\operatorname{rank}(\mathbf{S}(t)) < t$.

Only Lines 2 and 7 can result in a miscorrected codeword. If the decoder outputs $\hat{\mathbf{C}}$ on Line 2, we have $\hat{\mathbf{C}} \neq \mathbf{C}$ as $t > 0$. Further, in this case $\mathbf{S}(t) = \mathbf{0}$, so $\operatorname{rank}(\mathbf{S}(t)) = 0 < t$.

If the algorithm outputs a miscorrected codeword $\hat{\mathbf{C}} \neq \mathbf{C}$ on Line 7, the error positions in $\mathbf{R} - \hat{\mathbf{C}}$ correspond to a $\Lambda^\star(x)$ whose coefficients $\boldsymbol{\Lambda}^\star$ are a solution to $\mathfrak{S}(t^\star)$ and hence also to $\mathfrak{S}(t)$. Thus $\mathfrak{S}(t)$ has two different solutions $\boldsymbol{\Lambda}^\star$ and $\boldsymbol{\Lambda}$, which are different since $\hat{\mathbf{C}} \neq \mathbf{C}$, and it follows that $\operatorname{rank}(\mathbf{S}(t)) < t$.

Sufficient condition: We show that unsuccessful decoding follows from $\operatorname{rank}(\mathbf{S}(t)) < t$.

Only Lines 2 and 7 can result in the output of a valid codeword. Let us assume that $\operatorname{rank}(\mathbf{S}(t)) < t$ but the decoding was successful, i.e., $\hat{\mathbf{C}} = \mathbf{C}$. If $\mathbf{C}$ was found in Line 2 then $\mathbf{R} = \mathbf{C}$ and the number of errors is $t = 0$, which contradicts the assumption $t > 0$. If the correct $\mathbf{C} = \hat{\mathbf{C}}$ was the result of Line 7, then the *minimal* $t^\star$ is equal to the *actual* number of errors t and $\boldsymbol{\Lambda}^\star = \boldsymbol{\Lambda}$; otherwise it is not possible for the polynomial $\Lambda^\star(x)$, which is of degree $t^\star$, to have t distinct roots. Since, by assumption, the algorithm did not fail, it follows from Line 4 that in this case $\mathfrak{S}(t)$ has a unique solution which contradicts our assumption that $\operatorname{rank}(\mathbf{S}(t)) < t$. □

Remark 7.3 (Application of Lemma 7.1 to interleaved RS codes)**.** *It was shown in [SSB09b, Lemma 2] that Algorithm 1 returning a* `decoding failure` *is a sufficient condition for the matrix* $\mathbf{S}(t)$ *to be rank deficient. Therefore, an upper bound on the probability of* $\operatorname{rank}(\mathbf{S}(t)) < t$ *provides an upper bound on the probability of a decoding failure. In Lemma 7.1 we extend this argument by showing that the decoder does not succeed* if and only if $\operatorname{rank}(\mathbf{S}(t)) < t$*. This implies that any bound on the probability of* $\operatorname{rank}(\mathbf{S}(t)) < t$ *is not only a bound on the probability of a decoding failure, but an upper bound on sum of the probability of a decoding failure and the probability that the decoder returns a miscorrection. As this is a property of the decoder and therefore not specific* *to interleaved alternant codes, it follows that the upper bound on the probability of a decoding failure for interleaved RS codes of [SSB09b, Theorem 7] is in fact an upper bound on the probability of the decoder being unsuccessful, i.e., a bound on* $1 - P_{\mathsf{suc}}$.

With the help of Lemma 7.1, we now present the formal condition for a decoding success, which is the basis of the bounds presented in Section 7.4.

Lemma 7.2. *Let* $\mathcal{C}^{\times \ell}$ *be an* ℓ*-interleaved alternant code with* $\mathcal{C} \in \mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$ *and* $\mathcal{E} = \{j_1, j_2, \dots, j_t\} \subset [n]$ *be a set of* $|\mathcal{E}| = t$ *error positions. For a codeword* $\mathbf{C} \in \mathcal{C}^{\times \ell}$*, an error matrix* $\widetilde{\mathbf{E}} \in \mathbb{F}_q^{\ell \times n}$ *with* $\operatorname{colsupp}(\widetilde{\mathbf{E}}) \coloneqq \mathcal{E}$ *and* $\mathbf{E} \coloneqq \widetilde{\mathbf{E}}|_{\mathcal{E}} \in \mathbb{E}_q^{(\ell,t)}$*, and a received word* $\mathbf{R} \coloneqq \mathbf{C} + \widetilde{\mathbf{E}}$*, Algorithm 1* succeeds*, i.e., returns* $\hat{\mathbf{C}} = \mathbf{C}$*, if and only if*

$$\nexists \mathbf{v} \in \mathbb{F}_{q^m}^t \setminus \{\mathbf{0}\} \text{ such that } \mathbf{H} \cdot \operatorname{diag}(\mathbf{v}) \cdot \mathbf{E}^\top = \mathbf{0} \ , \tag{7.5}$$

where $\mathbf{H} \in \mathbb{F}_{q^m}^{d_{\min}-t-1 \times t}$ *is a parity-check matrix of the code* $\mathsf{GRS}(t, d_{\min} - t, \boldsymbol{\beta}|_{\mathcal{E}}, \mathbf{1})_{q^m}$*.*

Proof. We extend and adapt the proof for interleaved RS codes from [SSB09b].

According to Lemma 7.1, Algorithm 1 may only yield a `decoding failure` or a miscorrection $\hat{\mathbf{C}} \neq \mathbf{C}$ if $\operatorname{rank}(\mathbf{S}(t)) < t$, with $\mathbf{S}(t)$ as in Eq. (7.3). In other words, the decoding may only be unsuccessful, if there exists a nonzero vector $\mathbf{u} \in \mathbb{F}_{q^m}^t$ such that $\mathbf{S}(t) \cdot \mathbf{u} = \mathbf{0}$, i.e.,

$$\exists \mathbf{u} \in \mathbb{F}_{q^m}^t \setminus \{\mathbf{0}\} \text{ such that } \mathbf{S}^{(i)}(t) \cdot \mathbf{u} = \mathbf{0} \ , \ \forall i \in [\ell] \ . \tag{7.6}$$

It is known (see [PW72, Theorem 9.9][SSB09b]) that a syndrome matrix $\mathbf{S}^{(i)}(t)$ can be decomposed into

$$\mathbf{S}^{(i)}(t) = \mathbf{H} \cdot \mathbf{F}^{(i)} \cdot \mathbf{D} \cdot \mathbf{V} \ ,$$

where the matrix $\mathbf{H}$ is defined as in the lemma statement (see also Definition 2.3),

$$\begin{aligned}
\mathbf{V} &= \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \beta_{j_1} & \beta_{j_2} & \cdots & \beta_{j_t} \\ \beta_{j_1}^2 & \beta_{j_2}^2 & \cdots & \beta_{j_t}^2 \\ \vdots & \vdots & & \vdots \\ \beta_{j_1}^{t-1} & \beta_{j_2}^{t-1} & \cdots & \beta_{j_t}^{t-1} \end{pmatrix}^\top \in \mathbb{F}_{q^m}^{t \times t} \ , \\
\mathbf{F}^{(i)} &= \operatorname{diag}(\mathbf{E}[i, :]) \in \mathbb{F}_q^{t \times t} \ , \\
\mathbf{D} &= \operatorname{diag}(\boldsymbol{\nu}'|_{\mathcal{E}}) \in \mathbb{F}_{q^m}^{t \times t} \ ,
\end{aligned}$$

and $\boldsymbol{\nu}'$ are the column multiplier of the GRS code corresponding to the alternant code $\mathcal{C}$, i.e., $\mathsf{GRS}(n, d_{\min}, \boldsymbol{\beta}, \boldsymbol{\nu}')_{q^m} \cap \mathbb{F}_q = \mathcal{C}$.

Observe that the matrices $\mathbf{D}$ and $\mathbf{V}$ are both square and of full rank. Therefore, the product $\mathbf{v} = \mathbf{D} \cdot \mathbf{V} \cdot \mathbf{u}$ defines a one-to-one mapping $\mathbf{u} \to \mathbf{v}$, such that $\mathbf{0} \to \mathbf{0}$.

Consequently, the statement Eq. (7.6) is equivalent to the statements

$$\begin{aligned} & \exists \mathbf{v} \in \mathbb{F}_{q^m}^t \setminus \{\mathbf{0}\} \text{ such that } \mathbf{H} \cdot \mathrm{diag}(\mathbf{E}[i,:]) \cdot \mathbf{v} = \mathbf{0} \ , \ \forall i \in [\ell] \\ \iff \quad & \exists \mathbf{v} \in \mathbb{F}_{q^m}^t \setminus \{\mathbf{0}\} \text{ such that } \mathbf{H} \cdot \mathrm{diag}(\mathbf{v}) \cdot \mathbf{E}[i,:] = \mathbf{0} \ , \ \forall i \in [\ell] \ , \end{aligned}$$

and the lemma statement follows. □

Above we extended and adapted the first part of the proof of the upper bound on the failure probability for interleaved RS codes in [SSB09b], where the error matrix $\widetilde{\mathbf{E}}$ is assumed to be over $\mathbb{F}_{q^m}$ (the field of RS codes). Simulation results indicate that this bound is quite tight. However, for interleaved alternant codes, $\widetilde{\mathbf{E}}$ is over $\mathbb{F}_q$ (the *subfield* of the alternant code) and the bound from [SSB09b] is not valid in this case.

Lemma 7.2 gives a necessary and sufficient condition for Algorithm 1 to succeed for an error $\widetilde{\mathbf{E}}$ with fixed $\mathcal{E} = \mathrm{supp}(\widetilde{\mathbf{E}})$ and $\widetilde{\mathbf{E}}|_{\mathcal{E}} \in \mathbb{E}_q^{(\ell,t)}$. In Section 7.4 and Section 7.5 we bound the probability of successful decoding of Algorithm 1 for a random error matrix $\widetilde{\mathbf{E}}$ where $\widetilde{\mathbf{E}}|_{\mathcal{E}} \sim \mathbb{E}_q^{(\ell,t)}$.

7.3 Technical Preliminary Results

Before deriving the bounds on the success probability of decoding interleaved alternant codes in Section 7.4, we establish some technical preliminary results required for proving the bounds.

7.3.1 Maximization of Integer Distributions

To begin, we derive a simple upper bound on the maximization of a sum of integer powers, under a restriction on the base of the power.

Definition 7.1 (Majorization Relation). *Let $\mathcal{M} = \{\{M_1, M_2, \dots, M_c\}\}$ and $\mathcal{K} = \{\{K_1, K_2, \dots, K_c\}\}$ be two (finite) multi-sets of real numbers of the same cardinality. We say that the set $\mathcal{M}$* majorizes *the set $\mathcal{K}$ and write*

$$\mathcal{M} \succ \mathcal{K} \quad \text{or} \quad \mathcal{K} \prec \mathcal{M}$$

if, after a possible renumeration, $\mathcal{M}$ and $\mathcal{K}$ satisfy the following conditions:

(1) $M_1 \geq M_2 \geq \cdots \geq M_c$ and $K_1 \geq K_2 \geq \cdots \geq K_c$;

(2) $\sum_{i=1}^{j} M_i \geq \sum_{i=1}^{j} K_i, \ \forall 1 \leq j \leq c$;

We recall a well known result on multi-sets under this majorization relation.

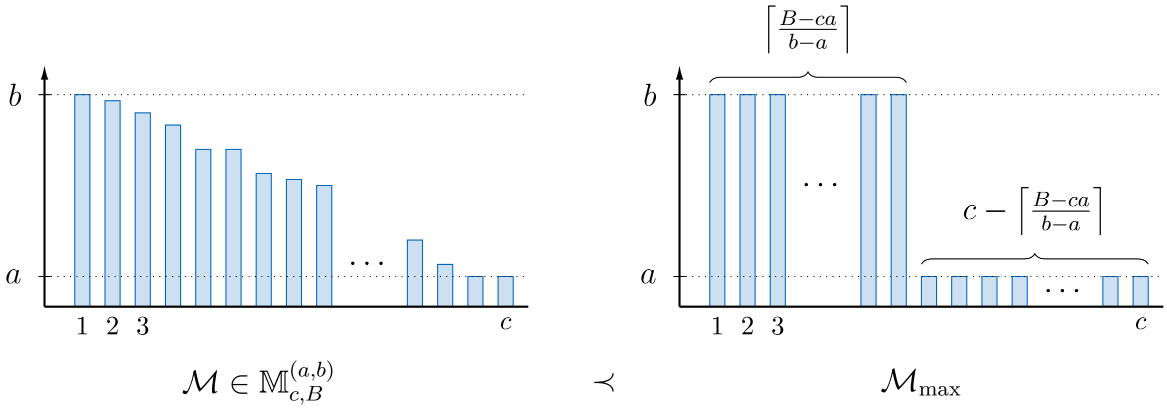


Figure 7.1: Illustration of the multiset $\mathcal{M}_{\max}$ as in the proof of Lemma 7.4, which majorizes all multisets $\mathcal{M} \in \mathbb{M}_{c,B}^{(a,b)}$.

Lemma 7.3 (Karamata's inequality [KDLM05, Theorem 1])**.** *Let $\mathcal{M} = \{\{M_1, M_2, \ldots, M_c\}\}$ and $\mathcal{K} = \{\{K_1, K_2, \ldots, K_c\}\}$ be two multi-sets of real numbers from an interval $[a, b]$. If the set $\mathcal{M} \succ \mathcal{K}$ and $f : [a, b] \to \mathbb{R}$ is a convex and nondecreasing function, then it holds that*

$$\sum_{i=1}^{c} f(M_i) \geq \sum_{i=1}^{c} f(K_i) \ . \tag{7.7}$$

For convenience of notation, we define a fixed notation for the set over which we will maximize in the following.

Definition 7.2. *Denote by $\mathbb{M}_{c,B}^{(a,b)}$ the set of all multi-sets $\mathcal{M} = \{\{M_1, \ldots, M_c\}\}$ of cardinality c with $b \geq M_1 \geq \ldots \geq M_c \geq a$ and $\sum_{M \in \mathcal{M}} M = B$.*

With these definitions established, we are now ready to give an upper bound on the sum over the results of a convex nondecreasing function evaluated on the elements of any multi-set in $\mathbb{M}_{c,B}^{(a,b)}$.

Lemma 7.4. *Let $a, c \geq 1$, $b \geq a$, $ca \leq B \leq cb$, and $\mathbb{M}_{c,B}^{(a,b)}$ be as in Definition 7.2. Then, for any function $f(x)$ that is convex and nondecreasing in the interval $a \leq x \leq b$, it holds that*

$$\max_{\mathcal{M} \in \mathbb{M}_{c,B}^{(a,b)}} \sum_{M \in \mathcal{M}} f(M) \leq \left(\frac{B - ca}{b - a} + 1 \right) (f(b) - f(a)) + cf(a)$$

Proof. By definition $\sum\limits_{M \in \mathcal{M}} M = \sum\limits_{M \in \mathrm{supp}(\mathcal{M})} \delta_{\mathcal{M}}^{M} M = B$, $\forall\, \mathcal{M} \in \mathbb{M}_{c,B}^{(a,b)}$ and it follows that for all $\mathcal{M} \in \mathbb{M}_{c,B}^{(a,b)}$ we have

$$\delta^b_{\mathcal{M}} = \frac{1}{b}\left(B - \sum_{M \in \operatorname{supp}(\mathcal{M})\setminus\{b\}} \delta^M_{\mathcal{M}} M\right) \leq \frac{B - (c - \delta^b_{\mathcal{M}})a}{b},$$
$$\delta^b_{\mathcal{M}} \leq \frac{B - ca}{b - a}.$$

Let $\mathcal{M}_{\max} = \{b, \dots, b, a, \dots, a\}$ be a multiset with $\delta^b_{\mathcal{M}_{\max}} = \left\lceil \frac{B-ca}{b-a} \right\rceil$ and $\delta^a_{\mathcal{M}_{\max}} = c - \delta^b_{\mathcal{M}_{\max}}$, as illustrated in Fig. 7.1. It can readily be seen that $\mathcal{M}_{\max} \succ \mathcal{M} \; \forall \; \mathcal{M} \in \mathbb{M}_{c,B}^{(a,b)}$ (note that $\mathcal{M}_{\max} \in \mathbb{M}_{c,B}^{(a,b)}$ if $(b-a)|(B-ca)$).

Since $f(x)$ is a convex nondecreasing function for $a \leq x \leq b$, it follows from Lemma 7.3 that

$$\sum_{M \in \mathcal{M}_{\max}} f(M) \geq \sum_{M \in \mathcal{M}} f(M) \; , \; \forall \; \mathcal{M} \in \mathbb{M}_{c,B}^{(a,b)} \; . \tag{7.8}$$

Hence, for the maximization over $\mathbb{M}_{c,B}^{(a,b)}$ we have

$$\begin{aligned} \max_{\mathcal{M} \in \mathbb{M}_{c,B}^{(a,b)}} \sum_{M \in \mathcal{M}} f(M) &\leq \sum_{M \in \mathcal{M}_{\max}} f(M) \\ &= \delta^b_{\mathcal{M}_{\max}} f(b) + (c - \delta^b_{\mathcal{M}_{\max}}) f(a) \\ &= \left\lceil \frac{B - ca}{b - a} \right\rceil (f(b) - f(a)) + cf(a) \end{aligned}$$

and the lemma statement follows. □

7.3.2 Sum over the Cardinalities of Alternant Codes

For specific subclasses of alternant codes, such as some BCH and Goppa codes, lower bounds on their dimension better than those in Lemma 2.1 are known [MS77] (see Remark 2.1). However, in general it is a difficult and open problem to predict the dimension of an alternant code for given column multipliers $\mathbf{v}$. On the other hand, the sum over the cardinality of subfield subcodes for all combinations of nonzero column multipliers is easily determined, not only for alternant codes, but for any linear code with a known weight distribution.

For a linear $[n, k, d_{\min}]_{q^m}$ code $\mathcal{C}$, define

$$B_{n,d_{\min},w}(\mathcal{C}) \coloneqq \sum_{\mathbf{v} \in (\mathbb{F}^\star_{q^m})^n} \left| \{\mathbf{c} \cdot \operatorname{diag}(\mathbf{v}) \mid \mathbf{c} \in \mathcal{C}, \operatorname{wt}(\mathbf{c}) = w\} \cap \mathbb{F}_q^n \right| .$$

Since every linear code contains the all-zero codeword and no other codeword of weight $< d_{\min}$, the sum over the cardinality of the subcodes for all combinations of non-zero column multipliers is given by

$$B_{n,d_{\min}}(\mathcal{C}) \coloneqq (q^m-1)^n + \sum_{w=d_{\min}}^{n} B_{n,d_{\min},w}(\mathcal{C}) = \sum_{\mathbf{v}\in(\mathbb{F}_{q^m}^{\star})^n} \left| \{\mathbf{c}\cdot \mathrm{diag}(\mathbf{v}) \mid \mathbf{c}\in\mathcal{C}\} \cap \mathbb{F}_q^n \right| .$$

Observe that if $\mathcal{C}$ is a $\mathsf{GRS}(n, d_{\min}, \boldsymbol{\beta}, \boldsymbol{\nu})$ code for some $\boldsymbol{\nu} \in (\mathbb{F}_{q^m}^{\star})^n$, then $B_{n,d_{\min},w}$ is the sum over the number of codewords of weight w in all alternant codes $\mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$ and $B_{n,d_{\min}}(\mathcal{C})$ is the sum over their cardinalities. Interestingly, while the weight enumerators and cardinality of a specific subfield subcode depend on $\mathbf{v}$, the sum of these values over all $\mathbf{v}$ only depends on the weight enumerators of $\mathcal{C}$.

Lemma 7.5. *Let $\mathcal{C}$ be an $[n,k,d_{\min}]_{q^m}$ code and denote by $A_w^{\mathcal{C}}$ the w-th weight enumerator of $\mathcal{C}$. Then,*

$$B_{n,d_{\min},w}(\mathcal{C}) = A_w^{\mathcal{C}} \cdot (q^m-1)^{n-w}(q-1)^w .$$

Proof. Let $\mathbf{c}$ be a codeword of $\mathcal{C}$. We have $\mathbf{c}\cdot\mathrm{diag}(\mathbf{v}) \in \mathbb{F}_q^n$ if and only if $c_i v_i \in \mathbb{F}_q$ for all $i \in [n]$. If $i \in \mathrm{supp}(\mathbf{c})$, then there are exactly $q-1$ choices of v_i for which $c_i v_i \in \mathbb{F}_q$. Else, any of the $q^m - 1$ possible values of v_i give $c_i v_i = 0 \in \mathbb{F}_q$. Hence, we have

$$\begin{aligned} B_{n,d_{\min},w}(\mathcal{C}) &= \sum_{\mathbf{v}\in(\mathbb{F}_{q^m}^{\star})^n} \left| \{\mathbf{c}\cdot\mathrm{diag}(\mathbf{v}) \mid \mathbf{c}\in\mathcal{C}, \mathrm{wt}(\mathbf{c})=w\} \cap \mathbb{F}_q^n \right| \\ &= \sum_{\substack{\mathbf{c}\in\mathcal{C} \\ \mathrm{wt}(\mathbf{c})=w}} \left| \{\mathbf{v}\in(\mathbb{F}_{q^m}^{\star})^n \mid c_i v_i \in \mathbb{F}_q \ \forall i\in[n]\} \right| \\ &= A_w^{\mathcal{C}} \cdot (q^m-1)^{n-w}(q-1)^w . \end{aligned}$$

□

The weight distribution of an MDS code only depends on its parameters, not the code itself (see Theorem 2.1). Hence, for an MDS code $\mathcal{C}$ we can omit the dependence on $\mathcal{C}$ and write

$$B_{n,d_{\min},w}^{\mathsf{MDS}} \coloneqq B_{n,d_{\min},w}(\mathcal{C}) \text{ and } B_{n,d_{\min}}^{\mathsf{MDS}} \coloneqq B_{n,d_{\min}}(\mathcal{C}) . \tag{7.9}$$

7.3.3 Probability of a Code Containing a Random Matrix

We now prove a technical lemma that bounds the probability that all rows of a randomly chosen matrix with no all-zero columns are in a code of a certain dimension. This is a refined version of [SSB09b, Lemma 3].

Lemma 7.6. *For some integers $\ell > 0, n \geq k \geq 0$, let $\mathcal{A}$ be an $[n,k]_q$ code and denote by $A_w^{\mathcal{A}}$ its w-th weight enumerator. Then, for $\mathbf{E} \sim \mathbb{E}_q^{(\ell,n)}$ we have*

$$\Pr_{\mathbf{E}\sim\mathbb{E}_q^{(\ell,n)}}\{\mathbf{E}[i,:] \in \mathcal{A} \; \forall i \in [\ell]\} \leq \frac{q^{k\ell}(q-1) - (q^\ell - 1)(q^k - 1 - A_n^{\mathcal{A}}) - (q-1)}{(q-1)(q^\ell-1)^n} \; .$$

Proof. Let $\mathcal{L} \subset \mathbb{F}_q^{\ell \times n}$ be the set of matrices whose rows are codewords of $\mathcal{A}$ and denote by $\mathcal{L}_0 \subset \mathcal{L}$ the subset of all matrices in $\mathcal{L}$ *with* at least one all-zero column. Observe that

$$\{\mathbf{E} \mid \mathbf{E}[1,:], \dots, \mathbf{E}[\ell,:] \text{ are } \mathbb{F}_q\text{-scalar multiples of } \mathbf{e}, \mathbf{e} \in \bar{\mathcal{A}} \cup \{0\}, \mathrm{wt}(\mathbf{e}) < n\} \subseteq \mathcal{L}_0 \; ,$$

where $\bar{\mathcal{A}}$ is a set of representatives[6] of $(\mathcal{A} \setminus \{0\})/\mathbb{F}_q^\star$, which is of cardinality $|\bar{\mathcal{A}}| = \frac{q^k-1}{q-1}$. If $\mathbf{e} = \mathbf{0}$ there is only one matrix, i.e., the all-zero matrix. For all other $\mathbf{e}$ with $\mathrm{wt}(\mathbf{e}) < n$ each row can be an $\mathbb{F}_q$-multiple of $\mathbf{e}$ and all these matrices are unique, if at least one row is not $\mathbf{0}$. The number of such choices is $q^\ell - 1$, so

$$\begin{aligned}|\mathcal{L}_0| &\geq (q^\ell - 1)(|\bar{\mathcal{A}}| - \underbrace{|\{\mathbf{c} \in \bar{\mathcal{A}} \mid \mathrm{wt}(\mathbf{c}) = n\}|}_{=:\frac{A_n^{\mathcal{A}}}{(q-1)}}) + 1 \\ &= \frac{(q^\ell-1)}{(q-1)}(q^k - 1 - A_n^{\mathcal{A}}) + 1 \; .\end{aligned}$$

Recall that $\mathbb{E}_q^{(\ell,n)}$ does not contain any matrices with all-zero columns by definition, so $\mathcal{L}_0 \cap \mathbb{E}_q^{(\ell,n)} = \emptyset$. As $\mathcal{L}_0 \subset \mathcal{L}$, it follows that

$$\Pr_{\mathbf{E}\sim\mathbb{E}_q^{(\ell,n)}}\{\mathbf{E}[i,:] \in \mathcal{A} \; \forall i = [\ell]\} = \frac{|\mathcal{L} \cap \mathbb{E}_q^{(\ell,n)}|}{|\mathbb{E}_q^{(\ell,n)}|} = \frac{|\mathcal{L} \setminus \mathcal{L}_0|}{|\mathbb{E}_q^{(\ell,n)}|} = \frac{|\mathcal{L}| - |\mathcal{L}_0|}{|\mathbb{E}_q^{(\ell,n)}|} \; .$$

The lemma statement follows from the observation that $|\mathcal{L}| = |\mathcal{A}|^\ell = q^{k\ell}$ and $|\mathbb{E}_q^{(\ell,n)}| = (q^\ell - 1)^n$.

□

If $|\mathcal{L}_0|$ is large, it is worthwhile to deduct it from $|\mathcal{L}|$ as in Lemma 7.6. However, for other parameters, (our best lower bound on) $|\mathcal{L}_0|$ becomes negligible compared to $|\mathcal{L}|$. Therefore, we also define a simplified version of this upper bound, where we only exclude the zero matrix from $\mathcal{L}$.

Corollary 7.1. *For some integers $\ell > 0, n \geq k \geq 0$, let $\mathcal{A}$ be an $[n,k]_q$ code. Then,*

[6] A common choice is the set of all nonzero codewords of $\mathcal{A}$ whose first nonzero entry is 1.

for $\mathbf{E} \sim \mathbb{E}_q^{(\ell,n)}$ *we have*

$$\Pr_{\mathbf{E}\sim\mathbb{E}_q^{(\ell,n)}}\{\mathbf{E}[i,:] \in \mathcal{A} \;\forall i \in [\ell]\} \leq \frac{|\mathcal{L} \setminus \{\mathbf{0}_{\ell\times n}\}|}{|\mathbb{E}_q^{(\ell,n)}|} = \frac{q^{k\ell}-1}{(q^\ell-1)^n} \; .$$

7.4 The Success Probability of Decoding Interleaved Alternant Codes

We now turn to the main topic of this chapter, namely providing bounds on the performance of the decoder of [FT91; SSB09b] (see Section 7.2) when applied to interleaved alternant codes. Recall that the success probability is given by

$$P_{\mathsf{suc}} = 1 - P_{\mathsf{fail}} - P_{\mathsf{misc}} \; ,$$

where P_{fail} and P_{misc} are the probability of a decoding failure and a miscorrection, respectively.

We begin by applying the technical results of Section 7.3 to obtain a lower bound on the success probability of decoding interleaved alternant codes that is valid for any interleaving order ℓ. The applied principle is a generalization of the approach in [SSB09b].

7.4.1 A Lower Bound on the Success Probability for any Interleaving Order ℓ

To begin, we relate the problem of bounding the probability of successful decoding to properties of the multisets $\mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$ of alternant codes for different parameters.

Theorem 7.2. *Let* $\mathcal{C}^{\times\ell}$ *be an* ℓ*-interleaved alternant code with* $\mathcal{C} \in \mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$ *and* $\mathcal{E} = \{j_1, j_2, \dots, j_t\} \subset [n]$ *be a set of* $|\mathcal{E}| = t$ *error positions. For a codeword* $\mathbf{C} \in \mathcal{C}^{\times\ell}$*, an error matrix* $\widetilde{\mathbf{E}} \in \mathbb{F}_q^{\ell\times n}$ *with* $\mathrm{supp}(\widetilde{\mathbf{E}}) \coloneqq \mathcal{E}$ *and* $\mathbf{E} \coloneqq \widetilde{\mathbf{E}}|_{\mathcal{E}} \sim \mathbb{E}_q^{(\ell,t)}$*, and a received word* $\mathbf{R} \coloneqq \mathbf{C} + \widetilde{\mathbf{E}}$*, Algorithm 1* succeeds*, i.e., returns* $\hat{\mathbf{C}} = \mathbf{C}$*, with probability*

$$P_{\mathsf{suc}}(\mathcal{C}^{\times\ell}, \mathcal{E}) \geq 1 - \sum_{w=d_{\min}-t}^{t} \sum_{\substack{\mathcal{V}\subseteq\mathcal{E}\\|\mathcal{V}|=w}} \sum_{\mathcal{A}\in\mathbb{A}(w,d_{\min}-t,\boldsymbol{\beta}|_{\mathcal{V}})} \left(\delta^{\mathcal{A}}_{\mathbb{A}(w,d_{\min}-t,\boldsymbol{\beta}|_{\mathcal{V}})}\right)^{-1} \cdot \Pr_{\mathbf{E}\sim\mathbb{E}_q^{(\ell,n)}}\{\mathbf{E}[i,\mathcal{V}] \in \mathcal{A} \;\forall\; i \in [\ell]\} \; ,$$

where $\delta^{\mathcal{A}}_{\mathbb{A}(w,d_{\min}-t,\boldsymbol{\beta}|_{\mathcal{V}})}$ *is the multiplicity of* $\mathcal{A}$ *in* $\mathbb{A}(w, d_{\min} - t, \boldsymbol{\beta}|_{\mathcal{V}})$*.*

Proof. By Lemma 7.2 the decoding of $\widetilde{\mathbf{E}}$ succeeds if and only if

$$\nexists \mathbf{v} \in \mathbb{F}_{q^m}^t \setminus \{\mathbf{0}\} \text{ such that } \mathbf{H} \cdot \mathrm{diag}(\mathbf{v}) \cdot \mathbf{E}^\top = \mathbf{0} \; ,$$

where $\mathbf{H} \in \mathbb{F}_{q^m}^{(d_{\min}-t-1)\times t}$ denotes the parity-check matrix of the code $\mathsf{GRS}(t, d_{\min} - t, \boldsymbol{\beta}|_{\mathcal{E}}, \mathbf{1})_{q^m}$, i.e., the RS codes of distance $d_{\min} - t$ with locators corresponding to the error positions.

Therefore, the probability of unsuccessful decoding is upper bounded by

$$\begin{aligned}
1-P_{\mathsf{suc}}(\mathcal{C}^{\times\ell}, \mathcal{E}) &\leq \Pr_{\mathbf{E}\sim\mathbb{E}_q^{(\ell,n)}}\{\exists\ \mathbf{v} \in \mathbb{F}_{q^m}^t \setminus \{\mathbf{0}\} \text{ s.t. } \mathbf{H}\cdot \operatorname{diag}(\mathbf{v}) \cdot \mathbf{E}^\top = \mathbf{0}\} \\
&\leq \sum_{w=1}^{t} \Pr_{\mathbf{E}\sim\mathbb{E}_q^{(\ell,n)}}\{\exists\ \mathbf{v} \in \mathbb{F}_{q^m}^t, \operatorname{wt}(\mathbf{v}) = w \text{ s.t. } \mathbf{H}\cdot \operatorname{diag}(\mathbf{v}) \cdot \mathbf{E}^\top = \mathbf{0}\} \qquad (7.10) \\
&\overset{(\mathsf{a})}{=} \sum_{w=d_{\min}-t}^{t} \Pr_{\mathbf{E}\sim\mathbb{E}_q^{(\ell,n)}}\{\exists\ \mathbf{v} \in \mathbb{F}_{q^m}^t, \operatorname{wt}(\mathbf{v}) = w \text{ s.t. } \mathbf{H}\cdot \operatorname{diag}(\mathbf{v}) \cdot \mathbf{E}^\top = \mathbf{0}\} \\
&= \sum_{w=d_{\min}-t}^{t} \sum_{\substack{\mathcal{V}\subseteq[\mathcal{E}] \\ |\mathcal{V}|=w}} \Pr_{\mathbf{E}\sim\mathbb{E}_q^{(\ell,n)}}\{\exists\ \mathcal{A} \in \mathbb{A}(w, d_{\min} - t, \boldsymbol{\beta}|_{\mathcal{V}}) \text{ s.t. } \mathbf{E}[i, \mathcal{V}] \in \mathcal{A}\ \forall\ i \in [\ell]\} \\
&\leq \sum_{w=d_{\min}-t}^{t} \sum_{\substack{\mathcal{V}\subseteq[\mathcal{E}] \\ |\mathcal{V}|=w}} \sum_{\mathcal{A}\in\mathbb{A}(w,d_{\min}-t,\boldsymbol{\beta}|_{\mathcal{V}})} (\delta^{\mathcal{A}}_{\mathbb{A}(w,d_{\min}-t,\boldsymbol{\beta}|_{\mathcal{V}})})^{-1} \Pr_{\mathbf{E}\sim\mathbb{E}_q^{(\ell,n)}}\{\mathbf{E}[i, \mathcal{V}] \in \mathcal{A}\ \forall\ i \in [\ell]\}\ ,
\end{aligned}$$

where (a) holds because any $d_{\min} - t - 1$ columns of $\mathbf{H}$ are linearly independent. □

With this connection between the multisets $\mathbb{A}(w, d_{\min} - t, \boldsymbol{\beta}|_{\mathcal{V}})$ and the probability of successful decoding $P_{\mathsf{suc}}(\mathcal{C}^{\times\ell}, \mathcal{E})$ established, we now apply the technical results of Section 7.3 to obtain a lower bound.

Theorem 7.3. *The probability of successful decoding $P_{\mathsf{suc}}(\mathcal{C}^{\times\ell}, \mathcal{E})$ as in Theorem 7.2 is lower bounded by*

$$\begin{aligned}
P_{\mathsf{suc}}(\mathcal{C}^{\times\ell}, \mathcal{E}) \geq 1 - \sum_{w=d_{\min}-t}^{t} \frac{\binom{t}{w}}{(q^m-1)(q^\ell-1)^w} \Bigg(& \frac{(q^\ell - 1)}{(q-1)} \Big(c_w + B^{\mathsf{MDS}}_{w,d_{\min}-t,w} - B^{\mathsf{MDS}}_{w,d_{\min}-t} \Big) \\
& - c_w + \left(\frac{B^{\mathsf{MDS}}_{w,d_{\min}-t} - c_w a_w}{b_w - a_w} + 1 \right) (b_w^\ell - a_w^\ell) + c_w a_w^\ell \Bigg),
\end{aligned}$$

where

$$\begin{aligned}
a_w &= \max\{1, q^{w-(d_{\min}-t-1)m}\}, \\
b_w &= q^{k_q^{\mathsf{opt.}}(w,d_{\min}-t)}, \\
c_w &= (q^m - 1)^w,
\end{aligned}$$

$B^{\mathsf{MDS}}_{w,d_{\min}-t}$ and $B^{\mathsf{MDS}}_{w,d_{\min}-t,w}$ are given in Eq. (7.9), and $k_{\mathsf{opt.}}(w, d_{\min} - t)$ is an upper bound on the dimension of a q-ary code of length w and minimum distance $d_{\min} - t$.

Proof. For a q-ary code $\mathcal{A}$ denote $k_{\mathcal{A}} \coloneqq \dim_q(\mathcal{A})$. From Theorem 7.2 we get

$$
\begin{aligned}
1 - P_{\mathsf{suc}}(\mathcal{C}^{\times \ell}, \mathcal{E}) &\leq \sum_{w=d_{\min}-t}^{t} \sum_{\substack{\mathcal{V}\subseteq[\mathcal{E}]\\ |\mathcal{V}|=w}} \sum_{\mathcal{A}\in\mathbb{A}(w,d_{\min}-t,\boldsymbol{\beta}|_{\mathcal{V}})} \left(\delta^{\mathcal{A}}_{\mathbb{A}(w,d_{\min}-t,\boldsymbol{\beta}|_{\mathcal{V}})}\right)^{-1} \\
&\qquad \cdot \Pr_{\mathbf{E}\sim\mathbb{E}_q^{(\ell,n)}} \{\mathbf{E}[i,\mathcal{V}] \in \mathcal{A} \ \forall\, i \in [\ell]\} \\
&\overset{(\mathsf{a})}{\leq} \sum_{w=d_{\min}-t}^{t} \sum_{\substack{\mathcal{V}\subseteq[t]\\ |\mathcal{V}|=w}} \sum_{\mathcal{A}\in\mathbb{A}(w,d_{\min}-t,\boldsymbol{\beta}|_{\mathcal{V}})} (q^m-1)^{-1} \frac{(q-1)q^{k_{\mathcal{A}}\ell} - (q^\ell-1)(q^{k_{\mathcal{A}}} - 1 - A_w^{\mathcal{A}}) - (q-1)}{(q-1)(q^\ell-1)^w} \\
&\overset{(\mathsf{b})}{=} \sum_{w=d_{\min}-t}^{t} \sum_{\substack{\mathcal{V}\subseteq[t]\\ |\mathcal{V}|=w}} \frac{1}{(q^m-1)(q^\ell-1)^w} \left(\frac{(q^\ell-1)}{(q-1)}(c_w + B^{\mathsf{MDS}}_{w,d_{\min}-t,w}) - c_w \right. \\
&\qquad \left. + \left(\sum_{\mathcal{A}\in\mathbb{A}(w,d_{\min}-t,\boldsymbol{\beta}|_{\mathcal{V}})} q^{k_{\mathcal{A}}\ell} - \frac{(q^\ell-1)}{(q-1)} q^{k_{\mathcal{A}}} \right) \right) \\
&\overset{(\mathsf{c})}{\leq} \sum_{w=d_{\min}-t}^{t} \frac{\binom{t}{w}}{(q^m-1)(q^\ell-1)^w} \left(\frac{(q^\ell-1)}{(q-1)}(c_w + B^{\mathsf{MDS}}_{w,d_{\min}-t,w}) - c_w \right. \\
&\qquad \left. + \max_{\mathcal{M}\in\mathbb{M}^{[a_w,b_w]}_{c_w,B^{\mathsf{MDS}}_{w,d_{\min}-t}}} \sum_{M\in\mathcal{M}} M^\ell - \frac{(q^\ell-1)}{(q-1)} M \right) \\
&= \sum_{w=d_{\min}-t}^{t} \frac{\binom{t}{w}}{(q^m-1)(q^\ell-1)^w} \left(\frac{(q^\ell-1)}{(q-1)}(c_w + B^{\mathsf{MDS}}_{w,d_{\min}-t,w} - B^{\mathsf{MDS}}_{w,d_{\min}-t}) - c_w \right. \\
&\qquad \left. + \max_{\mathcal{M}\in\mathbb{M}^{[a_w,b_w]}_{c_w,B^{\mathsf{MDS}}_{w,d_{\min}-t}}} \sum_{M\in\mathcal{M}} M^\ell \right) ,
\end{aligned}
$$

where (a) holds by Eq. (2.7) and Lemma 7.6, (b) holds as $\sum_{\mathcal{A}\in\mathbb{A}(w,d_{\min}-t,\boldsymbol{\beta}|_{\mathcal{V}})} A_w^{\mathcal{A}} = B^{\mathsf{MDS}}_{w,d_{\min}-t,w}$ (see Eq. (7.9)) and $|\mathbb{A}(w,d_{\min}-t,\boldsymbol{\beta}|_{\mathcal{V}})| = c_w$ (see Eq. (2.6)), and (c) holds as a_w and b_w are lower and upper bounds on the cardinality of all codes $\mathcal{A} \in \mathbb{A}(w, d_{\min}-t, \boldsymbol{\beta}|_{\mathcal{V}})$ (see Lemma 2.1) and because $\sum_{\mathcal{A}\in\mathbb{A}(w,d_{\min}-t,\boldsymbol{\beta}|_{\mathcal{V}})} q^{k_{\mathcal{A}}} = B^{\mathsf{MDS}}_{w,d_{\min}-t}$ by Lemma 7.5. The theorem statement follows by Lemma 7.4. □

By the use of Corollary 7.1 instead of Lemma 7.6 in inequality (a) we get a slightly simplified (though worse) lower bound.

Corollary 7.2. *The probability of successful decoding $P_{\mathsf{suc}}(\mathcal{C}, \mathcal{E})$ as in Theorem 7.2 is* *lower bounded by*

$$P_{\mathsf{suc}}(\mathcal{C}^{\times \ell}, \mathcal{E}) \geq 1 - \sum_{w=d_{\min}-t}^{t} \frac{\binom{t}{w}}{(q^m-1)(q^\ell-1)^w} \cdot \left(- c_w + \left(\frac{B^{\mathsf{MDS}}_{w,d_{\min}-t} - c_w a_w}{b_w - a_w} + 1 \right) (b_w^\ell - a_w^\ell) + c_w a_w^\ell \right) ,$$

where

$$\begin{aligned} a_w &= \max\{1, q^{w-(r-t)m}\} \\ b_w &= q^{k_q^{\mathsf{opt.}}(w,d_{\min}-t)} \\ c_w &= (q^m-1)^w, \end{aligned}$$

$B^{\mathsf{MDS}}_{w,d_{\min}-t}$ *is given in Eq. (7.9), and* $k_q^{\mathsf{opt}}(w, d_{\min}-t)$ *is an upper bound on the dimension of a* q*-ary code of length* w *and minimum distance* $d_{\min} - t$*.*

7.5 An Upper Bound on the Probability of Successful Decoding

For interleaved GRS codes it is known [SSB09b] that the probability of a decoding failure, and by Lemma 7.1 also the probability of unsuccessful decoding, decreases exponentially in the difference between the number of errors and the maximal decoding radius of Eq. (7.4). While the numerical results show that this probability is larger for interleaved alternant codes, it nevertheless quickly drops to values out of range for simulation. To evaluate the performance of the lower bounds of Section 7.4, we derive an upper bound on the probability of a decoding success, by showing that for a certain set of error matrices the decoder given in Algorithm 1 is *never* successful and then analyzing its cardinality.

We begin with a technical statement on the cardinality of the set of these "bad" matrices.

Lemma 7.7. *Denote by* $\mathbb{E}_{w\text{-}\mathsf{bad}}$ *the set of matrices* $\mathbf{E} \in \mathbb{E}_q^{(\ell,t)}$ *for which there exists a* *vector* $\mathbf{e} \in \mathbb{F}_q^\ell$ *that is a scalar multiple of at least* w *columns of* $\mathbf{E}$*. Then*

$$\begin{aligned}
\max_{w \leq \xi \leq t}\{Z^\xi\} \leq |\mathbb{E}_{w\text{-}\mathsf{bad}}| &\leq (t-w+1) \max_{w \leq \xi \leq t}\{Z^\xi\} \\
Z^\xi &:= \sum_{j=1}^{\lfloor \frac{t}{\ell} \rfloor} (-1)^{j-1} \binom{\frac{q^\ell-1}{q-1}}{j} D_j^\xi \\
D_j^\xi &:= \left(\prod_{z=0}^{j-1} \binom{t-z\xi}{\xi} \right) (q-1)^{j\xi}(q^\ell - q^j)^{t-j\xi} \ .
\end{aligned}$$

Proof. Consider the equivalence relation $\equiv_q$ on $\mathbb{F}_q^\ell \setminus \{\mathbf{0}\}$ defined by $\mathbf{v} \equiv_q \mathbf{u}$ if there exists a $\lambda \in \mathbb{F}_q^\star$ such that $\mathbf{v} = \lambda \mathbf{u}$. For a fixed vector $\mathbf{e} \in \mathbb{F}_q^\ell \setminus \{\mathbf{0}\}$ and a matrix $\mathbf{E} \in \mathbb{E}_q^{(\ell,w)}$ denote by $\delta_{\mathbf{E}}^{\mathbf{e}} = |\{i \mid E[:,i] \equiv_q \mathbf{e}\}|$ the multiplicity of $\mathbf{e}$ among the multiset of columns of $\mathbf{E}$ under the given equivalence relation. For a set of representatives $\mathcal{S} \subset \mathbb{F}_q^\ell \setminus \{\mathbf{0}\}$ under the given equivalence relation, we have

$$D_{|\mathcal{S}|}^\xi := |\{\mathbf{E} \mid \mathbf{E} \in \mathbb{E}_q^{(\ell,w)}, \delta_{\mathbf{E}}^{\mathbf{e}} = \xi \ \forall \, \mathbf{e} \in \mathcal{S}\}| = \left(\prod_{z=0}^{|\mathcal{S}|-1} \binom{t-z\xi}{\xi} \right) (q-1)^{|\mathcal{S}|\xi}(q^\ell - q^{|\mathcal{S}|})^{t-|\mathcal{S}|\xi} \ ,$$

where the first term accounts for the positions of the vectors of $\mathcal{S}$ in $\mathbf{E}$, the second term is the number of choices for the scalar coefficients of these positions, and the third term is the number of choices for the remaining columns, namely any nonzero vector that is not equivalent to any element of $\mathcal{S}$. By the principle of inclusion-exclusion we get

$$\begin{aligned}
\mathcal{Z}^\xi &:= \{\mathbf{E} \mid \exists \mathbf{e} \in \mathbb{F}_q^\ell \setminus \{\mathbf{0}\} \text{ s.t. } \delta_{\mathbf{E}}^{\mathbf{e}} = \xi, \mathbf{E} \in \mathbb{E}_q^{(\ell,w)}\} \\
Z^\xi &:= |\mathcal{Z}^\xi| = \sum_{j=1}^{\lfloor \frac{t}{\xi} \rfloor} (-1)^{j-1} \binom{\frac{q^\ell-1}{q-1}}{j} D_j^\xi \ .
\end{aligned}$$

The lemma statement follows from the observation that

$$\mathbb{E}_{w\text{-}\mathsf{bad}} = \bigcup_{j=w}^{t} \mathcal{Z}^j \ .$$

□

Using the lower bound on the cardinality of $\mathbb{E}_{w\text{-}\mathsf{bad}}$, we now derive an upper bound on the probability of successful decoding, by showing that the decoder never succeeds if the error matrix is in this set.

Theorem 7.4 (Upper Bound on P_{suc})**.** *Let* $\mathcal{C}^{\times \ell}$ *be an* ℓ*-interleaved alternant code with* $\mathcal{C} \in \mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$ *and* $\mathcal{E} = \{j_1, j_2, \ldots, j_t\} \subset [n]$ *be a set of* $|\mathcal{E}| = t$ *error positions. For a* *codeword* $\mathbf{C} \in \mathcal{C}^{\times \ell}$*, an error matrix* $\widetilde{\mathbf{E}} \in \mathbb{F}_q^{\ell \times n}$ *with* $\mathrm{supp}(\widetilde{\mathbf{E}}) \coloneqq \mathcal{E}$ *and* $\mathbf{E} \coloneqq \widetilde{\mathbf{E}}|_{\mathcal{E}} \sim \mathbb{E}_q^{(\ell,t)}$*, and a received word* $\mathbf{R} \coloneqq \mathbf{C} + \widetilde{\mathbf{E}}$ *Algorithm 1* succeeds*, i.e., returns* $\hat{\mathbf{C}} = \mathbf{C}$*, with probability*

$$P_{\mathsf{suc}}(\mathcal{C}^{\times \ell}, \mathcal{E}) \leq 1 - \frac{\max_{d_{\min}-t \leq \xi \leq t}\{Z^{\xi}\}}{(q^{\ell}-1)^t} \ ,$$

where Z^{ξ} *is given in Lemma 7.7.*

Proof. First observe that each summand in Eq. (7.10) gives a lower bound on the probability of unsuccessful decoding. Therefore, the fraction of matrices $\mathbf{E} \in \mathbb{E}_q^{(\ell,w)}$ that fulfills

$$\exists \mathbf{v} \in \mathbb{F}_{q^m}^t \text{ with } \mathrm{wt}(\mathbf{v}) = d_{\min} - t \text{ such that } \mathbf{H} \cdot \mathrm{diag}(\mathbf{v}) \cdot \mathbf{E}^{\top} = \mathbf{0} \ , \tag{7.11}$$

where $\mathbf{H} \in \mathbb{F}_{q^m}^{(d_{\min}-t-1)\times t}$ denotes the parity-check matrix of the code $\mathsf{GRS}(t, d_{\min} - t, \boldsymbol{\beta}|_{\mathcal{E}}, \mathbf{1})_{q^m}$, gives a lower bound on the probability of unsuccessful decoding $1 - P_{\mathsf{suc}}(\mathcal{C}^{\times \ell}, \mathcal{E})$. We denote by $\mathbb{E}_{w\text{-}\mathsf{bad}} \subset \mathbb{E}_q^{(\ell,t)}$ the set of matrices $\mathbf{E} \in \mathbb{E}_q^{(\ell,t)}$ that fulfills Eq. (7.11) and show that any error matrix $\mathbf{E} \in \mathbb{E}_q^{(\ell,t)}$ for which there exists a subset $\mathcal{L} \subset [t]$ of at least $d_{\min} - t$ columns such that $\mathrm{rank}(\mathbf{E}|_{\mathcal{L}}) = 1$ fulfills Eq. (7.11) and is therefore in $\mathbb{E}_{w\text{-}\mathsf{bad}}$.

Let $\mathcal{L} \subset [t]$ be a set of size $|\mathcal{L}| = d_{\min} - t$ and $\mathbf{v} \in \mathbb{F}_{q^m}^t$ be a vector with $\mathrm{supp}(\mathbf{v}) = \mathcal{L}$. Denote by $\bar{\mathbf{H}} = \mathbf{H}|_{\mathcal{L}}$, $\bar{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}|_{\mathcal{E}})|_{\mathcal{L}}$, $\bar{\mathbf{v}} = \mathbf{v}|_{\mathcal{L}}$, and $\bar{\mathbf{E}} = \mathbf{E}|_{\mathcal{L}}$ the respective restrictions to the support $\mathcal{L}$ of $\mathbf{v}$. Observe the equivalence

$$\mathbf{H} \cdot \mathrm{diag}(\mathbf{v}) \cdot \mathbf{E} = \mathbf{0} \quad \Leftrightarrow \quad \bar{\mathbf{H}} \cdot \mathrm{diag}(\bar{\mathbf{v}}) \cdot \bar{\mathbf{E}} = \mathbf{0} \ . \tag{7.12}$$

Recall that $\mathbf{E}$ has no all-zero columns by definition. As $\bar{\mathbf{H}} \cdot \mathrm{diag}(\bar{\mathbf{v}}) \in \mathbb{F}_{q^m}^{(d_{\min}-t-1)\times d_{\min}-t}$ is the parity check matrix of a GRS code, it is of full-rank $d_{\min} - t - 1$ and the dimension of its right kernel is exactly 1. We conclude that for any $\mathbf{E} \in \mathbb{E}_q^{(\ell,t)}$ that fulfills Eq. (7.12) there *necessarily* exists a subset $\mathcal{L}$ of $d_{\min} - t$ columns such that $\mathrm{rank}(\mathbf{E}|_{\mathcal{L}}) = 1$.

To show that this is also sufficient, first note that all rows $\bar{\mathbf{E}}[i,:]$ of this rank 1 matrix are necessarily scalar multiples of some vector $\mathbf{e} \in (\mathbb{F}_q^{\star})^{d_{\min}-t}$, where at least one scalar is nonzero (recall that $\mathbf{E}$ does not have any all-zero columns). For any fixed $\mathbf{v} \in \mathbb{F}_{q^m}^t$ with $\mathrm{supp}(\mathbf{v}) = \mathcal{L}$, the matrix $\bar{\mathbf{H}}$ is the parity-check matrix of a $[d_{\min} - t, 1, d_{\min} - t]$ GRS code, and therefore the $\mathbb{F}_{q^m}$-kernel of $\bar{\mathbf{H}}$ consists of the $\mathbb{F}_{q^m}$-scalar multiples of one vector $\mathbf{e}' \in (\mathbb{F}_q^{\star})^{d_{\min}-t}$. Further, as $\mathbf{v}$ can be any vector of support $\mathrm{supp}(\mathbf{v}) = \mathcal{L}$, there exists a $\mathbf{v}$ such that $\bar{\mathbf{H}} \cdot \mathrm{diag}(\bar{\mathbf{v}}) \cdot \mathbf{e}' = \mathbf{0}$ *for any* $\mathbf{e}' \in (\mathbb{F}_{q^m}^{\star})^{d_{\min}-t}$, and, in particular, for any $\mathbf{e} \in (\mathbb{F}_q^{\star})^{d_{\min}-t}$. It follows that there exists a $\mathbf{v}$ such that Eq. (7.12) is fulfilled and we conclude that the condition is also *sufficient.*

A set $\mathcal{L} \subset [t]$ with $|\mathcal{L}| = d_{\min} - t$ such that $\mathrm{rank}(\mathbf{E}|_{\mathcal{L}}) = 1$ exists if and only if a subset of $d_{\min} - t \leq \xi \leq t$ columns in $\mathbf{E}$ are equivalent. Thus, by Lemma 7.7, the probability of successful decoding is bounded from above by

$$P_{\mathsf{suc}}(\mathcal{C}^{\times \ell}, \mathcal{E}) \leq 1 - \frac{|\mathbb{E}_{w\text{-}\mathsf{bad}}|}{|\mathbb{E}_q^{(\ell,w)}|} \leq 1 - \frac{\max_{d_{\min}-t \leq \xi \leq t}\{Z^{\xi}\}}{(q^{\ell}-1)^t} \ .$$

□

7.6 Generalization of an Upper Bound on the Probability of Miscorrection

Before moving on to the numerical comparison between the derived bounds, we establish one last ingredient required for the interpretation of the results, namely, the probability of a miscorrection P_{misc}. To this end, we adapt the upper bound on P_{misc} for interleaved RS codes from [SSB09b] to interleaved alternant codes. The strategy of this bound applies for any decoder that possess the following property.

Definition 7.3 (ML certificate property [SSB09b, Definition 3]). *Consider a code $\mathcal{C}$ and a received word $\mathbf{R} = \mathbf{C} + \widetilde{\mathbf{E}}$ with $\mathbf{C} \in \mathcal{C}$. A decoder of $\mathcal{C}$ is said to have the* ML certificate property *if it always either returns $\hat{\mathbf{C}} = \arg\min_{\hat{\mathbf{C}} \in \mathcal{C}} |\operatorname{colsupp}(\hat{\mathbf{C}}, \mathbf{R})|$ or declares a* `decoding failure`*.*

It was shown in [SSB09b, Theorem 5] that the decoder of [SSB09b] for interleaved RS codes has the ML certificate property. As this is a property of the decoder, it clearly also holds when the decoder is applied to any subcode of the interleaved RS codes and, in particular, for interleaved alternant codes. However, the bound given in [SSB09b, Theorem 6] depends on the weight enumerators of the considered code, which are unknown for (interleaved) alternant codes. To circumvent this issue, we slightly generalize [SSB09b, Theorem 6] by employing general upper bounds on the weight enumerators, thereby making it independent of the specific linear code used.

Theorem 7.5 (Johnson Bound on Weight Enumerators [Bas65; Joh62] (see [MS77, Ch. 17])). *For any code $\mathcal{C}$ of length n and distance $d_{\min}$ over $\mathbb{F}_q$ it holds that*

$$A_w^{\mathcal{C}} \leq \frac{\theta_q d_{\min} n}{w^2 - \theta_q n(2w - d_{\min})}$$

with $\theta_q = 1 - \frac{1}{q}$, provided the denominator is positive.

This bound only applies if the denominator is positive, but we can also make a statement if this is not the case.

Theorem 7.6 (General Bound on Weight Enumerators [MS77, Theorem 4, Chapter 17])**.** *For any code $\mathcal{C}$ of length n and distance $d_{\min}$ it holds that*

$$A_w^{\mathcal{C}} \leq \frac{n}{w} \hat{A}_{w-1}^{[n-1,d_{\min}]} \ ,$$

where $\hat{A}_{w-1}^{[n-1,d_{\min}]}$ is an upper bound on the $(w-1)$-th weight enumerator of an arbitrary code of length $n-1$ and distance $d_{\min}$.

Proof. Note that [MS77, Theorem 4, Chapter 17] only considers binary codes. However, it is easy to see that it holds for any q by applying the same double counting argument for the number of nonzero positions, instead of the number of ones. □

Finally, we replace the explicit dependence on the weight enumerators in [SSB09b, Theorem 6] by the generic (code independent) bounds of Theorems 7.5 and 7.6, to obtain an upper bound on the probability of a miscorrection that is valid for any linear code and decoder that exhibits the ML certificate property.

Theorem 7.7 (Miscorrection Probability, generalizes [SSB09b, Theorem 6])**.** *Let $\mathcal{C}$ be a linear code of length n and minimum distance $d_{\min}$ over $\mathbb{F}_Q$ decoded with a decoder that exhibits the* ML certificate property *as in Definition 7.3. Assume that the decoding radius of this decoder is $t_{\max}$ and that it decodes a codeword that is corrupted by t errors. Let*

$$\hat{A}_w^{[n,d_{\min}]} = \begin{cases} \left\lfloor \frac{\theta_Q d_{\min} n}{w^2 - \theta_Q n(2w-d_{\min})} \right\rfloor, & \text{if } w^2 > \theta_Q n(2w-d_{\min}), \\ \hat{A}_{w-1}^{[n-1,d_{\min}]}, & \text{else,} \end{cases}$$

with $\theta_Q = 1 - \frac{1}{Q}$, and

$$U(Q,t,w,\rho) = \sum_{i=\lceil \frac{t+w-\rho}{2} \rceil}^{t+w-\rho} \binom{w}{i} \binom{i}{\rho-(t+w)+2i} \binom{n-w}{t-i} \cdot (Q-2)^{\rho-(t+w)+2i} (Q-1)^{t-i} \ ,$$

where $0^0 := 1$.

Then, the probability of a miscorrection is

$$P_{\mathsf{misc}}(\mathcal{C},t) \leq \frac{\sum_{w=d_{\min}}^{t+t_{\max}} \hat{A}_w^{[n,d_{\min}]} \sum_{\rho=0}^{\min\{t,t_{\max}\}} U(Q,t,w,\rho)}{\binom{n}{t}(Q-1)^t} \ ,$$

Proof. Trivially, the bound of [SSB09b, Theorem 6] is increasing in the weight enumerator A_w, so replacing them with the upper bound $\hat{A}_w^{[n,d_{\min}]}$ obtained from Theorems 7.5 and 7.6 results in a valid upper bound on P_{misc}. □

The bound of Theorem 7.7 is valid for any linear code. For completeness, we explicitly relate its parameters to those of interleaved alternant codes.

Corollary 7.3 (Miscorrection Probability of Interleaved Alternant Codes)**.** *Let $\mathcal{C}^{\times \ell}$ be an ℓ-interleaved alternant code with $\mathcal{C} \in \mathbb{A}(n, d_{\min}, \boldsymbol{\beta})$ and $\mathcal{E} = \{j_1, j_2, \ldots, j_t\} \subset [n]$ be a set of $|\mathcal{E}| = t$ error positions. For an error matrix $\widetilde{\mathbf{E}} \in \mathbb{F}_q^{\ell \times n}$ with* $\operatorname{supp}(\widetilde{\mathbf{E}}) := \mathcal{E}$ *and $\widetilde{\mathbf{E}}|_{\mathcal{E}} \sim \mathbb{E}_q^{(\ell,t)}$ and any decoder with the ML certificate property, the probability of a* miscorrection *when correcting $t \leq t_{\max}$ errors is upper bounded by Theorem 7.7 with $Q = q^{\ell}$.*

Proof. Fix a basis of $\mathbb{F}_{q^\ell}$ over $\mathbb{F}_q$ and regard the code $\mathcal{C}^{\times \ell}$ as a scalar code over $\mathbb{F}_{q^\ell}$. By Corollary 2.1 the minimum distance of the scalar code is $d_{\min}$ and as Theorem 7.7 holds for any linear scalar code, the statement follows. □

Corollary 7.3 presents a rather rough upper bound, as it is independent of both, the specific alternant code and its dimension. Nevertheless, it is sufficient for the purpose of showing that the probability of unsuccessful decoding of interleaved alternant codes is dominated by the failure probability, as evident from the numerical results presented in the next section.

7.7 Discussion and Numerical Results

In Sections 7.4 and 7.5 we have established lower and upper bounds on the probability of successful decoding

$$P_{\mathsf{suc}} = 1 - P_{\mathsf{fail}} - P_{\mathsf{misc}}$$

for the decoding algorithm of [FT91; SSB09b] when applied to interleaved alternant codes for uniformly distributed errors of a given weight. In the following we present and discuss some numerical results, where we compare these upper and lower bounds[7]. In order to better emphasize the individual contributions of failures and miscorrections, we further include an upper bound on the probability of miscorrection P_{misc}, given in Theorem 7.7, in the plots of Figs. 7.2 to 7.5. We label, summarize, and describe the different bounds and versions thereof in Table 7.1 and, for convenience and clarity, refer to them by their respective label for the remainder of this section. Further, we fix the code length to be $n = q^m - 1$, i.e., given the base field size q and extension degree m, we construct the longest possible RS/alternant codes, while excluding $\beta_i = 0$ as a code locator.

Aside from the comparison of the lower and upper bounds on the success probability, it is also interesting to see how the probability of successful decoding of an interleaved alternant codes compares to that of the corresponding interleaved GRS code over $\mathbb{F}_{q^m}$.

[7] For better presentation, we plot the respective bounds on the probability of *unsuccessful* decoding $1 - P_{\mathsf{suc}}$ instead of the bounds on P_{suc}.

Table 7.1: Overview of the bounds shown in Figs. 7.2 to 7.5

Label	Defined in	Description
L.RS	Theorem 7.8	Lower bound on the probability of successful decoding for interleaved RS codes
L.A	Theorem 7.3	Lower bound on the probability of successful decoding for interleaved alternant codes where the minimum of the Singleton, Griesmer, Hamming, Plotkin, Elias, and Linear Programming bound is used for k_q^{opt}.
L.A1	Theorem 7.3	Lower bound on the probability of successful decoding for interleaved alternant codes, where the Singleton bound is used for k_q^{opt}.
L.A2	Corollary 7.2	Simplified version of Theorem 7.3. The minimum of the Singleton, Griesmer, Hamming, Plotkin, Elias, and Linear Programming bound is used for k_q^{opt}.
M	Corollary 7.3	Upper bound on the probability of a miscorrection for interleaved alternant codes. We assume that the decoding radius of the interleaved decoder is $\left\lfloor \frac{\ell}{\ell+1}(d_{\min}-1) \right\rfloor$, i.e., the largest number of errors for which the *RS interleaved decoder*, given in Algorithm 1, would succeed (see Remark 7.4).
U	Theorem 7.4	Upper bound on the probability of successful decoding for interleaved alternant codes.
SIM	Remark 7.4	Threshold number of errors such that for all numbers of errors left of the indicated line, the interleaved alternant decoder succeeds with a probability of $P_{\mathsf{suc}} > 0.9$ obtained by simulation with 100 decoding iterations per parameter set.

Such a bound was derived[8] and shown to be close to the probability of successful decoding obtained from simulation in [SSB09b]. For the reader's convenience we restate it in Theorem 7.8 and assign it the label L.RS. Note that the decoder employed in [SSB09b] is equivalent to the decoder considered in this work (see Algorithm 1), however the error matrix $\widetilde{\mathbf{E}}$ is assumed to be over $\mathbb{F}_{q^m}$ (the field of the RS code) in Theorem 7.8.

Theorem 7.8 (Probability of successful decoding for interleaved RS codes [SSB09b, Theorem 7]). *Let $\mathcal{C}^{\times\ell}$ be an ℓ-interleaved GRS code with $\mathcal{C} \in \mathbb{G}(n, d_{\min}, \boldsymbol{\beta})_{q^m}$ as in Definition 2.3 and $\mathcal{E} = \{j_1, j_2, \dots, j_t\} \subset [n]$ be a set of $|\mathcal{E}| = t$ error positions. For a codeword $\mathbf{C} \in \mathcal{C}^{\times\ell}$, an error matrix $\widetilde{\mathbf{E}} \in \mathbb{F}_{q^m}^{\ell\times n}$ with* $\operatorname{supp}(\widetilde{\mathbf{E}}) := \mathcal{E}$ *and $\mathbf{E} := \widetilde{\mathbf{E}}|_{\mathcal{E}} \sim \mathbb{E}_{q^m}^{(\ell,t)}$, and a received word $\mathbf{R} := \mathbf{C} + \widetilde{\mathbf{E}}$, Algorithm 1* succeeds*, i.e., returns $\hat{\mathbf{C}} = \mathbf{C}$, with probability*

$$P_{\mathsf{suc}}(\mathcal{C}^{\times\ell}, \mathcal{E}) \geq 1 - \left(\frac{q^{m\ell} - \frac{1}{q^m}}{q^{m\ell} - 1}\right)^t \cdot \frac{q^{-m(\ell+1)(t_{\max,\mathsf{RS}} - t)}}{q^m - 1} \tag{7.13}$$

where $t_{\max,\mathsf{RS}} = \frac{\ell}{\ell+1}(d_{\min} - 1)$.

Before we discuss the numerical evaluations of the bounds, we make an important observation based on the *simulation* results.

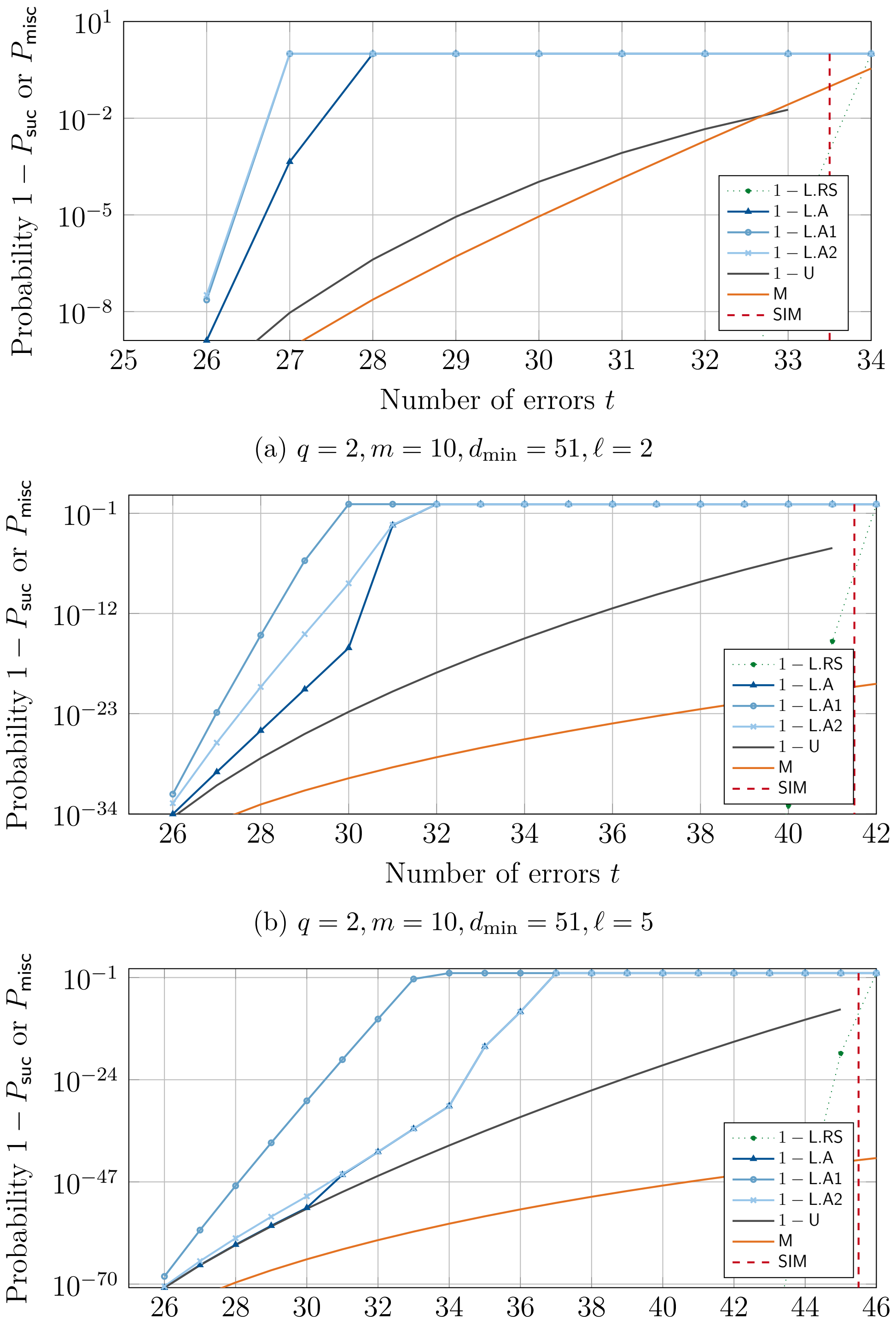
Remark 7.4. *For most parameters the provided lower bounds on the success probability of decoding interleaved alternant codes do not provide a nontrivial bound for the same decoding radius as the bounds for interleaved RS codes of [SSB09b]. To determine the real decoding threshold, i.e., the smallest number of errors for which the decoder succeeds with nonnegligible probability[9], we rely on simulation results. This threshold is indicated in the plots and labeled* SIM*. Notably, for all tested parameters,* the threshold for interleaved alternant codes is the same as for interleaved RS codes*, i.e., the simulation results imply that the collaborative decoding of errors in an interleaved alternant code succeeds w.h.p. for any number of errors t with*

$$t \leq \frac{\ell}{\ell+1}(d_{\min} - 1) = t_{\max,\mathsf{RS}} \ .$$

The numerical evaluations of the bounds are given in Figs. 7.2 to 7.5 for different base field size q, extension degree m, and distance $d_{\min}$, each for varying interleaving order ℓ:

[8] The bound in [SSB09b] is presented as a bound on the probability of failure, but it is in fact a bound on the probability of unsuccessful decoding (see Remark 7.3).

[9] We arbitrarily choose this probability to be $P_{\mathsf{suc}} > 0.9$ and run 100 decoding iterations for each parameter set to determine the decoding threshold.

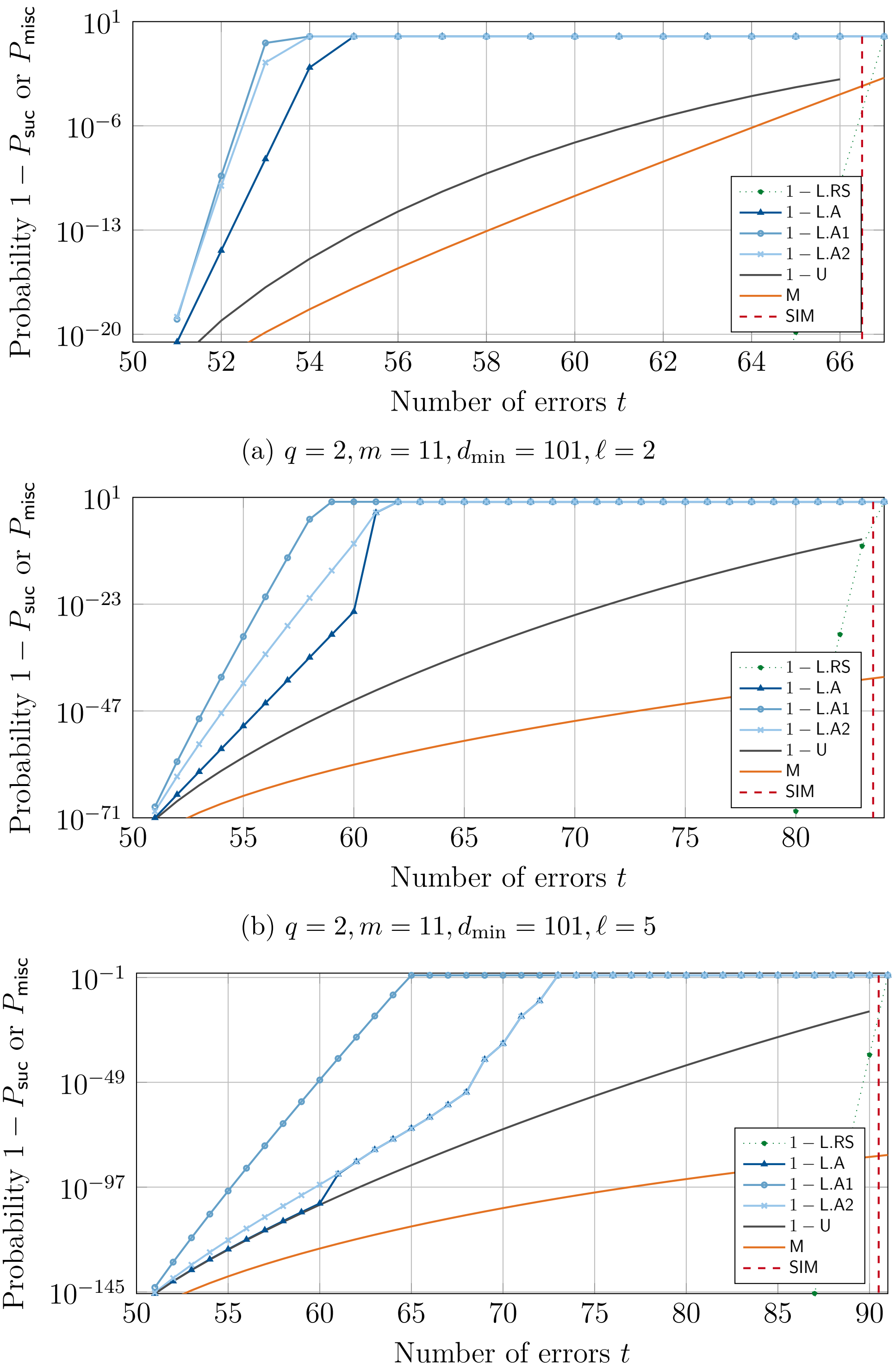


(a) $q = 2, m = 10, d_{\min} = 51, \ell = 2$

(b) $q = 2, m = 10, d_{\min} = 51, \ell = 5$

(c) $q = 2, m = 10, d_{\min} = 51, \ell = 10$

Figure 7.2: Comparison of the bounds for different parameters. For the bounds $\mathsf{L.RS}, \mathsf{L.A}, \mathsf{L.A1}, \mathsf{L.A2}$, and U on the success probability the respective probabilities of unsuccessful decoding $1 - P_{\mathsf{suc}}$ are shown.

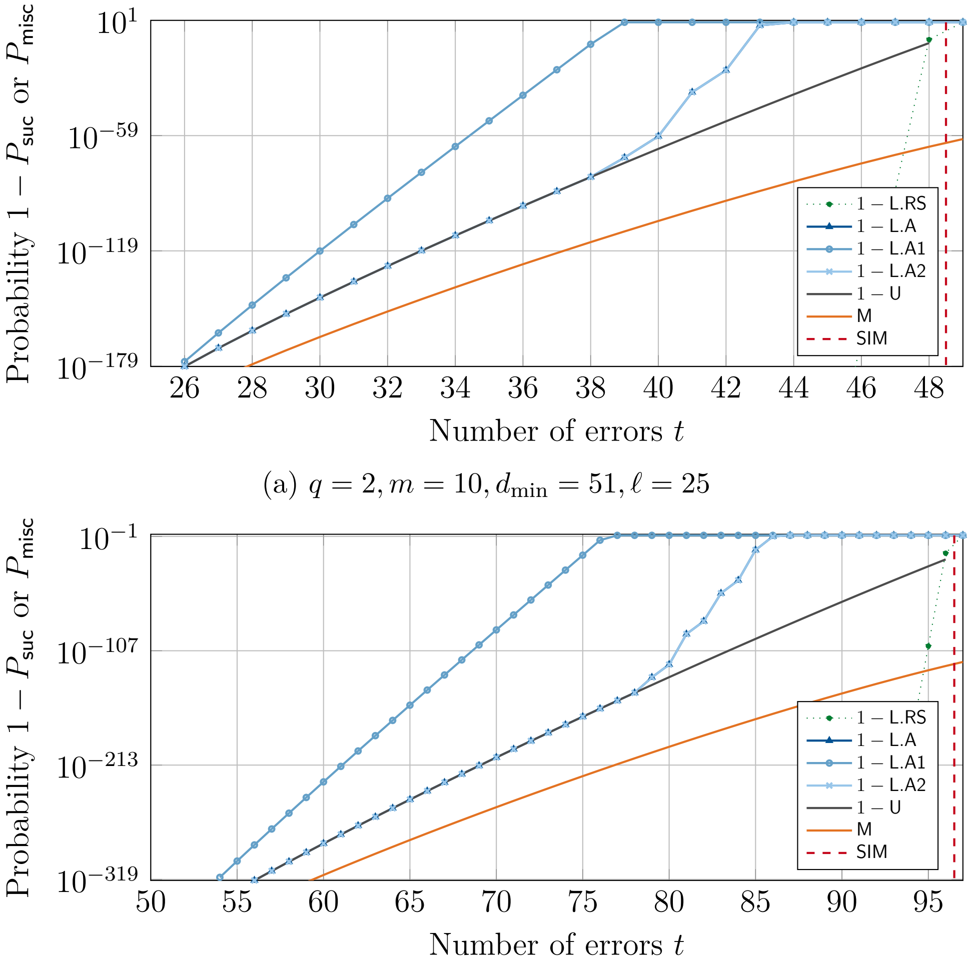


(a) $q = 2, m = 11, d_{\min} = 101, \ell = 2$

(b) $q = 2, m = 11, d_{\min} = 101, \ell = 5$

(c) $q = 2, m = 11, d_{\min} = 101, \ell = 10$

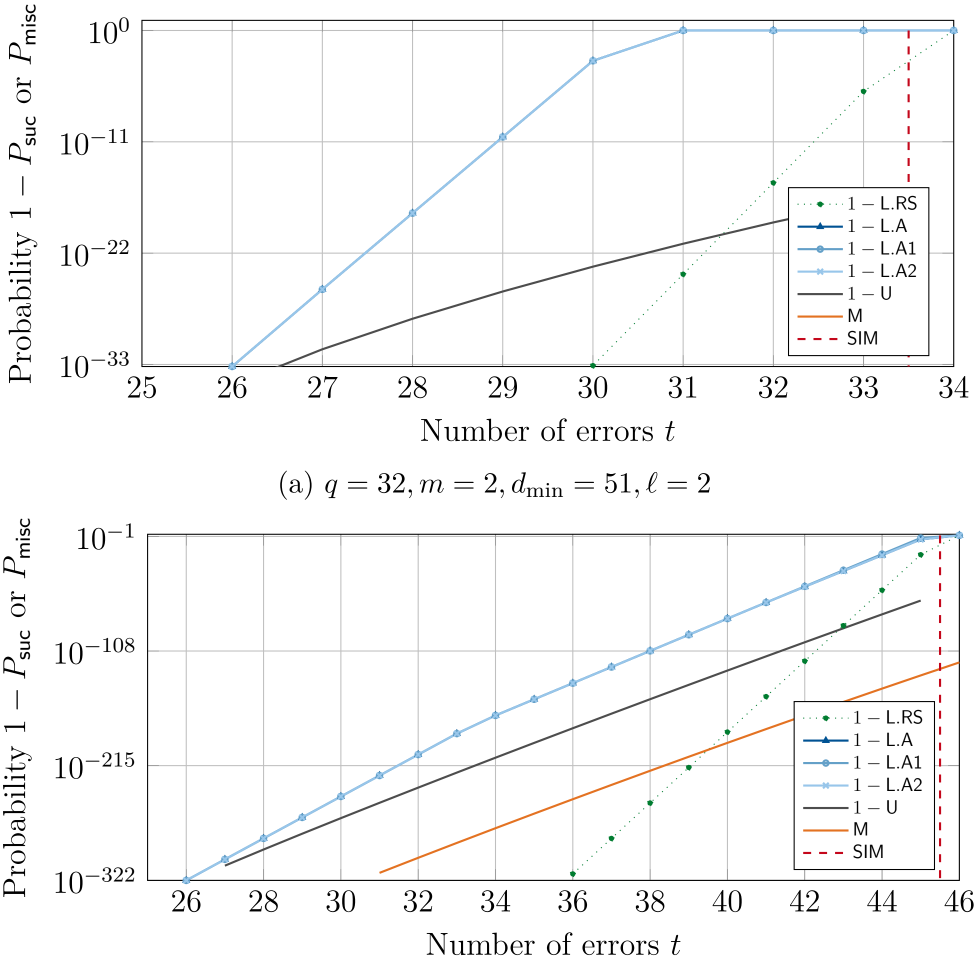
Figure 7.3: Comparison of the bounds for different parameters. For the bounds $\mathsf{L.RS}$, $\mathsf{L.A}$, $\mathsf{L.A1}$, $\mathsf{L.A2}$, and U on the success probability the respective probabilities of unsuccessful decoding $1 - P_{\mathsf{suc}}$ are shown.



(a) $q = 2, m = 10, d_{\min} = 51, \ell = 25$

(b) $q = 2, m = 11, d_{\min} = 101, \ell = 25$

Figure 7.4: Comparison of the bounds for different parameters. For the bounds $\mathsf{L.RS}, \mathsf{L.A}, \mathsf{L.A1}, \mathsf{L.A2}$, and U on the success probability the respective probabilities of unsuccessful decoding $1 - P_{\mathsf{suc}}$ are shown.



(a) $q = 32, m = 2, d_{\min} = 51, \ell = 2$

(b) $q = 32, m = 2, d_{\min} = 51, \ell = 10$

Figure 7.5: Comparison of the bounds for different parameters. For the bounds $\mathsf{L.RS}, \mathsf{L.A}, \mathsf{L.A1}, \mathsf{L.A2}$, and U on the success probability the respective probabilities of unsuccessful decoding $1 - P_{\mathsf{suc}}$ are shown.

- $\mathbf{q = 2, m = 10, d_{\min} = 51}$: For the parameters in Figs. 7.2a to 7.2c and 7.4a the code rate[10] is $R = \frac{k}{n} \approx 0.5$, assuming $k = n - (d_{\min} - 1)m$ (which holds for most alternant codes). For wild Goppa and BCH codes the rate is $R_{\mathsf{Gop./BCH}} \approx 0.75$ (see Remark 2.1).

- $\mathbf{q = 2, m = 11, d_{\min} = 101}$: For comparison to the parameters stated above, in Figs. 7.3a to 7.3c and 7.4b we fix the rate $R = \frac{k}{n} \approx 0.5$ ($R_{\mathsf{Gop./BCH}} \approx 0.75$), increase m, and vary $d_{\min}$ accordingly.

- $\mathbf{q = 32, m = 2, d_{\min} = 51}$: To illustrate the influence of the base field size q, in Figs. 7.5a and 7.5b we show some evaluations for $q = 32$.

We now briefly discuss the main observations taken from the numerical results. As L.A1 and L.A2 are simplifications of L.A and therefore strictly worse, we leave their comparison to each other until later in the section, and begin by only comparing L.RS, L.A, M, and U. All statements on the decoding failure, miscorrection, and success probability refer to the syndrome-based collaborative decoder of [FT91; SSB09b] given in Algorithm 1.

- For fixed q, m, and ℓ, the probability of a decoding success is significantly lower for interleaved (q-ary) alternant codes than for interleaved (q^m-ary) RS codes, as even the *upper* bound U on the success probability for interleaved alternant codes is in most cases smaller than the *lower* bound L.RS on the success probability for interleaved RS codes.

- The probability of unsuccessful decoding interleaved alternant codes $1 - P_{\mathsf{suc}}$ is dominated by the probability of failure P_{fail}, as $P_{\mathsf{misc}} \ll 1 - P_{\mathsf{suc}}$, i.e., the bound on the probability of a miscorrection P_{misc}, labeled M, is multiple orders of magnitude smaller than $1 - P_{\mathsf{suc}} = P_{\mathsf{misc}} + P_{\mathsf{fail}}$ for the best bound on P_{suc} given by L.A. This is consistent with the numerical results from [SSB09b] for the case of decoding interleaved RS codes.

- For higher interleaving order ℓ and relatively small number of errors t, the bound L.A essentially matches the upper bound of Theorem 7.4 (see Figs. 7.2c, 7.3c, 7.4a and 7.4b).

- For fixed q, m, and $d_{\min}$, the relative gap between the number of errors for which the lower bounds on the probability of decoding success become nontrivial, i.e., give $P_{\mathsf{suc}} > 0$, and the simulated decoding threshold decreases for increasing interleaving order ℓ (compare Figs. 7.2a to 7.2c and 7.4a or Figs. 7.3a to 7.3c and 7.4b).

[10] Recall that interleaving does not change the rate of the code.

Now consider the different versions of the bound in Theorem 7.3 labeled $\mathsf{L.A}, \mathsf{L.A1}$, and $\mathsf{L.A2}$.

- For small q, the performance of Theorem 7.3 is significantly worse when using a field size independent bound for k_q^{opt}, as evident from comparing $\mathsf{L.A}$ and $\mathsf{L.A1}$ in Fig. 7.2a to 7.4b. This can be expected due to the increasing gap between k_q^{opt} and the Singleton bound for decreasing q.
- For larger interleaving order ℓ, the simplified lower bound on the probability of successful decoding $\mathsf{L.A2}$ approaches the best version of the bound $\mathsf{L.A}$ (see Fig. 7.2c to 7.4b).

7.8 Summary and Open Problems

This chapter presented the first known lower and upper bounds for general parameters on the probability of successfully decoding interleaved alternant codes with the algorithm of [FT91; SSB09b]. The event of a decoding failure was shown to be the main cause of unsuccessful decoding, i.e., miscorrections are negligible in this sense. Numerical evaluations show that one of the provided lower bounds on this probability of successful decoding is tight for some parameters, as it matches the corresponding newly derived upper bound.

The most apparent open problem, in particular for smaller interleaving order, is closing the gap between the number of errors for which the bounds provide a nontrivial success probability and the simulated threshold for which the decoder succeeds. A closely related question, which is also of purely theoretical interest, is determining the distribution of the dimensions of all alternant codes for a given set of RS code locators. For specific applications, such as code-based cryptography, improvements of the bounds for other error distributions, arising, e.g., from an additional restriction to full-rank errors, could be of practical relevance. Finally, the simplification of the presented bound on the probability of decoding success, such that an analytical derivation of the maximal number of errors that result in a nontrivial bound is possible, as in the case of interleaved RS codes, is an interesting question to consider.

Part III

Private Information Retrieval

8

Towards the Capacity of PIR from Coded and Colluding Servers

Abstract

In this chapter, two practical concepts related to private information retrieval (PIR) are introduced and coined *full support-rank* PIR and *strongly linear* PIR. Being of full support-rank is a technical, yet natural condition required to prove a converse result for a capacity expression and satisfied by almost all currently known capacity-achieving schemes, while strong linearity is a practical requirement enabling implementation over small finite fields with low subpacketization degree.

The capacity of MDS-coded, linear, full support-rank PIR in the presence of colluding servers is derived, as well as the capacity of symmetric, linear PIR with colluding, adversarial, and nonresponsive servers for the recently introduced concept of matched randomness. This positively settles the capacity conjectures stated by Freij-Hollanti et al. and Tajeddine et al. in the presented cases. It is also shown that, further restricting to strongly linear PIR schemes with deterministic linear interference cancellation, the so-called star product scheme proposed by Freij-Hollanti et al. is essentially optimal and induces no capacity loss.

This chapter is based on the work [HFLH22] published in the IEEE Transactions on Information Theory*. In part, the results have been published in the proceedings of the* 2019 IEEE Information Theory Workshop (ITW) *[HFH19].*

8.1 Introduction

The previous chapters where concerned with the analysis, construction, and (efficient) decoding of different notions of linear codes. The application to DSSs has been the common motivation for these efforts, as linear codes can provide improved resilience against node failures with a relatively small storage overhead and interleaving is a natural consequence of the structure of these systems. With the amount of data stored in DSSs, this protection against the loss or corruption of data has become a topic of broad interest. However, assuring data integrity is not the only requirement of such systems and, in particular, systems that guarantee user privacy are in high demand.

As a result, PIR from storage systems has gained a lot of interest in recent years, where the goal of a user is to download a desired file without revealing the identity of the file to the servers. Specifically, this setting assumes a database consisting of a number of servers[1], which receives and answers queries of users for specific (functions of) files stored in the system. In the seminal work of [CGKS95] Chor et al. showed that information-theoretic privacy is not possible for systems comprised of a single server. This initiated the study of two main models of PIR. The first approach are systems consisting of a single server providing computational privacy, i.e., privacy under the assumption of an attacker with limited computational power. Several schemes have been proposed [KO97; Lip05; Lip05; AG07; YKPB12; GH19; ABFK16; ACLS18; KLL$^+$15; LP17], including one by the author of this work [HHW20] (see Section 9.2). However, these schemes incur a heavy computational load on the servers [SC07] and are prone to cryptanalytic attacks [LB16; BL21].

The main alternative to such single-server PIR schemes are schemes operating on systems consisting of multiple servers, which do not all communicate (collude) with each other. The advantage of this model is that it allows for schemes of low complexity and high rate, while providing perfect, information-theoretic privacy. Following the influential work of [SJ17], which derives the capacity of PIR from replicated storage, a multitude of different models emerged, such as PIR from MDS-coded storage [BU18], with colluding servers [SJ18b], with side information [HKS18; HGK$^+$18], and symmetric PIR (SPIR) [WS17b; WS17a; WS17c; SJ18c; WS19]. Here, *symmetric* refers to the property that the user is only able to decode the requested file and learns nothing about the other files.

In this chapter, we investigate the capacity of different notions of PIR from MDS-coded storage, i.e., the maximal achievable rate at which the private retrieval of a file from a database encoded with an MDS code is possible. Further, we consider symmetric privacy in the presence of adversaries and nonresponsive servers.

In the following, we denote nonsymmetric and symmetric PIR with t-collusion by TPIR and TSPIR. For the setting with additional b adversarial[2] (and possibly r nonreponsive) servers we write TBPIR and TBSPIR, respectively.

8.1.1 Known Results and Conjectures

For some settings the PIR capacity is known, e.g., for replicated storage without [SJ17] and with colluding servers [SJ18b], MDS-coded storage without collusion [BU18],

[1] We refer to the storage units in the system as servers instead of nodes in this chapter to imply the assumption that each of them can be contacted individually, which is not necessarily true for other kinds of nodes, such as individual hard drives within a rack.

[2] In PIR literature these adversarial servers are also referred to as *Byzantine* servers.

single-server PIR with side information [HKS18; HGK$^+$18], and SPIR [SJ18c; WS17b; WS17a; WS17c; WS19]. It has also been shown that the MDS property is not necessary for achieving the MDS–PIR capacity [FGH$^+$19; KLRA19].

For the reader's convenience, we summarize the known results relevant to this work in the following, along with some conjectures on the capacity of the open cases.

Nonrigorously, the rate of a PIR scheme with m files is denoted and defined (for a formal definition, see Definition 8.5) as

$$R_m = \frac{\text{size of the desired file}}{\text{size of the total download}} \ .$$

We denote by C_m the *capacity*, i.e., the largest achievable rate of a PIR scheme for m files under some given constraints. A collection of schemes defined for a varying number of files is said to be of asymptotic rate

$$R_\infty := \lim_{m\to\infty} R_m \ ,$$

and is called asymptotically capacity achieving if

$$R_\infty = \lim_{m\to\infty} C_m \ .$$

In a symmetric scheme, the servers need to share some amount of randomness to hide the undesired files from the user [GIKM00]. The amount of this randomness relative to the file size is referred to as the *secrecy rate* and denoted ρ (for a formal definition, see Definition 8.6).

Let us now assume $n > k+t+2b+r-1$, where n is the number of servers, k is the dimension of the storage code and t, b, r refer to the number of colluding, adversarial, and nonresponsive servers, respectively. For the remainder of this chapter we exclude the trivial case of a single file and assume $m \geq 2$.

Theorem 8.1 (Capacity of MDS-coded TSPIR [WS17a, Theorem 1] and uncoded TBSPIR [WS17b, Theorem 1])**.** *The capacity of MDS-TSPIR, i.e., symmetric PIR from $[n,k]$ MDS-coded storage with t colluding servers, is*

$$C^{\mathsf{MDS}}_{\mathsf{TSPIR}} = \begin{cases} 1-\frac{k+t-1}{n}, & \text{if } \rho^{\mathsf{MDS}}_{\mathsf{TSPIR}} \geq \frac{k+t-1}{n-k-t+1} \\ 0, & \text{otherwise} \end{cases} \ .$$

The capacity of uncoded TBSPIR, i.e., symmetric PIR from replicated storage with t colluding and b adversarial servers, is

$$C^{\mathsf{Rep}}_{\mathsf{TBSPIR}} = \begin{cases} 1-\frac{2b+t}{n}, & \text{if } \rho_{\mathsf{TBSPIR}} \geq \frac{t}{n-t-2b} \\ 0, & \text{otherwise} \end{cases} \ .$$

It is easy to check that when $t = 1$ or $k = 1$, the above SPIR capacity coincides with the asymptotic (in the number of files) capacity of PIR without server privacy [BU18; SJ18b]. The following conjectures describe the natural extension of this observation to more general settings.

Conjecture 8.1 (Asymptotic MDS-coded TBPIR [TGK+19, Conjecture 1])**.** *The asymptotic capacity of MDS-coded TBPIR, i.e., PIR from $[n,k]$ MDS-coded storage with t colluding, b adversarial, and r nonresponsive servers as the number of files $m \to \infty$, is*

$$C^{\mathsf{MDS}}_{\infty-\mathsf{TBPIR}} = 1 - \frac{k+t+2b+r-1}{n} \ .$$

Conjecture 8.2 (MDS-coded TBSPIR [TGK+19, Conjecture 2])**.** *The capacity of MDS-coded TBSPIR, i.e., symmetric PIR from $[n,k]$ MDS-coded storage with t colluding, b adversarial, and r nonresponsive servers, is*

$$C^{\mathsf{MDS}}_{\mathsf{TBSPIR}} = 1 - \frac{k+t+2b+r-1}{n} \ .$$

Remark 8.1. *In the original version of the above conjectures, the denominator is $n-r$ instead of n. This discrepancy is caused by our choice to include the nonresponsive servers in the calculation of the download cost. While it is reasonable to argue that nonresponsive servers do not incur any download and should therefore not be included in this cost, this depends on the particular system as, e.g., dropped packets on the side of the user could also cause a missing response, while clearly causing network traffic. Here, we also count the nonresponsive servers in the download cost, but point out that the results apply to both points of view.*

For the case of a finite number of files, the observation of the capacity expressions for the known cases of either $k=1$ or $t=1$ naturally leads to the following conjecture.

Conjecture 8.3 (MDS-coded TPIR [FGHK17, Conjecture 1])**.** *Let $\mathcal{C}$ be an $[n,k,d_{\min}]$ code with a generator matrix $\mathbf{G}$ that stores m files via the distributed storage system $\mathbf{Y} = \mathbf{X} \cdot \mathbf{G}$, and fix $1 \leq t \leq n-k$. Any PIR scheme for $\mathbf{Y}$ that protects against any t colluding servers has rate R_m at most*

$$R_m \leq \frac{1-\frac{k+t-1}{n}}{1-(\frac{k+t-1}{n})^m} \xrightarrow{m\to\infty} 1 - \frac{k+t-1}{n} \ .$$

Conjecture 8.3 in its full extent was disproved in [SJ18a], where the authors exhibited an explicit PIR scheme for $m=2$ files distributed over $n=4$ servers using a rate $1/2$ storage code, which protects against $t=2$ collusion. This scheme has rate $3/5$, while the conjectured capacity was $4/7$. However, as will be shown in the following, the

Table 8.1: Asymptotic capacity results and conjectures (in red). The maximum number of colluding, adversarial, and nonresponsive servers is denoted by t, b, r, respectively.

Restrictions	$[n,k]$ MDS-coded PIR	Reference
–	$1-\frac{k+t+2b+r-1}{n}$	[TGK+19]
$b=r=0$	$1-\frac{k+t-1}{n}$	[FGHK17]
$k=1, r=0$	$1-\frac{t+2b}{n}$	[BU19]
$t=1, b=r=0$	$1-\frac{k}{n}$	[BU18]
$k=1, b=r=0$	$1-\frac{t}{n}$	[SJ18b]

conjecture does hold for important subclasses of PIR schemes, which do not contain this counter-example.

In Table 8.1, we summarize the known asymptotic capacity results relevant to this work and show the conjectured results [FGHK17; TGK+19] in red.

8.1.2 Contributions and Outline

We begin this chapter by formalizing the considered model of a DSS and definitions of the different notions of PIR in Section 8.2.

The main contribution of this chapter, provided in Section 8.4, is the proof of the capacity of linear, full support-rank, MDS-coded PIR with colluding servers. Nonrigorously, linearity refers to the property that the responses are obtained as the inner product between the (encoded pieces of the) files stored at a node and a query vector. This appears to be a natural assumption as, to the best of our knowledge, *all* (asymptotically) capacity-achieving schemes are in fact linear [SJ17; FGHK17; SJ18a; SJ18b; BU18; TGK+19; FGH+19; KLRA19; BU19; DE19].

The seemingly technical assumption of full support-rank (see Definition 8.2) restricts the generality of the result, however, we demonstrate its practical relevance in two important regards. Firstly, the capacity achieving schemes for the special cases of $k=1$ (uncoded storage) or $t=1$ (no collusion) given in [SJ17; SJ18b; BU18; BU19] fulfill this definition. Second, the only scheme for general parameters achieving this newly proved capacity, introduced in [DE19], is also of full support-rank[3].

[3] We note that the necessary assumption was *not* made in the original paper [DE19], however, as we show in Section 8.4.2, it is in fact required to hold for the scheme to be private.

Further, and more importantly, the result provides insights towards the requirements for proving a general capacity expression. To better illustrate this, we take a high-level look at existing schemes: In the "simplest" approach, as utilized in [CGKS95; TGE18; TGK$^+$19; FGHK17], privacy is achieved through ensuring that each t-tuple of servers receives a set of vectors uniformly distributed over the respective linear vector space. The advantage of these schemes is that they achieve the respective asymptotic PIR capacity (for the cases where it is known, see Table 8.1), are relatively simple, and allow for small subpacketization. However, they fall short in achieving the capacity for a finite number of files.

The schemes able to achieve the capacity for a finite number of files [SJ17; SJ18b; BU18; BU19] are based on querying for specific, carefully chosen pieces of (encoded) files. In this case, the queries received by t-tuples are no longer uniformly distributed over *all* vectors since as, e.g., the all-zero vector will never be a query in these schemes. Furthermore, the equivalent of querying for specific pieces of files in a linear scheme is sending specific unit vectors as the queries. As querying for the same symbol multiple times is clearly suboptimal, these vectors are necessarily linearly independent. Similarly, the only general scheme achieving the newly derived capacity for the coded-colluding case $k, t > 1$, given in [DE19], is also based on constructing queries supported only on the positions corresponding to specific, carefully chosen files. As shown in Section 8.4.2, the natural choice to achieve privacy here, is requiring supported positions of the query to be linearly independent.

Our definition of full support-rank PIR (see Definition 8.2) captures this linear independence of the queries shared by these schemes. Thereby, the results we prove in the following show that in order to exceed the rate achieved by the scheme in [DE19], it is *necessary* for some restrictions of the queries to subsets of t servers to be *linearly dependent*. To further support this argument, we show in Section 8.4.2 that it is exactly this property that allows the scheme of [SJ18a], which is *not* of full support-rank, to exceed the (thereby disproved in full generality) conjectured capacity of [FGHK17, Conjecture 1].

In Section 8.5 we move on to the proof of the capacity of symmetric PIR from MDS-coded storage with colluding, adversarial, and nonresponsive servers. This confirms Conjecture 8.2 and shows that the symmetric PIR scheme of [TGK$^+$19] is capacity achieving.

Finally, in Section 8.6, we introduce a new notion of PIR schemes, which we coin *strongly linear*. Aside from the linearity of the query scheme, this class of schemes is characterized by the linearity and independence of their decoding functions, i.e., the (piece of) the desired file is obtained as a linear combination of the responses, which is independent of the query realization. While strong-linearity is a technical assumption, schemes of this class are of high practical relevance, as they allow for achieving the asymptotic capacity with small subpacketization. We show that all schemes in this class can be replaced by a star-product scheme as in [FGHK17] without loss in rate, thereby proving the capacity of these schemes.

8.2 Problem Setup

We now formalize the problem setup and recall some known results for different settings. For the purpose of this chapter we need to consider a generalized version of a DSS. In particular, as we derive information-theoretic expressions, we need to explicitly define the files, node storage, and related concepts as random variables. To this end, we first introduce some new notation, aimed to balance consistency in notation and readability of the presented results.

8.2.1 Notation

In the following we will define several random variables W, denoted by capital letters, whose realizations are (subsets of) a matrix space, i.e., $\mathrm{supp}(W) \subseteq \mathbb{F}^{m\times n}$. To establish the required technical results we also need the equivalent interpretation of these random variables as matrices of random variables over $\mathbb{F}$, i.e., matrices $\mathbf{W}$ where each $\mathbf{W}[i,j]$ is a random variable distributed over $\mathbb{F}$. For a set of random variables $W = \{W_1, W_2, \dots)$, we denote the corresponding matrix containing the W_j as a rows or columns by $\mathbf{W} = (\mathbf{W}_1^\top, \mathbf{W}_2^\top, \ldots)^\top$ or $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \ldots)$, where the applicable interpretation will be explicitly provided or clear from context. Semantically, the rows/columns of such a matrix $\mathbf{W}$ corresponding to each W_j belong together. To avoid double indexing, we restrict ourselves to only using one method of indexing, i.e., either super-/subscripts or square brackets, at a time. When necessary, we refer to such sets of rows/columns as *thick rows/columns* and to index them, we define a map from the indices of such thick rows/columns, to sets of normal rows/columns. For a set $\mathcal{I} \subseteq [n]$ define

$$\psi_\beta(\mathcal{I}) = \bigcup_{i\in\mathcal{I}} \{(i-1)\beta + 1, \dots, i\beta\} \ . \tag{8.1}$$

Then, for an $m \times n\beta$ matrix $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_n)$, where each $\mathbf{W}_j$ is an $m \times \beta$ matrix, the restriction $\mathbf{W}[:, \psi_\beta(\mathcal{I})]$ indexes the $|\mathcal{I}|$ *β-thick columns* given by $\mathcal{I}$, where a thick column is a submatrix consisting of β consecutive columns of $\mathbf{W}$. Note that this is equivalent to the set of random variables $W_{\mathcal{I}}$. The same notation is used to index thick rows.

8.2.2 Problem Setup

Consider a distributed storage system consisting of n servers storing m files $X = \{X^1, X^2, \dots, X^m\}$, where each X^l is a random variable uniformly distributed over $\mathbb{F}^{\alpha\times k}$. Interpreted as a matrix, the *data matrix* is denoted by $\mathbf{X} \in \mathbb{F}^{\alpha m\times k}$, where each block of α consecutive rows corresponds to a file. This matrix is encoded with an $[n,k]$ MDS storage code and server j stores the j-th thick column (see Section 2.1) of

$$\mathbf{Y} = \mathbf{X} \cdot \mathbf{G} = \begin{pmatrix} X^1 \\ X^2 \\ \vdots \\ X^m \end{pmatrix} \cdot G = \begin{pmatrix} Y_1^1 & Y_2^1 & \cdots & Y_n^1 \\ Y_1^2 & Y_2^2 & \cdots & Y_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ Y_1^m & Y_2^m & \cdots & Y_n^m \end{pmatrix} \in \mathbb{F}_q^{\alpha m \times n} ,$$

where $\mathbf{G}$ is a generator matrix of the storage code. Note that $Y^l = X^l \mathbf{G}$ is the encoded version of the l-th file and Y_j^l is a random variable taking realizations over $\mathbb{F}_q^\alpha$ representing the symbols corresponding to this file stored at node j.

For $l \in [m]$, $j \in [n]$, and $\mathcal{I} \subset [n]$, the random variables Y^l, Y_j, and $Y_\mathcal{I}$ taking realizations over a distribution of matrices can equivalently be described as matrices of random variables taking realizations over $\mathbb{F}_q$, given by $\mathbf{Y}[\psi_\alpha(l), :]$ (the l-th thick row of $\mathbf{Y}$), $\mathbf{Y}[:, j]$ (the j-th column of $\mathbf{Y}$), and $\mathbf{Y}[:, \mathcal{I}]$ (the restriction of $\mathbf{Y}$ to the columns indexed by $\mathcal{I}$), respectively. Here, α gives the number of *stripes* of each file and we note that each stripe is encoded independently of other stripes. The m files are independent and each consists of k i.i.d. randomly drawn symbols from $\mathbb{F}_q^\alpha$. Hence, for the entropies[4] it holds that

$$\begin{aligned} H(X^l) &= k\alpha \log q, \ \forall \ l \in [m] \\ H(X^1, \dots, X^m) &= mk\alpha \log q \ . \end{aligned}$$

We consider MDS codes, so every subset of at least k servers recover all files, i.e., for any set $\mathcal{W} \subset [n]$ with $|\mathcal{W}| \geq k$ it holds that

$$\begin{aligned} H(Y_\mathcal{W}) = H(X^1, \dots, X^m) &= mk\alpha \log q \\ H(X^1, \dots, X^m \mid Y_\mathcal{W}) &= 0 \ . \end{aligned}$$

Further, we assume that the servers have access to a shared source of randomness, which has been shown [GIKM00] (see also [SJ18c, Footnote 2]) to be required for enforcing the property of symmetry, i.e., ensuring that the user learns nothing about the files other than the requested file. Formally, let $\mathcal{S}$ be a vector space over $\mathbb{F}_q$, and let

$$S = (S_1, \dots, S_n)$$

be a random variable with $\text{supp}(S) \subseteq \mathcal{S}^n$, where the symbols of S_j may be used by the j-th server.

[4] Note that the base of the logarithm here is irrelevant, as it only reflects the unit (bits, nats, ...) in which the entropy is measured. As we are interested in rates in this chapter, i.e., *relative* relations between entropies, the results are independent of this choice.

In general, a PIR scheme consists of a user desiring the file with index i, who picks the corresponding query

$$Q^i = \left(Q_1^i, \dots, Q_n^i\right)$$

from the set of all possible queries $\mathcal{Q}$, and sends Q_j^i to the j-th server. Every server returns a response A_j^i that is a β-tuple of symbols in $\mathbb{F}_q$. For an honest (nonadversarial) server, this response depends on the query Q_j^i, the symbols Y_j stored at server j, and the randomness S shared by the servers, in a way known to the user. The list of responses from all servers for a given query is denoted by

$$A^i = \left(A_1^i, \dots, A_n^i\right) \ .$$

The desired file X^i should now be recoverable from the responses, meaning that

$$H(X^i \mid Q^i, A^i, \mathcal{Q}, i) = 0 \ . \tag{8.2}$$

In this work we only consider PIR schemes in which the query functions, i.e., the function that each server applies to obtain its response A_j^i, are linear.

Definition 8.1 (Linear PIR)**.** *A PIR scheme is said to be* linear *if*

- *the query Q^i can be represented as a matrix $\mathbf{Q}^i \in F^{\alpha m \times \beta n}$, where each β-thick column $\mathbf{Q}^i[:, \psi_\beta(j)]$ corresponds to the query Q_j^i to server $j \in [n]$, and*
- *the responses A_j^i of server $j \in [n]$ are given by the vector*

$$\begin{aligned}\mathbf{A}^i[:, \psi_\beta(j)] &= \left(\left\langle \mathbf{Y}[:, j], \mathbf{Q}^i[:, (j-1)\beta + s]\right\rangle + \mathbf{S}[:, (j-1)\beta + s]\right)_{s \in [\beta]} \\ &= \mathbf{Y}[:, j]^\top \cdot \mathbf{Q}^i[:, \psi_\beta(j)] + \mathbf{S}[:, \psi_\beta(j)] \ ,\end{aligned}$$

where the vector $\mathbf{S} \in \mathbb{F}^{1 \times \beta n}$ depends on the j-th share S_j of the randomness S shared by the servers.

Briefly and nonrigorously, in a linear PIR scheme each server receives β query vectors and responds with the β inner products between these vectors and the column of $\mathbf{Y}$ that it stores (possibly plus an additional symbol given by the shared randomness). In the case of nonsymmetric PIR, the servers do not need any shared randomness and we may assume that $\mathbf{S} = \mathbf{0}$.

It is customary to think of the β coordinates of the queries Q_j^i as *iterations.* In this terminology, a linear PIR scheme consists of β iterations, where in iteration s the user sends for each $j \in [n]$ the query vector

$$\mathbf{Q}^i[:, (j-1)\beta + s] \quad \in \mathbb{F}^{\alpha m \times 1}$$

and receives a response row vector

$$\Big(\mathbf{A}[:,(j-1)\beta+s]\Big)_{j\in[n]} \quad \in \mathbb{F}^{1\times n}.$$

It is easy to see that it is suboptimal to send linearly dependent queries to servers. However, in general, submatrices of the query matrix may indeed be nontrivially linearly dependent (see [SJ18a] and Section 8.4.2), i.e., have supported columns that are linearly dependent. The technical assumption we make in the following, given below in Definition 8.2, restricts all supported columns of the query for a subset of less than or equal to t servers to be linearly independent, even when restricting to an arbitrary subset of files. We therefore coin these schemes as full support-rank PIR schemes.

Definition 8.2 (Full support-rank PIR)**.** *A linear PIR scheme is said to be of* full support-rank *if for every query realization* $\mathbf{q} \in \operatorname{supp}(\mathbf{Q}^i) \subset \mathbb{F}^{\alpha m\times\beta n}$*, any subset* $\mathcal{T} \subseteq [n]$ *of* $|\mathcal{T}| \leq t$ *servers, and any file index* $j \in [m]$ *it holds that*

$$\operatorname{rank}(\mathbf{q}[\psi_\alpha(j),\psi_\beta(\mathcal{T})]) = |\operatorname{colsupp}(\mathbf{q}[\psi_\alpha(j),\psi_\beta(\mathcal{T})])|.$$

Most PIR schemes in the literature are indeed of full support-rank, including those in [SJ17; SJ18c; SJ18b; BU18; BU19; DE19; ZTSL20; LKH20; TSC19; ZYQT19]. A notable example of a scheme that is not of full support-rank is the counter-example to Conjecture 8.3 given in [SJ18a]. For a more detailed discussion regarding the applicability of Definition 8.2 to existing schemes, see Section 8.4.2.

In general, the goal of information-theoretic private information retrieval with t-collusion is for the user to retrieve a file such that any set of t storage servers learns nothing about the index of the desired file. This is referred to as *user privacy.*

Definition 8.3 (User Privacy with t-Collusion)**.** *Any t colluding servers shall not be able to obtain any information about the index of the requested file, i.e., the mutual information*

$$I(i; Q^i_{\mathcal{T}}, A^i_{\mathcal{T}}, Y_{\mathcal{T}}, S) = 0, \quad \forall\ \mathcal{T} \subset [n], |\mathcal{T}| = t\ . \tag{8.3}$$

We also consider symmetric PIR (SPIR), where the user is not supposed to learn any information about the files other than the requested one.

Definition 8.4 (Server Privacy)**.** *The user shall learn no information about files other than the requested one, i.e.,*

$$I(X^{[m]\setminus\{i\}}; Q^i, A^i, \mathcal{Q}, i) = 0\ . \tag{8.4}$$

A scheme that satisfies Eq. (8.2) and Eq. (8.3) is called a PIR scheme. If the scheme in addition satisfies Definition 8.4, then it is called an SPIR scheme. We are interested in the capacities of linear PIR and SPIR with collusion and adversaries (denoted with a prefix T and/or B, respectively), i.e., the highest achievable rate at which a desired file can be retrieved under these constraints.

Definition 8.5 ((S)PIR Rate and Capacity)**.** *The* rate *of an (S)PIR scheme is the number of information bits of the requested file retrieved per downloaded answer bits, i.e.,*

$$R_{\mathsf{(S)PIR}} = \frac{H(X^i)}{\sum_{j=1}^{n} H(A_j^i)} \ .$$

In order to achieve symmetric privacy, the servers require some amount of shared randomness [GIKM00].

Definition 8.6 (Secrecy Rate)**.** *The secrecy rate is the amount of common randomness shared by the storage servers relative to the file size, i.e.,*

$$\rho_{\mathsf{SPIR}} = \frac{H(S)}{H(X^i)} \ .$$

8.3 Preliminary Lemmas

We begin the technical part of this chapter by introducing some intermediate notions and lemmas which will be required in both Sections 8.4 and 8.5.

We will repeatedly use Han's inequality for joint entropies [CT91], which we state here for completeness. Let $W = \{W_1, \dots W_n\}$ be a set of random variables defined on the same probability space. Denote by $\binom{[n]}{k}$ the set of all subsets of $[n]$ with cardinality k. Then

$$\frac{k}{n} H(W_1, \dots W_n) \leq \frac{1}{\binom{n}{k}} \sum_{\mathcal{T} \in \binom{[n]}{k}} H(W_{\mathcal{T}}). \tag{8.5}$$

Our proofs of linear, full support-rank MDS-TPIR in Section 8.4 and MDS-TBSPIR capacity in Section 8.5 are partly based on the proofs of TBSPIR capacity in a replicated setting [WS17b] as well as the proofs of SPIR capacity [WS17c] and TSPIR capacity [WS17a] from MDS-coded storage. We first prove the intermediate results for a set of servers that is free of adversaries and then, similar to [WS17b], argue that the entropy of the adversarial responses has to be the same as for honest (nonadversarial) servers to obtain the capacity. For completeness, the proofs of the intermediate steps are included, though some of the proofs can be taken, with only minor adaptations, from [WS17a] and [WS17b].

Similar to the replicated case in [WS17b, Lemma 6], in the following we argue that when considering zero error probability, i.e., guaranteeing that the user can decode if the number of corrupted answers is less than or equal to b and the number of nonresponsive servers is less than or equal to r, every realization of $n-2b-r$ authentic answers has to be unique.

Lemma 8.1. *In an optimal scheme with zero error probability for b adversarial and r nonresponsive servers it holds that*

$$H(X^i \mid A^i_{\mathcal{H}}, \mathcal{Q}) = 0 \ ,$$

for any set $\mathcal{H} \subseteq [n]$ of honest servers with $|\mathcal{H}| \geq n-2b-r$.

Proof. The proof is similar to the replicated case of [WS17b, Lemma 6] and included for completeness. We show that the response of any $n-2b-r$ honest servers must suffice to correctly recover the desired file by proving that the corresponding responses must be unique for any realization of file i. Denote by $A^i_j(X^i = \mathbf{x}^i)$ the *honest* response of the j-th server for the realization $X^i = \mathbf{x}^i \in \mathbb{F}_q^{\alpha \times k}$ of the i-th file. For a contradiction, assume that for a set $\mathcal{R} \subset [n]$ with $|\mathcal{R}| = r$ of nonresponsive servers and a set $\mathcal{H} \subset [n] \setminus \mathcal{R}$ of honest servers with $|\mathcal{H}| = n-2b-r$ it holds that $A^i_{\mathcal{H}}(X^i = \mathbf{x}^i) = A^i_{\mathcal{H}}(X^i = \tilde{\mathbf{x}}^i)$ for two different realizations $\mathbf{x}^i \neq \tilde{\mathbf{x}}^i$ of file i. Partition the $2b$ remaining servers $\mathcal{B} = [n] \setminus (\mathcal{H} \cup \mathcal{R})$ into two subsets $\mathcal{B}_1$ and $\mathcal{B}_2$, each of size b, and denote their responses by $A^i_{\mathcal{B}_1}$ and $A^i_{\mathcal{B}_2}$, respectively. Now consider the following cases:

- The realization of file i is $X^i = \mathbf{x}^i$. The servers of $\mathcal{B}_1$ are *adversarial* and reply with $A^i_{\mathcal{B}_1}(X^i = \tilde{\mathbf{x}}^i)$. The servers of $\mathcal{B}_2$ are *honest*, i.e., they reply with $A^i_{\mathcal{B}_2}(X^i = \mathbf{x}^i)$.

- The realization of file i is $X^i = \tilde{\mathbf{x}}^i$. The servers of $\mathcal{B}_1$ are *honest*, i.e., they reply with $A^i_{\mathcal{B}_1}(X^i = \tilde{\mathbf{x}}^i)$. The servers of $\mathcal{B}_2$ are *adversarial* and reply with $A^i_{\mathcal{B}_2}(X^i = \mathbf{x}^i)$.

As $A^i_{\mathcal{H}}(X^i = \mathbf{x}^i) = A^i_{\mathcal{H}}(X^i = \tilde{\mathbf{x}}^i)$ by assumption, the user receives exactly the same responses from the servers in both cases and is therefore not able to differentiate between the two realizations. Hence unique decoding would fail, thereby violating the zero error probability requirement. Note that, as we require *zero* decoding error probability it is not necessary for the adversarial servers to know the index i. Instead, in each case it suffices that the probability of the adversarial servers replying with the respective responses is nonzero. We conclude that for any two different realizations $\mathbf{x}^i \neq \tilde{\mathbf{x}}^i$ of file i we have $A^i_{\mathcal{H}}(X^i = \mathbf{x}^i) \neq A^i_{\mathcal{H}}(X^i = \tilde{\mathbf{x}}^i)$, and the statement of the lemma follows. □

The following basic lemma will also be required in multiple proofs and applies to both the symmetric and nonsymmetric setting.

Lemma 8.2. *For any set $\mathcal{N} \subset [n]$ of honest (nonadversarial) servers*

$$H(A^i_{\mathcal{N}} \mid \mathcal{Q}, X^i, Q^i_{\mathcal{N}}) = H(A^i_{\mathcal{N}} \mid X^i, Q^i_{\mathcal{N}}) \ .$$

Proof. We first show that $I(A^i_{\mathcal{N}}; \mathcal{Q} \mid X^i, Q^i_{\mathcal{N}}) \leq 0$, as follows

$$\begin{aligned} I(A^i_{\mathcal{N}}; \mathcal{Q} \mid X^i, Q^i_{\mathcal{N}}) &\leq I(A^i_{\mathcal{N}}, X^{[m]}, S; \mathcal{Q} \mid X^i, Q^i_{\mathcal{N}}) \\ &\overset{(\mathsf{a})}{=} I(X^{[m]}, S; \mathcal{Q} \mid X^i, Q^i_{\mathcal{N}}) \\ &= H(X^{[m]}, S \mid X^i, Q^i_{\mathcal{N}}) - H(X^{[m]}, S \mid X^i, Q^i_{\mathcal{N}}, \mathcal{Q}) \\ &\overset{(\mathsf{b})}{=} H(X^{[m]}, S \mid X^i) - H(X^{[m]}, S \mid X^i) = 0 \end{aligned}$$

where (a) follows because the answers $A^i_{\mathcal{N}}$ are a function of the queries $Q^i_{\mathcal{N}}$, the files $X^{[m]}$, and the shared randomness S (for the nonsymmetric case S can be thought of as a constant, e.g., $S = \mathbf{0}$), and (b) holds because the files $X^{[m]}$ and shared randomness S are independent of the queries. As mutual information is nonnegative, it follows that

$$\begin{aligned} I(A^i_{\mathcal{N}}; \mathcal{Q} \mid X^i, Q^i_{\mathcal{N}}) &= H(A^i_{\mathcal{N}} \mid X^i, Q^i_{\mathcal{N}}) - H(A^i_{\mathcal{N}} \mid \mathcal{Q}, X^i, Q^i_{\mathcal{N}}) = 0 \\ &\Rightarrow H(A^i_{\mathcal{N}} \mid X^i, Q^i_{\mathcal{N}}) = H(A^i_{\mathcal{N}} \mid \mathcal{Q}, X^i, Q^i_{\mathcal{N}}) \ . \end{aligned}$$

□

8.4 The Capacity of Linear, Full Support-Rank MDS-TPIR

In this section we prove the capacity of linear, full support-rank PIR from MDS-coded storage with collusion.

As we are only concerned with nonsymmetric PIR here, we assume $\mathbf{S} = \mathbf{0}$ for the remainder of this section.

8.4.1 Converse

A novel formulation of the key Lemma 8.7, which is slightly stronger than the corresponding lemmas in [WS17c; WS17a], allows us to induct over the number of files, without requiring the symmetry assumption. We then use this induction result to prove the MDS-TPIR capacity for linear, full support-rank schemes. The same proof also yields an upper bound for the capacity in the presence of adversarial servers. However, the upper bound for MDS-TBPIR does not correspond to any known scheme constructions, and does not agree with the MDS-TBSPIR capacity asymptotically as the number of files grows to infinity.

The main technical difficulty in the derivation of this result is captured in Lemma 8.6, which describes how sets of as many as $k + t - 1$ servers will give responses that are independent of the index of the desired file, even when conditioned on an arbitrary subset of files. In order to show this we need some additional technical results on the rank of the Khatri-Rao product [KR68] of certain matrices (see Section 2.1.1).

For completeness, we note that the following results also hold if the thick columns are not all of the same size β, but instead each consist of a (possibly) different number of columns. However, to not complicate the notation even further, we restrict ourselves to PIR schemes that query each node exactly β times, which corresponds to equal sized thick columns, each consisting of β columns.

Lemma 8.3. *Let $\mathcal{C}$ be a $[k+t-1,k]$ MDS code with generator matrix $\mathbf{G} \in \mathbb{F}_q^{k\times(k+t-1)}$ and $\mathbf{q} \in \mathbb{F}_q^{\alpha\times\beta(k+t-1)}$ be a matrix such that for any set $\mathcal{T} \subset [k+t-1]$ with $|\mathcal{T}| = t$ we have*

$$\operatorname{rank}(\mathbf{q}[:,\psi_\beta(\mathcal{T})]) = |\operatorname{colsupp}(\mathbf{q}[:,\psi_\beta(\mathcal{T})])| \ . \tag{8.6}$$

Then

$$\operatorname{rank}((\mathbf{G}\otimes \mathbf{1}_\beta)\odot \mathbf{q}) = |\operatorname{colsupp}(\mathbf{q})| \ .$$

Proof. By a similar argument as in [Ran13, Proof of Lemma 6], we determine the rank of this matrix by proving that the unit vectors $\mathbf{e}_l \in \mathbb{F}^{(k+t-1)\beta}, l \in \operatorname{colsupp}(\mathbf{q})$ are contained in the row span of the matrix $(\mathbf{G}\otimes \mathbf{1}_\beta)\odot \mathbf{q}$, where $\mathbf{1}_\beta$ denotes the all-one vector of length β. First observe that for any full-rank matrices $\mathbf{P}_1 \in \mathbb{F}^{k\times k}$ and $\mathbf{P}_2 \in \mathbb{F}^{\alpha\times\alpha}$ we have

$$\begin{aligned}
\operatorname{rank}\Big((\mathbf{G}\otimes \mathbf{1}_\beta)\odot \mathbf{q}\Big) &= \dim\Big(\langle(\mathbf{G}\otimes \mathbf{1}_\beta)\odot \mathbf{q}\rangle_{\mathsf{row}}\Big)\\
&\overset{(\mathsf{a})}{=} \dim\Big(\langle\mathbf{G}\otimes \mathbf{1}_\beta\rangle_{\mathsf{row}}\odot\langle\mathbf{q}\rangle_{\mathsf{row}}\Big)\\
&\overset{(\mathsf{b})}{=} \dim\Big(\langle\mathbf{P}_1\cdot(\mathbf{G}\otimes \mathbf{1}_\beta)\rangle_{\mathsf{row}}\odot\langle\mathbf{P}_2\cdot\mathbf{q}\rangle_{\mathsf{row}}\Big)\\
&\overset{(\mathsf{c})}{=} \dim\Big(\langle(\mathbf{P}_1\cdot\mathbf{G})\otimes \mathbf{1}_\beta\rangle_{\mathsf{row}}\odot\langle\mathbf{P}_2\cdot\mathbf{q}\rangle_{\mathsf{row}}\Big)\\
&= \operatorname{rank}\Big(((\mathbf{P}_1\cdot\mathbf{G})\otimes \mathbf{1}_\beta)\odot(\mathbf{P}_2\cdot\mathbf{q})\Big)\ ,
\end{aligned} \tag{8.7}$$

where (a) follows from Eq. (2.3), (b) holds because the left-multiplication by a full-rank matrix does not change the row space, and (c) holds by Eq. (2.2). To obtain the unit vectors $\mathbf{e}_l \in \mathbb{F}^{(k+t-1)\beta}, l \in \psi_\beta(1)\cap\operatorname{colsupp}(\mathbf{q})$, choose $\mathbf{P}_1$ such that $\mathbf{P}_1\mathbf{G}$ is in systematic form, i.e., its first k columns are an identity matrix. This is always possible, since $\mathbf{G}$ is the generator matrix of an MDS code. Now consider the set $\mathcal{T} = \{1, k+1, \ldots, k+t-1\}$ and choose $\mathbf{P}_2$ such that the submatrix consisting of the t columns indexed by $\psi_\beta(\mathcal{T})$, i.e., the β-thick columns $\mathcal{T}$, contain the unit vectors $\mathbf{e}_l \in \mathbb{F}^{t\beta}, l \in \operatorname{colsupp}(\mathbf{q}[:,\psi_\beta(\mathcal{T})])$ as rows. Eq. (8.6) guarantees that such a matrix exists. Fig. 8.1 illustrates these matrices and the transformation step.

Now the first row of the matrix $(\mathbf{P}_1\cdot\mathbf{G})\otimes\mathbf{1}_\beta$ is a vector that is only (and exactly,

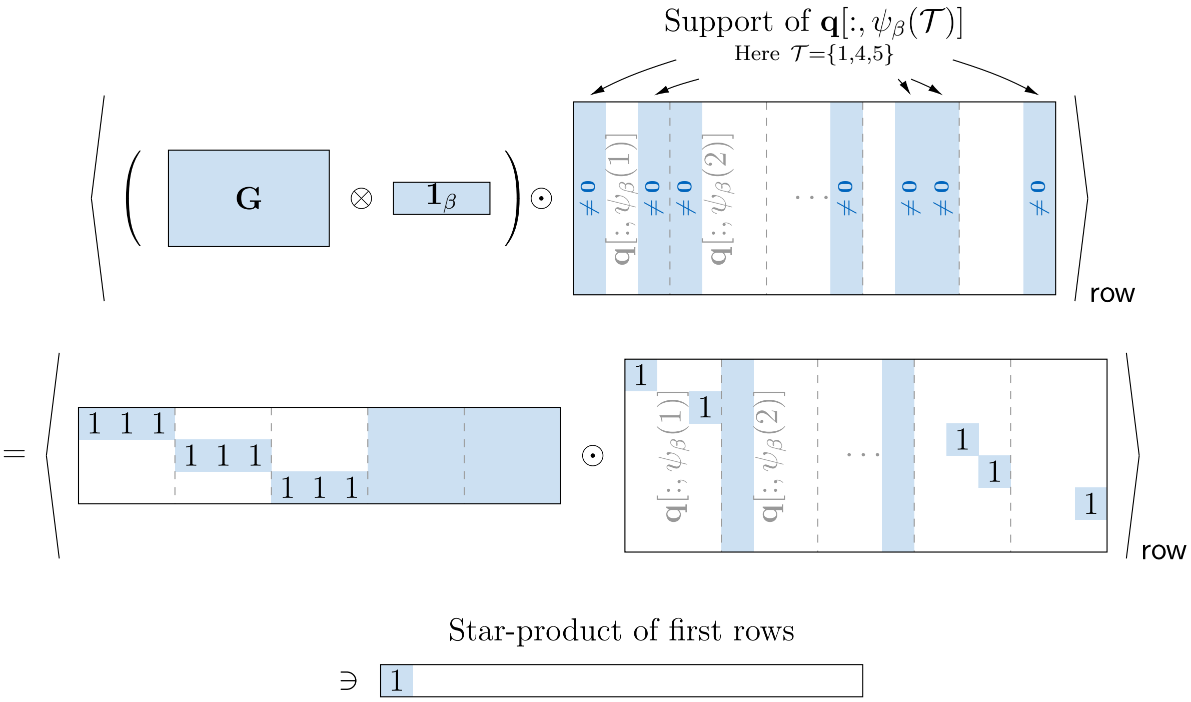

Support of $\mathbf{q}[:, \psi_\beta(\mathcal{T})]$
Here $\mathcal{T}$={1,4,5}
$\mathbf{G}$
$\otimes$
$\mathbf{1}_\beta$
$\odot$
$\neq \mathbf{0}$
$\mathbf{q}[:, \psi_\beta(1)]$
$\mathbf{q}[:, \psi_\beta(2)]$
...
row
=
1 1 1
1 1 1
1 1 1
1
1
1
1
1
row
Star-product of first rows
$\ni$
1

Figure 8.1: Illustration of the proof of Lemma 8.3 for $k = t = \beta = 3$. The blue areas indicate positions that are potentially nonzero, white areas contain only zeros. Columns in the support of $\mathbf{q}$, i.e., nonzero columns of $\mathbf{q}$, are indicated by $\neq \mathbf{0}$. The second line corresponds to Eq. (8.7) when $\mathbf{P}_1$ and $\mathbf{P}_2$ are chosen as described in the proof for $\mathcal{T} = \{1, 4, 5\}$. The third line is the first unit vector, given by the star-product of the first rows of the two matrices.

as the code is MDS) supported on the positions $\psi_\beta(\mathcal{T})$. Further, by the choice of $\mathbf{P}_2$, for any $l \in \mathrm{colsupp}(\mathbf{q})$, there exists a row in the matrix $\mathbf{P}_2 \cdot \mathbf{q}$ of support $\mathcal{S} \subset \{l\} \cup \psi_\beta(\{2, 3, \dots, k\})$ and $l \in \mathcal{S}$. Hence, the star-product of these rows, which, by definition of the column-wise Khatri-Rao product, is a row of $(\mathbf{P}_1 \cdot (\mathbf{G} \otimes \mathbf{1}_\beta)) \odot (\mathbf{P}_2 \cdot \mathbf{q})$, is the l-th unit vector.

By the same approach we can show that all the unit vectors $\mathbf{e}_l \in \mathbb{F}^{(k+t-1)\beta}, l \in \mathrm{colsupp}(\mathbf{q})$ are contained in the row span[5] of $(\mathbf{G} \otimes \mathbf{1}_\beta) \odot \mathbf{q}$ and the lemma statement follows. □

With this technical lemma established, we now link the entropy of the answers of any subset of $k + t - 1$ servers to the column support of the query.

Lemma 8.4. *Let $\mathcal{C}$ be an $[n, k]$ MDS code with generator matrix $\mathbf{G} \in \mathbb{F}_q^{k \times n}$ and $\mathbf{Y} = \mathbf{X} \cdot \mathbf{G} \in \mathbb{F}_q^{\alpha m \times n}$, where $\mathbf{X}$ is chosen uniformly at random from all $\mathbb{F}^{\alpha m \times k}$ matrices. Further, let $\mathbf{q} \in \mathbb{F}^{\alpha m \times \beta n}$ be a matrix such that for any set $\mathcal{T} \subset [n]$ with $|\mathcal{T}| = t$ and nonempty set $\mathcal{F} \subset [m]$ we have*

$$\mathrm{rank}(\mathbf{q}[\psi_\alpha(\mathcal{F}), \psi_\beta(\mathcal{T})]) = |\,\mathrm{colsupp}(\mathbf{q}[\psi_\alpha(\mathcal{F}), \psi_\beta(\mathcal{T})])|\ .$$

Then for any set $\mathcal{N} \subset [n]$ with $|\mathcal{N}| = k + t - 1$ it holds that

$$H\Big(\sum_{l \in \psi_\alpha(\mathcal{F})} (\mathbf{Y}[l, \mathcal{N}] \otimes \mathbf{1}_\beta) \star \mathbf{q}[\psi_\alpha(\mathcal{F}), \psi_\beta(\mathcal{N})] \Big) = |\,\mathrm{colsupp}(\mathbf{q}[\psi_\alpha(\mathcal{F}), \psi_\beta(\mathcal{N})])|\ .$$

Proof. Let $\mathbf{I}_m$ denote the $m \times m$ identity matrix. We begin with some transformation steps:

$$\begin{aligned} &\sum_{l \in \psi_\alpha(\mathcal{F})} (\mathbf{Y}[l, \mathcal{N}] \otimes \mathbf{1}_\beta) \star \mathbf{q}[\psi_\alpha(\mathcal{F}), \psi_\beta(\mathcal{N})] \\ &\quad= \mathbf{1}_{|\psi_\alpha(\mathcal{F})|} \cdot \Big(((\mathbf{X}[\psi_\alpha(\mathcal{F}), :] \cdot \mathbf{G}|_{\mathcal{N}}) \otimes \mathbf{1}_\beta) \star (\mathbf{I}_{|\psi_\alpha(\mathcal{F})|} \cdot \mathbf{q}[\psi_\alpha(\mathcal{F}), \psi_\beta(\mathcal{N})]) \Big) \\ &\quad\overset{(2.2)}{=} \mathbf{1}_{|\psi_\alpha(\mathcal{F})|} \cdot \Big((\mathbf{X}[\psi_\alpha(\mathcal{F}), :] \cdot (\mathbf{G}|_{\mathcal{N}} \otimes \mathbf{1}_\beta)) \star (\mathbf{I}_{|\psi_\alpha(\mathcal{F})|} \cdot \mathbf{q}[\psi_\alpha(\mathcal{F}), \psi_\beta(\mathcal{N})]) \Big) \\ &\quad\overset{(2.1)}{=} \mathbf{1}_{|\psi_\alpha(\mathcal{F})|} \cdot \Big((\mathbf{X}[\psi_\alpha(\mathcal{F}), :] * \mathbf{I}_{|\psi_\alpha(\mathcal{F})|}) \cdot ((\mathbf{G}|_{\mathcal{N}} \otimes \mathbf{1}_\beta) \odot \mathbf{q}[\psi_\alpha(\mathcal{F}), \psi_\beta(\mathcal{N})]) \Big) \\ &\quad= \Big(\mathbf{1}_{|\psi_\alpha(\mathcal{F})|} \cdot (\mathbf{X}[\psi_\alpha(\mathcal{F}), :] * \mathbf{I}_{|\psi_\alpha(\mathcal{F})|}) \Big) \cdot \Big((\mathbf{G}|_{\mathcal{N}} \otimes \mathbf{1}_\beta) \odot \mathbf{q}[\psi_\alpha(\mathcal{F}), \psi_\beta(\mathcal{N})] \Big)\ . \end{aligned}$$

By Eq. (2.4) and the definition of $\mathbf{X}$, the vector $\mathbf{1}_{|\psi_\alpha(\mathcal{F})|} \cdot (\mathbf{X}[\psi_\alpha(\mathcal{F}), :] * \mathbf{I}_{|\psi_\alpha(\mathcal{F})|})$ is

[5] Observe that the matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ are chosen to show that a specific unit vector is contained as a row of the matrix $(\mathbf{P}_1 \cdot (\mathbf{G} \otimes \mathbf{1}_\beta)) \odot (\mathbf{P}_2 \cdot \mathbf{q})$, which implies that it also in the span of $(\mathbf{G} \otimes \mathbf{1}_\beta) \odot \mathbf{q}$. As we are interested in showing which unit vectors are in the *span*, we do *not* require the matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ to be the same for all unit vectors $\mathbf{e}_l \in \mathbb{F}^{(k+t-1)\beta}, l \in \mathrm{colsupp}(\mathbf{q})$. Instead, it suffices that for each of these unit vectors there *exists* a choice of $\mathbf{P}_1$ and $\mathbf{P}_2$ such that it is a row of the resulting matrix.

uniformly distributed over $\mathbb{F}^{1\times k|\psi_\alpha(\mathcal{F})|}$ and it follows that

$$\begin{aligned}
&H\Big(\sum_{l\in\psi_\alpha(\mathcal{F})}(\mathbf{Y}[l,\mathcal{N}]\otimes\mathbf{1}_\beta)\star\mathbf{q}[l,\psi_\beta(\mathcal{N})]\Big)\\
&\quad= H\Big(\big(\mathbf{1}_{|\psi_\alpha(\mathcal{F})|}\cdot(\mathbf{X}[\psi_\alpha(\mathcal{F}),:]*\mathbf{I}_{|\psi_\alpha(\mathcal{F})|})\big)\cdot\big((\mathbf{G}|_\mathcal{N}\otimes\mathbf{1}_\beta)\odot\mathbf{q}[\psi_\alpha(\mathcal{F}),\psi_\beta(\mathcal{N})]\big)\Big)\\
&\quad= \operatorname{rank}\Big((\mathbf{G}|_\mathcal{N}\otimes\mathbf{1}_\beta)\odot\mathbf{q}[\psi_\alpha(\mathcal{F}),\psi_\beta(\mathcal{N})]\Big)\\
&\quad= |\operatorname{colsupp}(\mathbf{q}[\psi_\alpha(\mathcal{F}),\psi_\beta(\mathcal{N})])|\ ,
\end{aligned}$$

where the last equality holds by Lemma 8.3. □

One key to the proof of the capacity of full support-rank schemes is that while it is generally not possible to make a statement on the expected rank of a query solely based on the requirement that a PIR scheme is t-private, it *is possible* to make such a statement on the expected size of the support of the query.

Lemma 8.5. *For any PIR scheme, file indices $i,i'\in[m]$, and any $\mathcal{F}\subset[m]$ it holds that*

$$\mathop{\mathbb{E}}_{\mathbf{q}\in\operatorname{supp}(Q^i)}\Big(|\operatorname{colsupp}(\mathbf{q}[\psi_\alpha(\mathcal{F}),:])|\Big)=\mathop{\mathbb{E}}_{\mathbf{q}\in\operatorname{supp}(Q^{i'})}\Big(|\operatorname{colsupp}(\mathbf{q}[\psi_\alpha(\mathcal{F}),:])|\Big)\ .$$

Proof. As the scheme is private, the query Q^i_j to each individual server $j\in[n]$ must be independent of the index i, i.e., $\mathbf{Q}^i[:,\psi_\beta(j)]$ and $\mathbf{Q}^{i'}[:,\psi_\beta(j)]$ must have the same probability distribution. Trivially, this implies that the $(|\mathcal{F}|\alpha\times\beta)$-matrices $\mathbf{Q}^i[\psi_\alpha(\mathcal{F}),\psi_\beta(j)]$ and $\mathbf{Q}^{i'}[\psi_\alpha(\mathcal{F}),\psi_\beta(j)]$ also have the same probability distribution and therefore

$$\mathop{\mathbb{E}}_{\mathbf{q}\in\operatorname{supp}(Q^i)}\Big(|\operatorname{colsupp}(\mathbf{q}[\psi_\alpha(\mathcal{F}),:])\cap\psi_\beta(j)|\Big)=\mathop{\mathbb{E}}_{\mathbf{q}\in\operatorname{supp}(Q^{i'})}\Big(|\operatorname{colsupp}(\mathbf{q}[\psi_\alpha(\mathcal{F}),:])\cap\psi_\beta(j)|\Big)\ .$$

Writing the column support as a disjoint union, we get

$$|\operatorname{colsupp}(\mathbf{q}[\psi_\alpha(\mathcal{F}),:])|=\sum_{j\in[n]}|\operatorname{colsupp}(\mathbf{q}[\psi_\alpha(\mathcal{F}),:])\cap\psi_\beta(j)|,$$

and so by additivity of the expectation we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{q}\in\operatorname{supp}(Q^i)}\Big(|\operatorname{colsupp}(\mathbf{q}[\psi_\alpha(\mathcal{F}),:])|\Big)&=\sum_{j\in[n]}\mathop{\mathbb{E}}_{\mathbf{q}\in\operatorname{supp}(Q^i)}\Big(|\operatorname{colsupp}(\mathbf{q}[\psi_\alpha(\mathcal{F}),:])\cap\psi_\beta(j)|\Big)\\
&=\sum_{j\in[n]}\mathop{\mathbb{E}}_{\mathbf{q}\in\operatorname{supp}(Q^{i'})}\Big(|\operatorname{colsupp}(\mathbf{q}[\psi_\alpha(\mathcal{F}),:])\cap\psi_\beta(j)|\Big)\\
&=\mathop{\mathbb{E}}_{\mathbf{q}\in\operatorname{supp}(Q^{i'})}\Big(|\operatorname{colsupp}(\mathbf{q}[\psi_\alpha(\mathcal{F}),:])|\Big)\ .
\end{aligned}$$ □

With these technical preliminary lemmas established, we are now ready to show that subsets of servers of size $k+t-1$ treat the desired and undesired files similarly for full support-rank schemes.

Lemma 8.6. *Let $\mathcal{N} \subset [n]$ with $n > k+t-1$ be a set of $|\mathcal{N}| = k+t-1$ honest (nonadversarial) servers and $\mathcal{F} \subsetneq [m]$ be any proper subset of the thick rows of the MDS-coded storage system. For any optimal linear, full support-rank PIR scheme, and any $i, i' \in [m]$, it holds that*

$$H(A^i_{\mathcal{N}} \mid X^{\mathcal{F}}, Q^i_{\mathcal{N}}) = H(A^{i'}_{\mathcal{N}} \mid X^{\mathcal{F}}, Q^{i'}_{\mathcal{N}}) \ . \tag{8.8}$$

Proof. First note that an equivalent problem formulation[6] is given by (cf. [CT91, Section 2.2, Eq. (2.10)])

$$\begin{aligned}&\mathop{\mathbb{E}}_{\mathbf{q}[:,\psi_\beta(\mathcal{N})]\in\mathrm{supp}(Q^i_{\mathcal{N}})}\Big(H(A^i_{\mathcal{N}} \mid X^{\mathcal{F}}, Q^i_{\mathcal{N}} = \mathbf{q}[:,\psi_\beta(\mathcal{N})])\Big)\\ &\qquad\qquad = \mathop{\mathbb{E}}_{\mathbf{q}[:,\psi_\beta(\mathcal{N})]\in\mathrm{supp}(Q^{i'}_{\mathcal{N}})}\Big(H(A^i_{\mathcal{N}} \mid X^{\mathcal{F}}, Q^{i'}_{\mathcal{N}} = \mathbf{q}[:,\psi_\beta(\mathcal{N})])\Big) \ .\end{aligned}$$

Further, observe that by Definition 8.1 the responses of servers $j \in [\mathcal{N}]$ can be expressed as the star-product (Hadamard product) between rows of the restricted query matrix $\mathbf{Q}^i[:,\psi_\beta(\mathcal{N})]$ and the restricted storage matrix with each column repeated β times, i.e., $\mathbf{Y} \otimes \mathbf{1}_\beta$, where $\otimes$ denotes the Kronecker product. Specifically, we have

$$\begin{aligned}A^i_{\mathcal{N}} &= \mathbf{A}^i[:,\psi_\beta(\mathcal{N})]\\ &= \Big(\Big\langle \mathbf{Y}[:,j], \mathbf{Q}^i[:,(j-1)\beta+s]\Big\rangle\Big)_{s\in[\beta], j\in\mathcal{N}}\\ &= \Big(\mathbf{Y}[:,j]^\top \cdot \mathbf{Q}^i[:,\psi_\beta(j)]\Big)_{j\in\mathcal{N}}\\ &= \sum_{l\in\psi_\alpha([m])}\Big((\mathbf{Y}[l,j]\otimes\mathbf{1}_\beta)\star\mathbf{Q}^i[l,\psi_\beta(j)]\Big)_{j\in\mathcal{N}}\\ &= \sum_{l\in\psi_\alpha([m])}\Big((\mathbf{Y}[l,\mathcal{N}]\otimes\mathbf{1}_\beta)\star\mathbf{Q}^i[l,\psi_\beta(\mathcal{N})]\Big) \ .\end{aligned} \tag{8.9}$$

Next, we show that for every query realization, the entropies only depend on the size

[6] We choose to refer to the realizations of $\mathcal{Q}^i_{\mathcal{N}}$ as $\mathbf{q}[:,\psi_\beta(\mathcal{N})]$ to be consistent with notation and indexing, i.e., we treat the realizations of $\mathcal{Q}^i_{\mathcal{N}}$ as a submatrix consisting of $k+t-1$ thick columns of the realizations $\mathbf{q}$ of $\mathcal{Q}^i$.

of the support of the query realization as

$$\begin{aligned}
&H(A^i_{\mathcal{N}} \mid X^{\psi_\alpha(\mathcal{F})}, Q^i_{\mathcal{N}} = \mathbf{q}[:, \psi_\beta(\mathcal{N})]) \\
&\quad = H\Big(\sum_{l \in \psi_\alpha([m])} \Big((\mathbf{Y}[l, \mathcal{N}] \otimes \mathbf{1}_\beta) \star \mathbf{q}[l, \psi_\beta(\mathcal{N})] \Big) \;\Big|\; X^{\mathcal{F}}, Q^i_{\mathcal{N}} = \mathbf{q}[:, \psi_\beta(\mathcal{N})] \Big) \\
&\quad = H\Big(\sum_{l \in \psi_\alpha([m] \setminus \mathcal{F})} \Big((\mathbf{Y}[l, \mathcal{N}] \otimes \mathbf{1}_\beta) \star \mathbf{q}[l, \psi_\beta(\mathcal{N})] \Big) \;\Big|\; Q^i_{\mathcal{N}} = \mathbf{q}[:, \psi_\beta(\mathcal{N})] \Big) \\
&\quad \overset{(\mathsf{a})}{=} \Big| \operatorname{colsupp} \Big(\mathbf{q}[\psi_\alpha([m] \setminus \mathcal{F}), \psi_\beta(\mathcal{N})] \Big) \Big| \;,
\end{aligned}$$

where (a) holds by Lemma 8.4. Taking the expectation over the support of $Q^i_{\mathcal{N}}$ gives

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathbf{q}[:, \psi_\beta(\mathcal{N})] \in \operatorname{supp}(Q^i_{\mathcal{N}})} \Big(H(A^i_{\mathcal{N}} \mid X^{\mathcal{F}}, Q^i_{\mathcal{N}} = \mathbf{q}[:, \psi_\beta(\mathcal{N})]) \Big) \\
&\qquad = \mathop{\mathbb{E}}_{\mathbf{q}[:, \psi_\beta(\mathcal{N})] \in \operatorname{supp}(Q^i_{\mathcal{N}})} \Big(| \operatorname{colsupp}(\mathbf{q}[\psi_\alpha([m] \setminus \mathcal{F}), \psi_\beta(\mathcal{N})]) | \Big) \\
&\qquad \overset{(\mathsf{a})}{=} \mathop{\mathbb{E}}_{\mathbf{q}[:, \psi_\beta(\mathcal{N})] \in \operatorname{supp}(Q^{i'}_{\mathcal{N}})} \Big(| \operatorname{colsupp}(\mathbf{q}[\psi_\alpha([m] \setminus \mathcal{F}), \psi_\beta(\mathcal{N})]) | \Big) \\
&\qquad = \mathop{\mathbb{E}}_{\mathbf{q}[:, \psi_\beta(\mathcal{N})] \in \operatorname{supp}(Q^{i'}_{\mathcal{N}})} \Big(H(A^{i'}_{\mathcal{N}} \mid X^{\mathcal{F}}, Q^{i'}_{\mathcal{N}} = \mathbf{q}[:, \psi_\beta(\mathcal{N})]) \Big) \;,
\end{aligned}$$

where (a) follows from Lemma 8.5. □

While the previous lemma holds for any pair of indices $i, i' \in [m]$, the interesting case is when $i \in \mathcal{F}$, $i' \notin \mathcal{F}$. Intuitively, in this case the statement implies that $(k + t - 1)$-tuples of servers handle desired and undesired files equally, which will be used in the inductive proof of Lemma 8.7. Also note that the property of full support-rank was needed in the proof of Lemma 8.3, the key technical ingredient to the proof of Lemma 8.4 and thereby also to Lemma 8.6, as it ensures that the given entropy expression is equal to the size of the column support of the query restricted to the respective rows and columns.

Remark 8.2. *The formulation of the server responses used in Lemma 8.6 implies a novel formulation of the PIR problem with linear decoding functions. As shown in Eq.* (2.4) *and Lemma 8.4, the received responses are given by (to simplify the notation we assume* $\alpha = 1$ *here)*

$$\begin{aligned}
\mathbf{A}^i = (\mathbf{X}[1,1], \mathbf{X}[2,1], \ldots, \mathbf{X}[m,1], \mathbf{X}[1,2], \mathbf{X}[2,2], \ldots, \mathbf{X}[m,2], \ldots, \mathbf{X}[m,k]) \quad (8.10) \\
\cdot \Big((\mathbf{G} \otimes \mathbf{1}_\beta) \odot \mathbf{Q}^i \Big) \;,
\end{aligned}$$

where $\odot$ *denotes the column-wise Khatri-Rao product [KR68] and* $\mathbf{G}$ *is a generator matrix of the storage code. When restricting to linear decoding functions, the application of a decoder* $\mathfrak{D}$ *such that* $\mathfrak{D}(\mathbf{A}) = \mathbf{X}^i$*, is equivalent to performing linear combinations* *of the received responses $\mathbf{A}^i$, which, in turn, is equivalent to performing linear combinations of the columns of $(\mathbf{G} \otimes \mathbf{1}_\beta) \odot \mathbf{Q}^i$. It is easy to see that the l-th symbol of the information vector can be obtained exactly if the l-th unit vector $\mathbf{e}_l$ is in the column span of this matrix. Therefore, the problem of linear PIR with linear decoding functions can be defined solely based on operations from linear algebra: For each $i \in [m]$ determine a distribution of query matrices $\mathbf{Q}^i$ such that*

$$\begin{aligned} \mathbf{e}_l \in \left\langle (\mathbf{G} \otimes \mathbf{1}_\beta) \odot \mathbf{Q}^i \right\rangle_{\mathsf{col}} \quad & \forall\, l \in \{i, m+i, 2m+i, \ldots, (k-1)m+i\} \\ \Pr(\mathbf{Q}^i_{\mathcal{T}} = \mathbf{q}) = \Pr(\mathbf{Q}^{i'}_{\mathcal{T}} = \mathbf{q}) \quad & \forall\, i, i' \in [m], \mathcal{T} \subset [n], |\mathcal{T}| \leq t \ . \end{aligned}$$

The first condition guarantees decodability, as the given set indexes the symbols of file i in the data vector of Eq. (8.10)*, while the second condition guarantees t-privacy.*

The following lemma will be used to prove the upper bounds on the nonsymmetric MDS-TPIR capacity.

Lemma 8.7. *Consider an optimal linear (S)PIR scheme, and let $\mathcal{H} \subset [n]$ be a minimal set (set of smallest possible cardinality) such that the requested file i can be obtained from the respective responses, i.e.,*

$$H(X^i \mid A^i_{\mathcal{H}}, \mathcal{Q}) = 0 \ .$$

For $1 \leq s \leq m$, let

$$h_s = \frac{n}{|\mathcal{H}|} H(A^s_{\mathcal{H}} \mid Q, X^{[s-1]})$$

and $h_{m+1} = 0$. Then, for all $1 \leq s \leq m$,

$$h_s \geq \frac{n}{n-2b-r} \left(H(X^s) + \frac{k+t-1}{n} h_{s+1} \right) \ .$$

Proof. By Lemma 8.1, we have $|\mathcal{H}| \leq n - 2b - r$. By Han's inequality (see Eq. (8.5)), the average value of $H(A^{s+1}_{\mathcal{N}} \mid Q, X^{[s]})$ over all sets $\mathcal{N} \subseteq \mathcal{H}$ with $|\mathcal{N}| = k+t-1$ is at least

$$\frac{k+t-1}{|\mathcal{H}|} H(A^{s+1}_{\mathcal{H}} \mid Q, X^{[s]}) \ .$$

Hence, we can choose a set $\mathcal{N} \subseteq \mathcal{H}$ with $|\mathcal{N}| = k+t-1$ such that

$$\begin{aligned} H(A^{s+1}_{\mathcal{N}} \mid Q, X^{[s]}) &\geq \frac{k+t-1}{|\mathcal{H}|} H(A^{s+1}_{\mathcal{H}} \mid Q, X^{[s]}) \\ &= \frac{k+t-1}{n} h_{s+1} \ . \end{aligned}$$

By independence of the files and the queries, we have

$$H(X^s \mid Q, X^{[s-1]}) = H(X^s) \ .$$

We thus get

$$\begin{aligned} h_s &= \frac{n}{|\mathcal{H}|} H(A^s_{\mathcal{H}} \mid Q, X^{[s-1]}) \\ &= \frac{n}{|\mathcal{H}|} \left(H(X^s) + H(A^s_{\mathcal{H}} \mid Q, X^{[s]}) \right) \\ &\geq \frac{n}{|\mathcal{H}|} \left(H(X^s) + H(A^s_{\mathcal{N}} \mid Q, X^{[s]}) \right) \\ &\overset{(\mathsf{a})}{=} \frac{n}{|\mathcal{H}|} \left(H(X^s) + H(A^{s+1}_{\mathcal{N}} \mid Q, X^{[s]}) \right) \\ &\geq \frac{n}{|\mathcal{H}|} \left(H(X^s) + \frac{k+t-1}{n} h_{s+1} \right) \\ &\geq \frac{n}{n-2b-r} \left(H(X^s) + \frac{k+t-1}{n} h_{s+1} \right) \ , \end{aligned}$$

where (a) follows from Lemma 8.6. □

Setting $b = r = 0$, we are now ready to prove the capacity of linear, full support-rank MDS-TPIR. Note that this settles Conjecture 8.3 under this technical assumption.

Theorem 8.2 (Capacity of linear, Full Support-Rank MDS-TPIR)**.** *Let $n, k, t,$ and m be integers with $n > k + t - 1$ and $m \geq 2$. The capacity of linear, full support-rank MDS-TPIR, i.e., PIR from $[n, k]$ MDS-coded storage with t colluding servers where the queries fulfill Definition 8.2, is*

$$C^{\mathsf{MDS}}_{\mathsf{FSR-TPIR}} = \frac{1 - \frac{k+t-1}{n}}{1 - \left(\frac{k+t-1}{n}\right)^m} \quad \xrightarrow{m \to \infty} \quad 1 - \frac{k+t-1}{n} \ .$$

Proof. Achievability: An explicit scheme achieving the rate is constructed in [DE19] by "lifting" the star product scheme of [FGHK17]. To be private, this scheme needs to fulfill the definition of Definition 8.2, as argued in Section 8.4.2.

Converse: Let $\mathcal{H} \subset [n]$ be a minimal set such that

$$H(X^i \mid A^i_{\mathcal{H}}, \mathcal{Q}) = 0,$$

and for $s \in [m]$, let

$$h_s = \frac{n}{|\mathcal{H}|} H(A^s_{\mathcal{H}} \mid Q, X^{[s-1]})$$

as in Lemma 8.7. Denote $L = H(X^i)$ and notice that this is equal for all files i. By definition and Lemma 8.2, the rate R of the scheme satisfies

$$\begin{aligned}\frac{1}{R} &= \frac{\sum_{j\in[n]} H(A_j^s)}{L} \\ &\geq \frac{\sum_{j\in[n]} H(A_j^s \mid Q)}{L} \\ &\geq \frac{h_1}{L} \ ,\end{aligned}$$

where the last equation follows by minimality of $\mathcal{H}$. It is thus enough to show that

$$\frac{h_s}{L} \geq \frac{1-\left(\frac{k+t-1}{n}\right)^{m-s+1}}{1-\frac{k+t-1}{n}} \tag{8.11}$$

holds for all $1 \leq s \leq m$. We will prove this by backwards induction on s.

As the base case consider $s = m$ and observe that Eq. (8.11) simplifies to $h_s \geq L$ in this case. Recall that A^m is a function of the files X and the queries Q. As we have $b = r = 0$, Lemma 8.7 gives

$$h_m = H(A_{\mathcal{H}}^m \mid Q, X^{[m-1]}) = H(A_{\mathcal{H}}^m, X^m \mid Q, X^{[m-1]}) = H(X^m) = L \ .$$

It follows that Eq. (8.11) is correct for $s = m$. Now assume as an induction hypothesis that

$$\frac{h_{s'}}{L} \geq \frac{1-\left(\frac{k+t-1}{n}\right)^{m-s'+1}}{1-\frac{k+t-1}{n}} \ ,$$

and let $s = s' - 1$. Then Lemma 8.7 yields

$$\begin{aligned}\frac{h_s}{L} &\geq 1 + \left(\frac{k+t-1}{n}\frac{h_{s'}}{L}\right) \\ &\geq 1 + \frac{\frac{k+t-1}{n} - \left(\frac{k+t-1}{n}\right)^{m-s'+2}}{1-\frac{k+t-1}{n}} \\ &= \frac{1-\left(\frac{k+t-1}{n}\right)^{m-s+1}}{1-\frac{k+t-1}{n}} \ .\end{aligned}$$

This proves Eq. (8.11) for all $1 \leq s \leq m$ by induction. The case $s = 1$ is the statement of the theorem. □

Remark 8.3. *By similar techniques, we get an upper bound*

$$C^{\mathsf{MDS}}_{\mathsf{FSR-TBPIR}} \leq \left(1 - \frac{2b+r}{n}\right) \cdot \frac{1 - \frac{k+t-1}{n}}{1 - \left(\frac{k+t-1}{n}\right)^m} \tag{8.12}$$

for the case where we also have adversarial and nonresponsive servers. However, we believe this to be a loose upper bound. If the bound Eq. (8.12) *were to be tight, the result of [WS17a; WS17b] (see Theorem 8.3) would imply that in this setting symmetric PIR has a strictly lower capacity than PIR even as the number of files goes to infinity. This would be in sharp contrast to the known cases of TPIR/TSPIR and MDS-PIR/MDS-SPIR with and without adversarial/nonresponsive servers, where the nonsymmetric capacity converges (from above) to the symmetric capacity as the number of files increases.*

8.4.2 Known PIR Schemes and the Definition of Full Support-Rank

With the capacity of full support-rank PIR schemes established, we now discuss two schemes related to this result, beginning with the approach of refinement and lifting introduced in [DE19], a method for increasing the rate of a class of PIR schemes. Next, we discuss the only known scheme that exceeds the rate achievable by full support-rank schemes.

Lifted PIR Schemes

In this section we aim to clarify some of the details of the refinement operation of [DE19]. This operation is based on choosing vectors such that their respective inner product with the stored vectors are "linearly independent random variables". Given the application of these rules in [DE19, Example 7], this appears to mean that the corresponding columns in the column-wise Khatri-Rao product of the matrix of storage vectors and the matrix of the query vectors are linearly independent. However, as we discuss in the following, this is not sufficient for the scheme to be private.

We consider [DE19, Example 7] for the setting $n = 4$ and $k = t = 2$. There and in the following, file 1 is assumed to be desired by the user. The storage code is a $[4, 2]$ MDS code over $\mathbb{F}_3$ with generator matrix (cf. [DE19, Table VII])

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \end{pmatrix} .$$

Following the notation of [DE19], the linear combinations used to obtain $\mathbf{x}_3^2$ and $\mathbf{x}_4^2$

can be written as

$$[\mathbf{x}_1^2, \mathbf{x}_2^2, \mathbf{x}_3^2, \mathbf{x}_4^2] = [\mathbf{x}_1^2, \mathbf{x}_2^2] \cdot \mathbf{G} \tag{8.13}$$

and we therefore also refer to the code generated by $\mathbf{G}$ as the *query code*[7]. For the desired file 1, the vectors $\mathbf{x}_j^1$ are chosen uniformly at random from all query vectors of $\mathbb{F}_3^{\alpha 2\times 1}$ supported only on file 1 (cf. [DE19, Definition 1]) and such that the $\left\langle \mathbf{Y}_j, \mathbf{x}_j^1 \right\rangle$ are "linearly independent random variables". For the undesired file 2, the vectors $\mathbf{x}_1^2$ and $\mathbf{x}_2^2$ are chosen uniformly at random from all query vectors supported only on file 2 and such that $\langle \mathbf{Y}_1, \mathbf{x}_1^2 \rangle$ and $\langle \mathbf{Y}_2, \mathbf{x}_2^2 \rangle$ are "linearly independent random variables". The vectors $\mathbf{x}_3^2$ and $\mathbf{x}_4^2$ are given by Eq. (8.13).

We set the subpacketization to be $\alpha = 2$, i.e., the storage is a length 4 vector, where the first two positions correspond to file 1 and the other two positions to file 2. Now assume the following realizations of $\mathbf{x}^1$ and $\mathbf{x}^2$ (the j-th column of $\mathbf{x}^l$ gives $\mathbf{x}_j^l$)

$$\mathbf{x}^1 = \left(\begin{array}{cccc} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ \hdashline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \quad \mathbf{x}^2 = \left(\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hdashline 1 & 2 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right) .$$

By [DE19, Lemma 1] and written in terms of our notation[8], the query is then given by

$$\mathbf{q} = \left(\begin{array}{cc:cc:cc:c} 1 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 2 \\ \hdashline 0 & 1 & 0 & 2 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) ,$$

where server j receives the j-th thick column, as indicated by the dashed lines. We make the following observations:

- The $\left\langle \mathbf{Y}_j, \mathbf{x}_j^1 \right\rangle, j = 1, 2, 3, 4$, are indeed linearly independent, as any two columns of the storage code are linearly independent.
- By the same argument, $\langle \mathbf{Y}_1, \mathbf{x}_1^2 \rangle$ and $\langle \mathbf{Y}_2, \mathbf{x}_2^2 \rangle$ are linearly independent.
- The third and fourth columns of $\mathbf{x}^2$, i.e., $\mathbf{x}_3^2$ and $\mathbf{x}_4^2$, are as in Eq. (8.13).

As $\mathbf{x}^1$ and $\mathbf{x}^2$ are chosen uniformly at random such that these properties are fulfilled, this is a query realization with nonzero probability. However, since $\langle \mathbf{Y}_3, \mathbf{x}_3^2 \rangle = \langle \mathbf{Y}_3, \mathbf{0} \rangle = 0$, the query $\mathbf{x}_3^2$ is not a valid query if file 2 is the desired file. Hence,

[7] In general, the storage and query code do not need to be the same.

[8] Here, the fourth server only receives one query, so the fourth "thick" column is only one column wide.

upon receiving the queries $\mathbf{x}_3^1$ and $\mathbf{x}_3^2$, server 3 is able to deduce that file 2 is not the desired file. Further, observe that here we have $\mathbf{x}^1[\{1,2\},1] = \mathbf{x}^2[\{3,4\},1]$ and $\mathbf{x}^1[\{1,2\},2] = \mathbf{x}^2[\{3,4\},2]$, so simply excluding this case for the undesired file is not an option, as this would allow servers one and two to deduce that file 1 is the desired file.

It is easy to see the problem here is that while $\langle \mathbf{Y}_1, \mathbf{x}_1^2 \rangle$ and $\langle \mathbf{Y}_2, \mathbf{x}_2^2 \rangle$ give linearly independent random variables, the vectors $\mathbf{x}_1^2$ and $\mathbf{x}_2^2$ themselves are not linearly independent. This leads to an $\mathbf{x}_3^2$ that trivially results in a "linearly dependent random variable". One solution to this problem is requiring the vectors $\mathbf{x}_1^2$ and $\mathbf{x}_2^2$ themselves to be linearly independent. In this case, assuming the query code is MDS[9], any t-subset of columns in $\mathbf{x}^2$ is linearly independent, which guarantees that it is also a valid choice for the desired file. In other words, the submatrix of $\mathbf{x}^2$ corresponding to file 2, i.e., the two bottom rows, has to be chosen as a (random) basis of the MDS query code. Obviously, the privacy requirement implies that $\mathbf{x}^1$ also needs to be chosen such that any t-subset of columns is linearly independent.

In conclusion, a "fix" to this ambiguity, which ensures the privacy of this scheme, is given by requiring that the supported columns of any subset of t thick columns of $\mathbf{x}^j$ are linearly independent, exactly as required in Definition 8.2. Note that, our proposed fix allows the scheme to achieve the highest rate possible (for this specific scheme, not necessarily in general). Hence, albeit it might be possible to find a different distribution that also results in a private scheme, there is no advantage to be gained in terms of rate for this class of lifted schemes.

A Scheme that does not fulfill Definition 8.2

In [SJ18a], a linear PIR scheme from $[n = 4, k = 2]$ MDS-coded storage with $t = 2$ colluding servers and $m = 2$ files was presented, achieving a PIR rate $3/5$. This rate exceeds the one in Conjecture 8.3, thereby providing a counter-example that disproves it in its full generality. In the following, we briefly introduce this counter-example with a focus on the query construction and show that it does not fulfill Definition 8.2.

Each of the two files is assumed to be comprised of 12 symbols from $\mathbb{F}_p$ for a large prime p and the subpacketization level is set to $\alpha = 6$. Let

$$\begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_6 \end{pmatrix}, \begin{pmatrix} \mathbf{U}_0 \\ \mathbf{U}_1 \\ \vdots \\ \mathbf{U}_5 \end{pmatrix}$$

be two random full-rank 6×6 matrices over $\mathbb{F}_p$. Without loss of generality, suppose

[9] Note that, while it is unclear whether this is strictly necessary, this assumption does not lower the PIR rate, as can be seen from applying the refinement/lifting operation of [DE19] to the star-product scheme of [FGHK17] utilizing Reed–Solomon codes.

that the first file is desired. The queries to servers 1 and 2 are, respectively,

$$Q_1^1 = \begin{pmatrix} \mathcal{L}_{11}(\mathbf{V}_1^\top,\mathbf{V}_2^\top,\mathbf{V}_3^\top) & \mathcal{L}_{12}(\mathbf{V}_1^\top,\mathbf{V}_2^\top,\mathbf{V}_3^\top) & & & \mathcal{L}_{13}(\mathbf{V}_1^\top,\mathbf{V}_2^\top,\mathbf{V}_3^\top) \\ & & \mathcal{L}_{11}(\mathbf{U}_0^\top,\mathbf{U}_6^\top,\mathbf{U}_8^\top) & \mathcal{L}_{12}(\mathbf{U}_0^\top,\mathbf{U}_6^\top,\mathbf{U}_8^\top) & \mathcal{L}_{13}(\mathbf{U}_0^\top,\mathbf{U}_6^\top,\mathbf{U}_8^\top) \end{pmatrix} ,$$

and

$$Q_2^1 = \begin{pmatrix} \mathcal{L}_{21}(\mathbf{V}_1^\top,\mathbf{V}_4^\top,\mathbf{V}_5^\top) & \mathcal{L}_{22}(\mathbf{V}_1^\top,\mathbf{V}_4^\top,\mathbf{V}_5^\top) & & & \mathcal{L}_{23}(\mathbf{V}_1^\top,\mathbf{V}_4^\top,\mathbf{V}_5^\top) \\ & & \mathcal{L}_{21}(\mathbf{U}_0^\top,\mathbf{U}_7^\top,\mathbf{U}_9^\top) & \mathcal{L}_{22}(\mathbf{U}_0^\top,\mathbf{U}_7^\top,\mathbf{U}_9^\top) & \mathcal{L}_{23}(\mathbf{U}_0^\top,\mathbf{U}_7^\top,\mathbf{U}_9^\top) \end{pmatrix} ,$$

where $\mathcal{L}_{ij}(\mathbf{a},\mathbf{b},\mathbf{c})$ denotes some linear combinations of $\mathbf{a},\mathbf{b},\mathbf{c}$ (see P1 and P2 in [SJ18a, Pg. 1004] for more details on the requirements on the coefficients of the involved linear combinations), and

$$\begin{aligned} \mathbf{U}_6 &= \mathbf{U}_1 + \mathbf{U}_2, \quad & \mathbf{U}_7 &= \mathbf{U}_1 + 2\mathbf{U}_2, \\ \mathbf{U}_8 &= \mathbf{U}_3 + \mathbf{U}_4, \quad & \mathbf{U}_9 &= \mathbf{U}_3 + 2\mathbf{U}_4. \end{aligned}$$

Note that this definition includes the processing step done at the servers in [SJ18a] as part of the query, which is necessary to describe the scheme as a linear scheme as in Definition 8.1. Then, in our notation for the query, we have for $\mathcal{F} = \{1\}$ and $\mathcal{T} = \{1,2\}$

$$\mathbf{q}[\psi_\alpha(\mathcal{F}),\psi_\beta(\mathcal{T})] = \mathbf{q}[\psi_\alpha(1),\psi_\beta(\{1,2\})] = \mathbf{q}[[6],[12]] =$$
$$\begin{pmatrix} \mathcal{L}_{11}(\mathbf{V}_1^\top,\mathbf{V}_2^\top,\mathbf{V}_3^\top) & \mathcal{L}_{12}(\mathbf{V}_1^\top,\mathbf{V}_2^\top,\mathbf{V}_3^\top) & \mathbf{0}_{6\times 2} & \mathcal{L}_{13}(\mathbf{V}_1^\top,\mathbf{V}_2^\top,\mathbf{V}_3^\top) & \mathcal{L}_{21}(\mathbf{V}_1^\top,\mathbf{V}_4^\top,\mathbf{V}_5^\top) & \mathcal{L}_{22}(\mathbf{V}_1^\top,\mathbf{V}_4^\top,\mathbf{V}_5^\top) & \mathbf{0}_{6\times 2} & \mathcal{L}_{23}(\mathbf{V}_1^\top,\mathbf{V}_4^\top,\mathbf{V}_5^\top) \end{pmatrix} ,$$

where $\mathbf{0}_{6\times 2}$ denotes the 6×2 zero matrix. The matrix $\mathbf{q}[\psi_\alpha(1),\psi_\beta(\{1,2\})]$ is a 6×10 matrix with 6 nonzero columns that are linear combinations of the 5 vectors $\mathbf{V}_1^\top, \mathbf{V}_2^\top, \mathbf{V}_3^\top, \mathbf{V}_4^\top$, and $\mathbf{V}_5^\top$. Therefore, we have

$$\operatorname{rank}(\mathbf{q}[\psi_\alpha(1),\psi_\beta(\{1,2\})]) \leq 5 < 6 = |\operatorname{colsupp}(\mathbf{q}[\psi_\alpha(1),\psi_\beta(\{1,2\})])| ,$$

and conclude that the PIR scheme in [SJ18a] does not fulfill Definition 8.2.

While it might seem excessive to describe a scheme that does *not* fall into the class of full support-rank PIR schemes in this much detail, we would like to point out that this in fact further motivates our definition. The results presented in Section 8.4 show that *the* distinguishing feature of this scheme is in fact the low rank of the queries, when restricting to a subset of thick columns and rows, thereby strongly hinting at what a scheme for general parameters and of a PIR rate that exceeds the one in Conjecture 8.3 and Theorem 8.2 must fulfill.

8.5 Capacity of MDS-coded TBSPIR for Schemes with Additive Randomness

In this section, we prove the capacity of MDS-coded TBSPIR for the specific system model considered in [WS19]. Recent works [WSS19; WS19] have shown that it is crucial to specify the distribution of the randomness shared by the servers when deriving the capacity of such systems. We begin by shortly reviewing the results presented in these works. In [WSS19] the authors derive the capacity of MDS-coded SPIR with *mismatched randomness*, meaning that they assume the complete randomness to be available to all servers. It is shown that this assumption of sharing of the complete randomness among the servers leads to a strictly larger rate than when the randomness is also coded with the MDS storage code, referred to as *matched randomness*. The resulting capacity approaches the capacity of coded, matched SPIR when the number of files tends to infinity and is always strictly lower than the capacity of coded PIR without the symmetry requirement.

In [WS19] the authors derive the capacity of MDS-coded SPIR with and without collusion for the case of matched randomness, i.e., where the randomness is also encoded with the MDS storage code. Further, they consider the special case of schemes with additive randomness independent of the queries. Specifically, the authors show

- the capacity of SPIR from $[n, k]$ MDS-coded storage, where for any k servers the randomness is independent (matched randomness), to be

$$C^{\mathsf{MDS}}_{\mathsf{matched-SPIR}} = 1 - \frac{k}{n} \ .$$

- the capacity of uncoded TSPIR, i.e., symmetric PIR from n servers encoded with a repetition code ($k = 1$) where up to t servers (TSPIR) can collude, to be

$$C^{\mathsf{Rep}}_{\mathsf{TSPIR}} = 1 - \frac{t}{n} \ .$$

- the capacity of $[n, k]$ MDS-coded TSPIR, for schemes where the servers add the randomness to the responses and the randomness is independent of the queries, to be

$$C^{\mathsf{MDS}}_{\mathsf{add.-TSPIR}} = 1 - \frac{k+t-1}{n} \ .$$

In this section we consider the extension of the results from [WS19] to the MDS-TBSPIR setting, i.e., to symmetric PIR from coded databases in the presence of up to b adversarial servers and r nonresponsive servers.

We begin by formally defining the considered setting.

Definition 8.7 (Matched SBPIR [WS19])**.** *We say a BSPIR scheme is* matched *if the randomness shared by the servers is independent for every subset of k servers.*

Definition 8.8 (Additive randomness TBSPIR [WS19])**.** *We define a scheme to be an* additive randomness *TBSPIR scheme if the responses are of the form*

$$A_j^i = f_j(Q_j^i, Y_j) + S_j \ ,$$

where f_j is an arbitrary function and S_j is independent of the received query Q_j^i.

In Lemma 8.1 we have shown that the desired file must be recoverable from any subset of $\geq n - 2b - r$ honest and responsive servers. The following lemma establishes the final ingredient required for proving the converse of Theorem 8.3.

Lemma 8.8. *For any MDS-TBSPIR scheme and for any set of honest (nonadversarial) servers $\mathcal{N} \subset [n]$ with $|\mathcal{N}| = k + t - 1$ it holds that*

$$H(A_{\mathcal{N}}^i \mid X^i, Q_{\mathcal{N}}^i) = H(A_{\mathcal{N}}^i \mid Q_{\mathcal{N}}^i) \ ,$$

if the randomness is additive as in Definition 8.8 or $t = 1$.

Proof. The proof for the case of additive randomness follows directly from the proof of [WS19, Lemma 8], as it is independent of the total number of servers and, by definition, all servers in $\mathcal{N}$ are honest. By the same argument the proof of [WS19, Lemma 7] also applies here for the case of $t = 1$. □

We are now ready to present the main statement of this section, the capacity of linear MDS-TSPIR for the shared randomness distributions of Definition 8.8 and Definition 8.7.

Theorem 8.3. *The capacity of linear MDS-TBSPIR, i.e., PIR from $[n, k]$ MDS-coded storage with t colluding, b adversarial, and r nonresponsive servers, is*

$$C_{\mathsf{TBSPIR}}^{\mathsf{MDS}} = 1 - \frac{k + t + 2b + r - 1}{n} \ ,$$

if the randomness is additive as defined in Definition 8.8 or, for $t = 1$, as in Definition 8.7.

Proof. Achievability: The symmetric version of the scheme introduced in [TGK⁺19], which generalizes the scheme of [FGHK17], achieves the presented upper bound on the PIR rate. Note that this scheme fulfills both Definition 8.7 and Definition 8.8, since the symmetry is achieved by adding a random codeword of the $[n, k]$ MDS storage code to the answers.

Converse: Let $\mathcal{H} \subset [n]$ and $\mathcal{N} \subset \mathcal{H}$ be sets of honest, responsive servers with $|\mathcal{H}| = n - 2b - r$ and $|\mathcal{N}| = k + t - 1$. Then

$$\begin{aligned}
H(X^i) &\overset{(\mathsf{a})}{=} H(X^i \mid \mathcal{Q}) \\
&\overset{(\mathsf{b})}{=} H(X^i \mid \mathcal{Q}) - H(X^i \mid A^i_{\mathcal{H}}, \mathcal{Q}) \\
&= I(X^i; A^i_{\mathcal{H}} \mid \mathcal{Q}) \\
&= H(A^i_{\mathcal{H}} \mid \mathcal{Q}) - H(A^i_{\mathcal{H}} \mid X^i, \mathcal{Q}) \\
&\overset{(\mathsf{c})}{\leq} H(A^i_{\mathcal{H}} \mid \mathcal{Q}) - H(A^i_{\mathcal{N}} \mid X^i, \mathcal{Q}, Q^i_{\mathcal{N}}) \\
&\overset{(\mathsf{d})}{=} H(A^i_{\mathcal{H}} \mid \mathcal{Q}) - H(A^i_{\mathcal{N}} \mid X^i, Q^i_{\mathcal{N}}) \\
&\overset{(\mathsf{e})}{=} H(A^i_{\mathcal{H}} \mid \mathcal{Q}) - H(A^i_{\mathcal{N}} \mid Q^i_{\mathcal{N}}) \\
&\leq H(A^i_{\mathcal{H}} \mid \mathcal{Q}) - H(A^i_{\mathcal{N}} \mid \mathcal{Q}) \ ,
\end{aligned}$$

where equality (a) holds because the files are independent of the queries, (b) holds by Lemma 8.1, (c) holds because $\mathcal{N} \subset \mathcal{H}$, (d) holds by Lemma 8.2, and (e) holds by Lemma 8.8.

Averaging over all sets $\mathcal{N}$ gives

$$H(X^i) \leq H(A^i_{\mathcal{H}} \mid \mathcal{Q}) - \frac{1}{\binom{n-2b-r}{k+t-1}} \sum_{\substack{\mathcal{N} \subset \mathcal{H} \\ |\mathcal{N}| = k+t-1}} H(A^i_{\mathcal{N}} \mid \mathcal{Q})$$

and by Han's inequality (see Eq. (8.5))

$$\frac{1}{\binom{n-2b-r}{k+t-1}} \sum_{\substack{\mathcal{N} \subset \mathcal{H} \\ |\mathcal{N}| = k+t-1}} H(A^i_{\mathcal{N}} \mid \mathcal{Q}) \geq \frac{k+t-1}{n-2b-r} H(A^i_{\mathcal{H}} \mid \mathcal{Q}).$$

Hence, there exists an $h \in \mathcal{H}$ such that

$$\begin{aligned}
H(X^i) &\leq H(A^i_{\mathcal{H}} \mid \mathcal{Q}) - \frac{k+t-1}{n-2b-r} H(A^i_{\mathcal{H}} \mid \mathcal{Q}) \\
&= \frac{n-k-2b-r-t+1}{n-2b-r} H(A^i_{\mathcal{H}} \mid \mathcal{Q}) \\
&\leq \frac{n-k-2b-r-t+1}{n-2b-r} (n-2b-r) H(A^i_h \mid \mathcal{Q}) \ .
\end{aligned}$$

Since the adversaries could otherwise be easily identified, we can assume that the answers of the adversarial servers are of the same entropy as the nonadversarial answers.

We conclude that in this setting the PIR rate is bounded by

$$\begin{aligned}\frac{H(X^i)}{\sum_{j=1}^n H(A_j^i)} &= \frac{H(X^i)}{n \cdot H(A_h^i \mid \mathcal{Q})} \\ &\leq \frac{H(X^i)}{n} \cdot \frac{n-k-2b-r-t+1}{H(X^i)} \\ &= \frac{n-k-2b-r-t+1}{n} \ .\end{aligned} \tag{8.14}$$

□

Remark 8.4. *Similar to Conjectures 8.1 and 8.2, the denominator in the rate expression in [TGK+19] is $n-r$ instead of n (see also Remark 8.1) because the nonresponsive servers are not included in the calculation of the download cost. Here, we also include the nonresponsive servers in the calculation, but note that this can be modified by changing the upper limit of the sum in Eq.* (8.14) *to $n-r$.*

Finally, we derive the secrecy rate of TBSPIR, i.e., the minimal amount of shared randomness required by the servers for symmetry to be achievable. We combine the proofs of [WS17a, Theorem 7] and [WS17b, Theorem 1].

Theorem 8.4. *The secrecy rate of a linear TBSPIR scheme from $[n,k]$ MDS-coded storage fulfills*

$$\rho \geq \frac{k+t-1}{n-k-t-2b-r+1} \ ,$$

if the randomness is additive as defined in Definition 8.8 or, for $t=1$, as in Definition 8.7.

Proof. Let $\mathcal{H} \subset [n]$ and $\mathcal{N} \subset \mathcal{H}$ be sets of honest, responsive servers with $|\mathcal{H}| = n-2b-r$ and $|\mathcal{N}| = k+t-1$. First, observe that

$$\begin{aligned}H(A_{\mathcal{H}}^i \mid \mathcal{Q}) &= H(X^i) + H(A_{\mathcal{H}}^i \mid X^i, Q) \\ &\geq H(X^i) + H(A_{\mathcal{N}}^i \mid X^i, Q) \\ &\geq H(X^i) + H(A_{\mathcal{N}}^i \mid Q).\end{aligned}$$

Averaging over all sets $\mathcal{N} \subset \mathcal{H}$ with $|\mathcal{N}| = k+t-1$ we get

$$H(A_{\mathcal{H}}^i \mid \mathcal{Q}) \geq H(X^i) + \frac{k+t-1}{n-2b-r} H(A_{\mathcal{H}}^i \mid Q). \tag{8.15}$$

Let $\mathcal{H} \subset [n]$ and $\mathcal{N} \subset \mathcal{H}$ be sets of honest, responsive servers with $|\mathcal{H}| = n-2b-r$

and $|\mathcal{N}| = k + t - 1$. By server privacy,

$$\begin{aligned}
0 &= I(X^{[m]\setminus i}; A_{\mathcal{H}} \mid \mathcal{Q}) \\
&= H(X^{[m]\setminus i} \mid \mathcal{Q}) - H(X^{[m]\setminus i} \mid A^i_{\mathcal{H}}, \mathcal{Q}) \\
&= H(X^{[m]\setminus i} \mid X^i, \mathcal{Q}) - H(X^{[m]\setminus i} \mid A^i_{\mathcal{H}}, X^i, \mathcal{Q}) \\
&= I(X^{[m]\setminus i}; A^i_{\mathcal{H}} \mid \mathcal{Q}, X^i) \\
&\geq I(X^{[m]\setminus i}; A^i_{\mathcal{N}} \mid \mathcal{Q}, X^i) \\
&= H(A^i_{\mathcal{N}} \mid X^i, \mathcal{Q}) - H(A^i_{\mathcal{N}} \mid X^{[m]}, \mathcal{Q}) + H(A^i_{\mathcal{N}} \mid S, X^{[m]}, \mathcal{Q}) \\
&= H(A^i_{\mathcal{N}} \mid X^i, \mathcal{Q}) - I(S; A^i_{\mathcal{N}} \mid X^{[m]}, \mathcal{Q}) \\
&\geq H(A^i_{\mathcal{N}} \mid X^i, Q^i_{\mathcal{N}}, \mathcal{Q}) - H(S) \\
&= H(A^i_{\mathcal{N}} \mid Q^i_{\mathcal{N}}) - H(S) \\
&\geq H(A^i_{\mathcal{N}} \mid \mathcal{Q}) - H(S)\,.
\end{aligned}$$

Averaging over all sets $\mathcal{N}$, we get by Eq. (8.15) that

$$\begin{aligned}
H(S) &\geq \frac{1}{\binom{n-2b-r}{k+t-1}} \sum_{\substack{\mathcal{N}\subset\mathcal{H} \\ |\mathcal{N}|=k+t-1}} H(A^i_{\mathcal{N}} \mid \mathcal{Q}) \\
&\geq \frac{k+t-1}{n-2b-r} H(A^i_{\mathcal{H}} \mid \mathcal{Q}) \\
&\geq \frac{k+t-1}{n-k-t-2b-r+1} H(X^i)\ ,
\end{aligned}$$

Thus, the bound on the secrecy rate is given by

$$\rho = \frac{H(S)}{H(X^i)} \geq \frac{k+t-1}{n-k-t-2b-r+1}\ .$$

□

8.6 Strongly-Linear PIR Capacity

We have seen that, for a symmetric linear scheme as in Theorem 8.3 and regardless of the number of files, the rate cannot be larger than that obtained by the scheme in [TGK⁺19], a generalization of the star-product scheme of [FGHK17]. Further, Theorem 8.2 shows that as the number of files grows, the rate of the star product scheme in [FGHK17] approaches the full support-rank capacity. We will now show that, under stronger linearity assumptions, this is also true for a finite number of files and without assuming server privacy. In essence, we define a PIR scheme to be *strongly linear* if all interference cancellation is linear and deterministic, and where every computation uses only one response symbol from each server. This is a highly natural assumption, that also has practical implications as it allows for decoding at the user to occur without delay, even when queries are sent sequentially. However, the assumption is not true for schemes achieving the capacity for a finite number of files, such as those in [SJ18b; BU18].

Definition 8.9 (Strongly Linear PIR)**.** *We say that a linear PIR scheme is* strongly linear *if each symbol of the desired file is obtained as a deterministic linear function over $\mathbb{F}$ of a response vector consisting of one response symbol from each server*

$$\Big(\mathbf{A}[:,(j-1)\beta+s]\Big)_{j\in[n]} ,$$

for some $s \in [\beta]$. Specifically, the reconstruction function does not depend on the randomness used to produce the queries.

From a practical point of view $\Big(\mathbf{A}[:,(j-1)\beta+s]\Big)_{j\in[n]}$ can be considered the response obtained in the s-th iteration of the PIR scheme.

Remark 8.5. *Note that a full support-rank PIR scheme does not have to be strongly linear. However, the rate of every optimal strongly linear scheme is upper bounded by the rate of the star product scheme [FGHK17], which agrees with the asymptotic capacity of a full support-rank PIR scheme with corresponding parameters. This result is proved in Theorem 8.5. Hence, a full support-rank scheme can always be replaced by a strongly linear scheme (e.g., a star product scheme) without a loss in the asymptotic rate.*

Remark 8.6. *We would like to emphasize that strongly linear schemes form a very relevant and practical case, namely the respective capacity result is known to be achievable [FGHK17; TGK+19] by a small field size $q \geq n$, which is that of a GRS code. Moreover, the subpacketization level is independent of m and is (at most) quadratic in n [FGHK17, Eq. (17)]. This is in contrast to the schemes in [SJ18b; SJ17; BU18; ZX18], where each file is assumed to be subdivided into a number of packets that grows exponentially with the number of files m. It was shown in [ZX18] that an exponential (in m) number of packets per file is necessary for a PIR scheme with optimal download rate, under the assumption that all servers respond to the queries and the responses have the same size. In [ZTSL20] a scheme was presented that achieves the capacity with only $O(n)$ packets by making a weaker assumption on the size of the responses than in [ZX18].*

Lemma 8.9. *Consider a strongly linear PIR scheme from a linear storage code $\mathcal{C}$, and fix an index $s \in [\beta]$. For all $l \in [m]$, let $\mathcal{D}^{i,l} \subseteq \mathbb{F}^n$ be the linear span of the row vectors that can occur as the l-th row of the s-th iteration of a query matrix Q^i, i.e.,*

$$\mathcal{D}^{i,l} = \Big\langle \{\mathbf{q}[l,\{s,\beta+s,\ldots,(n-1)\beta+s\}] \mid \mathbf{q} \in \operatorname{supp}(Q^i)\} \Big\rangle_{\mathsf{row}} .$$

Then the rate of the PIR scheme is at most

$$1 - \frac{\dim(\mathcal{C} \star (\sum_{l \notin \psi_\alpha(i)} \mathcal{D}^{i,l}))}{\dim(\mathcal{C} \star (\sum_l \mathcal{D}^{i,l}))} \ .$$

If the download from each server is of the same size, then the PIR rate is at most

$$\frac{\dim(\mathcal{C} \star (\sum_j \mathcal{D}^{i,j})) - \dim(\mathcal{C} \star (\sum_{l \notin \psi_\alpha(i)} \mathcal{D}^{i,j}))}{n} \ .$$

Proof. By Eq. (8.9) the responses in a linear PIR scheme as in Definition 8.1 can be described as the sum of the star product (Hadamard product) of rows of the query matrix and rows of the storage by

$$\begin{aligned}\Big(\mathbf{A}[:,(j-1)\beta+s]\Big)_{j\in[n]} &= \sum_{l=1}^{\alpha m} \Big(\mathbf{Y}[l,(j-1)\beta+s]\Big)_{j\in[n]} \star \mathbf{Q}^i[l,(j-1)\beta+s]\Big)_{j\in[n]}\Big) \\ &\in \sum_{l=1}^{\alpha m} \mathcal{D}^{i,l} \star \mathcal{C} \ .\end{aligned}$$

Let

$$\Phi : \Big(\mathbf{A}[:,(j-1)\beta+s]\Big)_{j\in[n]} \mapsto \mathbf{x} \in \mathbb{F}^\gamma$$

be the deterministic map from the responses in iteration s to γ coordinates of the desired file $\mathbf{X}$. Then for each $l \notin \psi_\alpha(i)$, Φ must be constant on each coset of $\mathcal{D}^{i,l} \star \mathcal{C}$, because otherwise changing the query matrix or the l-th row of $\mathbf{Y}$ would affect the value of $\Phi\Big(\Big(\mathbf{A}[:,(j-1)\beta+s]\Big)_{j\in[n]}\Big)$. As this holds for every $l \neq i$, Φ must be constant on each coset of $\sum_{j\neq i} \mathcal{D}^{i,j} \star \mathcal{C}$. Thus, the dimension of the range of Φ is

$$\begin{aligned}\gamma &= \dim\Big(\sum_j \mathcal{D}^{i,j} \star \mathcal{C}\Big) - \dim\Big(\ker(\Phi)\Big) \\ &\leq \dim\Big(\sum_j \mathcal{D}^{i,j} \star \mathcal{C}\Big) - \dim\Big(\sum_{j\notin\psi_\alpha(i)} \mathcal{D}^{i,j} \star \mathcal{C}\Big) \ .\end{aligned}$$

The n symbols $\Big(\mathbf{A}[:,(j-1)\beta+s]\Big)_{j\in[n]}$ can be reconstructed from the responses of $\dim\Big(\sum_j \mathcal{D}^{i,j} \star \mathcal{C}\Big)$ servers, or from n servers if we require to download equally much from each server. Dividing the number $|\mathcal{I}|$ of downloaded q-ary symbols from the desired file by the number of q-ary symbols in $\Big(\mathbf{A}[:,(j-1)\beta+s]\Big)_{j\in[n]}$, we get the claimed bounds on the PIR rate. This concludes the proof. □

For simplicity, we only consider schemes downloading the same number of symbols from all the servers in each iteration.

Before proceeding to the proof of the capacity of MDS-coded TBPIR under the assumption of strong linearity, we briefly recapitulate the star product PIR scheme of [FGHK17; TGK$^+$19], which will be used to show the achievability of the provided upper bound on the PIR rate. Consider a DSS storing m files encoded with an $[n, k]$ MDS storage code $\mathcal{C}$ and a user looking to retrieve file i privately in the presence of up to t colluding servers. For simplicity, we assume $n = 2k+t+2b+r-1$ and $\alpha = 1$ here, as this allows the recovery of the file in one iteration[10]. Further, for ease of notation, we only consider the case of all servers being responsive, i.e., $r = 0$. The extension to the case of nonresponsive servers is trivial. The star product scheme consists of the following steps:

1. The user chooses a *query code* $\mathcal{D}_Q$ with $d_{\mathcal{D}_Q^\perp} \geq t+1$, where $d_{\mathcal{D}_Q^\perp}$ denotes the minimum distance of the dual code $\mathcal{D}_Q^\perp$. From this code, she generates a matrix $\mathbf{D} \in F^{m\times n}$ whose m rows are codewords of $\mathcal{D}_Q$ chosen i.i.d. at random[11].

2. The *query matrix* is given by
$$\mathbf{Q}^i = \mathbf{D} + \mathbf{E} \ ,$$
where $\mathbf{E}$ is all-zero, except for the i-th row $\mathbf{E}[i,:]$, which is chosen to be the basis of an $[n, 1]$ code[12] $\mathcal{E}$.

3. The user sends the j-th column of $\mathbf{Q}^i$ to the j-th server. The server replies with
$$\mathbf{A}^i[1,j] = \left\langle \mathbf{Q}^i[:,j], \mathbf{Y}[:,j] \right\rangle + \mathbf{z}[1,j] \ ,$$
where $\mathbf{z}[1,j] = 0$ if the server is honest and arbitrary if the server is adversarial ($\mathbf{z}$ can be thought of as the received error vector).

[10] We would like to emphasize that the scheme discussed here is a special case of the star product PIR scheme of [FGHK17; TGK$^+$19], with parameters chosen for an illustrative purpose. The full scheme is not limited to this specific choice of n. For more details, see [FGHK17; TGK$^+$19].

[11] The fact that $d_{\mathcal{D}_Q^\perp} \geq t+1$ implies that any t positions in a codeword of $\mathcal{D}_Q$ are an information set. Hence, by choosing random codewords of $\mathcal{D}_Q$ as the rows, any t columns of $\mathbf{D}$ are i.i.d. distributed over $\mathbb{F}^{m\times t}$.

[12] Here and for the general scheme it is convenient to view this as a code instead of a vector. Note that for a different choice of n and α the dimension of this code could be larger than 1.

4. By Eq. (8.9), the user receives

$$\begin{aligned}\mathbf{A}^i &= \Big(\sum_{l\in[m]} \mathbf{Y}[l,:]\star\mathbf{Q}^i[l,:]\Big)+\mathbf{z}\\ &= \Big(\sum_{l\in[m]} \mathbf{Y}[l,:]\star\Big(\mathbf{D}[l,:]+\mathbf{E}[l,:]\Big)\Big)+\mathbf{z}\\ &= \underbrace{\Big(\sum_{l\in[m]} \mathbf{Y}[l,:]\star\mathbf{D}[l,:]\Big)}_{\in\mathcal{C}\star\mathcal{D}_Q}+\underbrace{\Big(\mathbf{Y}[i,:]\star\mathbf{E}[i,:]\Big)}_{\in\mathcal{C}\star\mathcal{E}}+\mathbf{z}\ .\end{aligned}$$

Recall that the Hamming weight of $\mathbf{z}$ is at most b, the number of adversarial servers. Hence, if the code $\mathcal{C}\star\mathcal{D}_Q+\mathcal{C}\star\mathcal{E}$ is of distance $d_{\mathcal{C}\star\mathcal{D}_Q+\mathcal{C}\star\mathcal{E}}\geq 2b+1$, the errors can be decoded and the user obtains

$$\underbrace{\Big(\sum_{l\in[m]} \mathbf{Y}[l,:]\star\mathbf{D}[l,:]\Big)}_{\in\mathcal{C}\star\mathcal{D}_Q}+\underbrace{\Big(\mathbf{Y}[i,:]\star\mathbf{E}[i,:]\Big)}_{\in\mathcal{C}\star\mathcal{E}}\ .$$

As $\mathbf{E}$ is chosen by the user, we only require that the codes $\mathcal{C}\star\mathcal{D}_Q$ and $\mathcal{C}\star\mathcal{E}$ intersect trivially to recover the vector $\mathbf{Y}[i,:]\star\mathbf{E}[i,:]\in\mathcal{C}\star\mathcal{E}$. Finally, the file X^i can be recovered from this vector, given that $\mathcal{C}\star\mathcal{E}$ is of dimension k.

It remains to determine codes $\mathcal{C}$, $\mathcal{D}_Q$, and $\mathcal{E}$ that fulfill the required properties for the given n. Conveniently, it has been shown [FGHK17; TGK$^+$19] that GRS codes (see Definition 2.3) provide such codes, however, these details are beyond the scope of this short summary.

We are now ready to show that any strongly linear scheme can be replaced by a star product scheme for the same privacy model, without losing in the PIR rate.

Theorem 8.5 (Capacity of Strongly Linear PIR)**.** *The capacity of strongly linear MDS-TBPIR, i.e., strongly linear PIR from $[n,k]$ MDS-coded storage with t colluding, b adversarial, and r nonresponsive servers, is*

$$C^{\mathsf{MDS}}_{\mathsf{SL-TBPIR}}=1-\frac{k+t+2b+r-1}{n}$$

for any number of files m.

Proof. Consider an arbitrary strongly linear PIR scheme. Like in Lemma 8.9, fix an iteration $s\in[\beta]$ and define

$$\mathcal{D}^{i,l}=\Big\langle\{\mathbf{q}[l,\{s,\beta+s,\ldots,(n-1)\beta+s\}]\mid\mathbf{q}\in\operatorname{supp}(Q^i)\}\Big\rangle$$

for $l \in [n]$, and $\mathcal{D} = \sum_{l \neq i} \mathcal{D}^{i,l}$. Let $\mathbf{E} \in \mathbb{F}^{\alpha m \times n}$ be a matrix such that $\mathbf{E}[\psi_\alpha(i), :]$ is an arbitrary realisation of $\mathbf{Q}^i[\psi_\alpha(i), \{s, s+\beta, \ldots, s+(n-1)\beta\}]$, and all other entries are zero. Let $\mathbf{D} \in \mathbb{F}^{\alpha m \times n}$ be a random matrix whose rows are selected uniformly at random from $\mathcal{D}$.

Now consider the star product scheme with query matrix $\mathbf{D} + \mathbf{E}$. This scheme has a set of feasible query matrices that is more restrictive in the row of the desired file, but less restrictive in the rows of the unwanted files, than the original strongly linear scheme. Thus, whatever privacy constraints were satisfied by the original scheme, including robustness against nonresponsive and adversarial servers, are also respected by the star product scheme. By design, all symbols that were decoded in the s-th iteration of the strongly linear scheme are also decoded in the star product scheme. Moreover, by construction the rate of the star product scheme is

$$1 - \frac{\dim(\mathcal{D})}{\dim(\sum_l \mathcal{D}^{i,l})} ,$$

which is at least the rate of the original strongly linear scheme by Lemma 8.9. So the rate of any strongly linear scheme is bounded from above by the rate of a star product scheme with the same privacy constraints, which is in turn bounded by $1 - \frac{k+2b+r+t-1}{n}$ as shown in [TGK+19]. The paper also presents a scheme achieving this bound via the star-product construction. □

Note that the capacity of strongly linear PIR is *independent* of the number of files. Hence, the above theorem also yields a proof for Conjecture 8.1 in the strongly linear case. The capacity of a strongly linear scheme also matches the asymptotic rate of Conjecture 8.3, hence proving the asymptotic expression for such schemes.

Remark 8.7. *Here, to simplify the notation, we have assumed that all the servers respond with equal size responses. However, by loosening this assumption, improvements for finite m are possible, along the same lines as in [ZTSL20]. The proof of the above theorem shows that, among strongly linear schemes as in Definition 8.9, the star product scheme [FGHK17; TGK+19] is optimal if the responses are of* equal size.

8.7 Summary and Open Problems

This chapter introduced the practical notions of full support-rank PIR and strongly linear PIR. The capacity of MDS-coded, linear, full support-rank PIR with colluding servers was proved as well as the capacity of symmetric linear PIR with MDS-coded, colluding, adversarial, and nonresponsive servers for the case of matched/additive randomness.

The results on full support-rank PIR are a significant step towards the general proof of the capacity of PIR from MDS-coded storage with colluding servers. Meanwhile, the results on strongly linear PIR bear high practical interest in that these schemes allow for small field sizes and low subpacketization levels. These simpler schemes also achieve the same asymptotic capacity as the full support-rank schemes. The main open problem that remains is proving the capacity of (linear) PIR with MDS-coded and colluding servers without the assumption of full support-rank. As explained in Section 8.1, the presented definition of full support-rank PIR isolates a property required for a scheme to achieve this capacity for general linear, MDS-coded PIR, namely for the restrictions of its queries to *not* be of full support-rank. Thereby, the results in this chapter provide a good starting point for both giving upper bounds on the PIR rate and constructing achieving schemes.

Another open problem is determining the capacity of TPIR for transitive storage codes, along the lines of [FGH$^+$19], by adapting the proofs of Lemma 8.3 Lemma 8.4 accordingly.

9 Other Results on PIR

This chapter briefly summarizes a selection of other works on PIR. For more details the interested reader is referred to the respective publications.

9.1 Private Streaming with Convolutional Codes

This abstract summarizes the results of [HFWH20] published in the IEEE Transactions on Information Theory. *In part, the results have been published in the proceedings of the* 2018 IEEE Information Theory Workshop (ITW) *[HFWH18].*

Recently, information-theoretic PIR from coded storage systems [CGKS95; SJ17] has gained a lot of attention. In this setting, the goal of the user is to retrieve a file from a database without revealing its index. However, in applications such as video streaming, the user is commonly interested in decoding parts of the file while the retrieval is still on-going. Such applications require low latency decoding of the received data blocks and it has been shown that, under such constraints, convolutional codes perform well [BKTA13; KB16] for different channels.

To fulfill this requirement in a private setting, this work studies the problem of *private streaming.* To this end, the star-product scheme of [FGHK17; TGK$^+$19] is adapted by introducing memory into the retrieval process. Specifically, a scheme for streaming from a database encoded with an RS code is proposed, where the user designs the queries such that the set of all replies received by from the servers has a block convolutional structure. Two schemes are proposed and shown to improve the resistance against erroneous decoding under two channel models related to nonresponsive (block erasure channel) and adversarial (AWGN-channel) servers, both in the baseline case as well as with colluding servers. The schemes can operate on the same database and the user can adapt the queries according to the current channel conditions.

The achieved PIR rates are derived and for the block erasure scheme shown to be either asymptotically optimal, or, for cases where the capacity is unknown, shown to coincide with conjectures on the asymptotic capacity. For the adversarial server model, the introduced scheme is shown to outperform the alternative scheme of downloading stripes of the desired file separately without memory.

9.2 Computational Code-Based Single-Server Private Information Retrieval

This abstract summarizes the results of the work [HHW20] published in the proceedings of the 2020 IEEE International Symposium on Information Theory (ISIT)*.*

Considerable attention has been directed towards the problem of private information retrieval in recent years. Chor et al.'s seminal paper [CGKS95] showed that perfect information-theoretic privacy cannot achieved with a single server. This lead to two main research directions—information-theoretic privacy from more than one server and computational privacy from a single server. While the former allows for high rates and computationally simple schemes, the underlying assumption that not all servers collude severely limits its use cases. Single-server PIR schemes, such as those proposed in [YKPB12; AG07; KLL+15; ABFK16; LP17; GH19; ABFK16; ACLS18] do not require this assumption, but their practicality is limited by their computational complexity on the server side.

This work proposes the first known computational PIR scheme based on codes, which can be seen as a counterpart to the lattice-based scheme of [AG07] along the same lines as code-based and lattice-based cryptography are connected in general. The query to the sever is a matrix whose rows contain corrupted codewords of a secret code, i.e., each row is similar to a ciphertext in the well-known McEliece cryptosystem. The server then responds with the scalar product of the query matrix and the files and the user can recover the requested file by erasure decoding. Depending on the parameters, the achieved PIR rates are comparable to the existing computational PIR schemes of [YKPB12; AG07]. The complexity, which is the bottleneck of current computational schemes, benefits from the fact that all calculations can be conducted over binary extension fields, which is advantageous for implementation.

Note that this computational PIR scheme has recently been broken for all relevant parameters. For details see [BL21].

9.3 Quantum Private Information Retrieval from Coded and Colluding Servers

This abstract summarizes the results of [AHPH20a] published in the IEEE Journal on Selected Areas in Information Theory *and the results of [ASH$^+$22], which has been accepted for publication in the* IEEE Journal on Selected Areas in Communications. *In part, the results of [AHPH20a] have been published in the proceedings of the* 2020 IEEE International Symposium on Information Theory (ISIT) *[AHPH20b] and parts of the results of [ASH$^+$22] in the proceedings of the* 2021 IEEE International Symposium on Information Theory (ISIT) *[AHPH21].*

In the classical PIR setup, as introduced by Chor et al. [CGKS95], a user wants to retrieve a file from a database or a distributed storage system (DSS) without revealing the file identity to the servers holding the data. To achieve information-theoretic privacy, the user poses queries to a number of servers, which the (set of colluding) servers can read without learning any information. This problem has also been considered in the context of quantum communication [KDW04; Gal11; GLM08]. More recently, Song et al. [SH20b] proposed a PIR scheme based on the properties of quantum communication. Under this model the user poses classical queries to the servers, which respond with quantum systems. By exploiting the quantum theoretic guarantee that such a system can only be measured once, the authors are able to show that this scheme achieves a rate of one, i.e., induces no communication overhead to ensure privacy. In [SH19] the same authors considered the case of replicated storage where all but one servers collude, showing that the retrieval rate can be essentially doubled compared to the classical setting.

In the work [AHPH20b] the quantum PIR protocol of [SH19] is generalized to allow for the private retrieval of a file from a storage system encoded with an MDS code. The proposed protocol works for any linear MDS code of length n and dimension k while tolerating up to t colluding servers, with the restriction that $t = n-k$ (note that $t = n-1$ and $k = 1$ corresponds to the setting of [SH19]). Similar to the quantum PIR schemes for replicated storage, the rates achieved are approximately double of the classical counterparts. Further, it is demonstrated how the protocol can be adapted to achieve significantly higher retrieval rates from DSSs encoded with an LRC with disjoint repair groups, each of which is an MDS code.

The work [ASH$^+$22] relaxes the parameter restrictions to allow for high-rate quantum PIR for any length n, dimension k, and collusion resistance t with $t \leq n-k$. For this setting the capacity of a subclass of quantum PIR schemes is proved. To show the achievability a new quantum PIR scheme is introduced, which combines the quantum PIR scheme of [SH20a] for replicated storage with the classical PIR scheme of [FGHK17] for coded storage, both with collusion, through the use of (weakly) self-dual GRS codes.

10
Conclusion and Outlook

This work investigated different aspects of algebraic coding theory with a focus on concepts related the distributed data storage.

Part I considered codes with different locality properties, starting with MR codes for grid-like topologies. While the understanding of (MR) LRCs has improved significantly in recent years, this class of codes still offers exciting research possibilities, as highlighted by the presented negative result on the set of correctable erasure patterns. This interest is not purely theoretical though, as MR codes are, by definition, the best choice to lessen the effects of the main downside of using codes with locality in distributed storage systems—their reduced overall erasure correction capability compared to MDS codes. The complexity of storage systems can be expected to increase with the growth of online services and the presented results take a step towards accommodating such more complex requirements in a coding theoretic framework.

The next challenged addressed in this part of the work is the combination of codes with locality and codes that offer bandwidth-efficient node repair, i.e., regenerating codes. Both concepts are motivated by distributed storage applications and the presented constructions of locally and globally regenerating PMDS codes combine them to provide classes of codes which are optimal in terms of their repair bandwidth and erasure correction capability, given the locality constraints. In this regard, an opportunity for future research is the construction more practical classes of codes, in particular regarding the subpacketization.

The final chapter in this first part highlighted the versatility of locality properties beyond the application in distributed storage. By exploiting the particularly strong locality and availability of lifted affine-invariant codes, a new, simple BD decoder for this class of codes was introduced. Further, the correction capability of this class of codes in the high-error regime was proved.

In Part II the first known bound on the success probability of decoding interleaved alternant codes was derived. Among these subcodes of RS codes are some of the most popular classes of codes over small alphabets, such as BCH and Goppa codes. While the remaining gap to the newly introduced lower bound leaves room for improvement, the utilized approach has potential to lead to even stronger bounds by means of a better understanding of the distribution of dimensions among alternant codes.

Finally, Part III addressed the problem of PIR in a DSS encoded with an MDS code. Some progress on the long-standing open problem of determining the capacity in this setting was made by introducing to new notions of PIR and proving their capacities. Aside from the implications for the considered class of schemes, these results also enhance the understanding of the requirements for any scheme to improve upon their rate.

Related Publications by the Author

[AHPH20a] Matteo Allaix, Lukas Holzbaur, Tefjol Pllaha, and Camilla Hollanti. "Quantum Private Information Retrieval from Coded and Colluding Servers". In: *IEEE Journal on Selected Areas in Information Theory* 1.2 (2020), pp. 599–610.

[AHPH20b] Matteo Allaix, Lukas Holzbaur, Tefjol Pllaha, and Camilla Hollanti. "Quantum Private Information Retrieval from MDS-coded and Colluding Servers". In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2020, pp. 1059–1064.

[AHPH21] Matteo Allaix, Lukas Holzbaur, Tefjol Pllaha, and Camilla Hollanti. "High-Rate Quantum Private Information Retrieval with Weakly Self-Dual Star Product Codes". In: *IEEE International Symposium on Information Theory (ISIT)*. 2021, pp. 1046–1051.

[ASH$^+$22] Matteo Allaix, Seunghoan Song, Lukas Holzbaur, Tefjol Pllaha, Masahito Hayashi, and Camilla Hollanti. "On the Capacity of Quantum Private Information Retrieval from MDS-Coded and Colluding Servers". In: *IEEE Journal on Selected Areas in Communications* (2022), (Early Access). DOI: 10.1109/JSAC.2022.3142363.

[HFH19] Lukas Holzbaur, Ragnar Freij-Hollanti, and Camilla Hollanti. "On the Capacity of Private Information Retrieval from Coded, Colluding, and Adversarial Servers". In: *IEEE Information Theory Workshop (ITW)*. IEEE. 2019, pp. 1–5.

[HFLH22] Lukas Holzbaur, Ragnar Freij-Hollanti, Jie Li, and Camilla Hollanti. "Toward the Capacity of Private Information Retrieval From Coded and Colluding Servers". In: *IEEE Transactions on Information Theory* 68.1 (2022), pp. 517–537.

[HFWH18] Lukas Holzbaur, Ragnar Freij-Hollanti, Antonia Wachter-Zeh, and Camilla Hollanti. "Private Streaming with Convolutional Codes". In: *IEEE Information Theory Workshop (ITW)*. IEEE. 2018, pp. 1–5.

[HFWH20] Lukas Holzbaur, Ragnar Freij-Hollanti, Antonia Wachter-Zeh, and Camilla Hollanti. "Private Streaming with Convolutional Codes". In: *IEEE Transactions on Information Theory* 66.4 (2020), pp. 2417–2429.

[HHW20] Lukas Holzbaur, Camilla Hollanti, and Antonia Wachter-Zeh. “Computational Code-Based Single-Server Private Information Retrieval”. In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2020, pp. 1065–1070.

[HKFW20] Lukas Holzbaur, Stanislav Kruglik, Alexey Frolov, and Antonia Wachter-Zeh. “Secrecy and Accessibility in Distributed Storage”. In: *IEEE Global Communications Conference (GLOBECOM)*. 2020, 1–6.

[HKFW21] Lukas Holzbaur, Stanislav Kruglik, Alexey Frolov, and Antonia Wachter-Zeh. “Secure Codes With Accessibility for Distributed Storage”. In: *IEEE Transactions on Information Forensics and Security* 16 (2021), pp. 5326–5337.

[HLN+20] Lukas Holzbaur, Hedongliang Liu, Alessandro Neri, Sven Puchinger, Johan Rosenkilde, Vladimir Sidorenko, and Antonia Wachter-Zeh. “Success Probability of Decoding Interleaved Alternant Codes”. In: *IEEE Information Theory Workshop (ITW)*. IEEE. 2020, pp. 1–5.

[HLN+21] Lukas Holzbaur, Hedongliang Liu, Alessandro Neri, Sven Puchinger, Johan Rosenkilde, Vladimir Sidorenko, and Antonia Wachter-Zeh. “Decoding of Interleaved Alternant Codes”. In: *IEEE Transactions on Information Theory* 67.12 (2021), pp. 8016–8033.

[HLPW19] Lukas Holzbaur, Hedongliang Liu, Sven Puchinger, and Antonia Wachter-Zeh. “On Decoding and Applications of Interleaved Goppa Codes”. In: *IEEE International Symposium on Information Theory (ISIT)*. 2019, pp. 1887–1891.

[HP20] Lukas Holzbaur and Nikita Polyanskii. “Decoding of Lifted Affine-Invariant Codes”. In: *IEEE Information Theory Workshop (ITW)*. IEEE. 2020, pp. 1–5.

[HPP+20] Lukas Holzbaur, Rina Polyanskaya, Nikita Polyanskii, Ilya Vorobyev, and Eitan Yaakobi. “On Lifted Multiplicity Codes”. In: *IEEE Information Theory Workshop (ITW)*. IEEE. 2020, pp. 1–5.

[HPP+21] Lukas Holzbaur, Rina Polyanskaya, Nikita Polyanskii, Ilya Vorobyev, and Eitan Yaakobi. “Lifted Reed-Solomon Codes and Lifted Multiplicity Codes”. In: *IEEE Transactions on Information Theory* 67.12 (2021), pp. 8051–8069.

[HPPV20] Lukas Holzbaur, Rina Polyanskaya, Nikita Polyanskii, and Ilya Vorobyev. “Lifted Reed-Solomon Codes with Application to Batch Codes”. In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2020, pp. 634–639.

[HPW19] Lukas Holzbaur, Sven Puchinger, and Antonia Wachter-Zeh. “On Error Decoding of Locally Repairable and Partial MDS Codes”. In: *IEEE Information Theory Workshop (ITW)*. IEEE. 2019, pp. 1–5.

[HPW21] Lukas Holzbaur, Sven Puchinger, and Antonia Wachter-Zeh. “Error Decoding of Locally Repairable and Partial MDS Codes”. In: *IEEE Transactions on Information Theory* 67.3 (2021), pp. 1571–1595.

[HPYW20] Lukas Holzbaur, Sven Puchinger, Eitan Yaakobi, and Antonia Wachter-Zeh. “Partial MDS Codes with Local Regeneration”. In: *IEEE International Symposium on Information Theory (ISIT)*. 2020, pp. 628–633.

[HPYW21] Lukas Holzbaur, Sven Puchinger, Eitan Yaakobi, and Antonia Wachter-Zeh. “Correctable Erasure Patterns in Product Topologies”. In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2021, pp. 2054–2059.

[HPYWZ21] Lukas Holzbaur, Sven Puchinger, Eitan Yaakobi, and Antonia Wachter-Zeh. “Partial MDS Codes With Regeneration”. In: *IEEE Transactions on Information Theory* 67.10 (2021), pp. 6425–6441.

[HW18] Lukas Holzbaur and Antonia Wachter-Zeh. “List Decoding of Locally Repairable Codes”. In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2018, pp. 1331–1335.

Bibliography

[ABFK16] Carlos Aguilar-Melchor, Joris Barrier, Laurent Fousse, and Marc-Olivier Killijian. "XPIR: Private information retrieval for everyone". In: *Proceedings on Privacy Enhancing Technologies* 2016.2 (2016), pp. 155–174.

[ACLS18] Sebastian Angel, Hao Chen, Kim Laine, and Srinath Setty. "PIR with compressed queries and amortized query processing". In: *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE. 2018, pp. 962–979.

[AG07] Carlos Aguilar-Melchor and Philippe Gaborit. "A lattice-based computationally-efficient private information retrieval protocol". In: *Cryptol. ePrint Arch., Report* 446 (2007).

[AS03] Sanjeev Arora and Madhu Sudan. "Improved low-degree testing and its applications". In: *Combinatorica* 23.3 (2003), pp. 365–426.

[Bas65] Leonid Alexandrovich Bassalygo. "New upper bounds for error correcting codes". In: *Problemy Peredachi Informatsii* 1.4 (1965), pp. 41–44.

[Ber03] Thierry P. Berger. "Isometries for Rank Distance and Permutation Group of Gabidulin Codes". In: *IEEE Transactions on Information Theory* 49.11 (2003), pp. 3016–3019.

[Ber73] Elwyn Berlekamp. "Goppa codes". In: *IEEE Transactions on Information Theory* 19.5 (1973), pp. 590–592.

[BHH13] Mario Blaum, James Lee Hafner, and Steven Hetzler. "Partial-MDS codes and their application to RAID type of architectures". In: *IEEE Transactions on Information Theory* 59.7 (2013), pp. 4510–4519.

[BK15] S.B. Balaji and P. Vijay Kumar. "On partial maximally-recoverable and maximally-recoverable codes". In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2015, pp. 1881–1885.

[BKTA13] Ahmed Badr, Ashish Khisti, Wai-Tian Tan, and John Apostolopoulos. "Streaming codes for channels with burst and isolated erasures". In: *2013 Proceedings IEEE INFOCOM*. IEEE. 2013, pp. 2850–2858.

[BKY03] Daniel Bleichenbacher, Aggelos Kiayias, and Moti Yung. “Decoding of Interleaved Reed–Solomon Codes Over Noisy Data”. In: *International Colloquium on Automata, Languages, and Programming*. Springer. 2003, pp. 97–108.

[BL21] Sarah Bordage and Julien Lavauzelle. “On the privacy of a code-based single-server computational PIR scheme”. In: *Cryptography and Communications* (2021), pp. 1–8.

[BLP11] Daniel J. Bernstein, Tanja Lange, and Christiane Peters. “Wild McEliece”. In: *Selected Areas in Cryptography*. Springer Berlin Heidelberg, 2011, pp. 143–158.

[BMS04] Andrew Brown, Lorenz Minder, and Amin Shokrollahi. “Probabilistic Decoding of Interleaved RS-Codes on the q-Ary Symmetric Channel”. In: *IEEE International Symposium on Information Theory (ISIT)*. 2004, pp. 326–326.

[BMS05] Andrew Brown, Lorenz Minder, and Amin Shokrollahi. “Improved Decoding of Interleaved AG Codes”. In: *IMA International Conference on Cryptography and Coding*. Springer. 2005, pp. 37–46.

[BPSY16] Mario Blaum, James S Plank, Moshe Schwartz, and Eitan Yaakobi. “Construction of partial MDS and sector-disk codes with two global parity symbols”. In: *IEEE Transactions on Information Theory* 62.5 (2016), pp. 2673–2681.

[BRC60] Raj Chandra Bose and Dwijendra K Ray-Chaudhuri. “On a class of error correcting binary group codes”. In: *Information and Control* 3.1 (1960), pp. 68–79.

[BSGM$^+$11] Eli Ben-Sasson, Elena Grigorescu, Ghid Maatouk, Amir Shpilka, and Madhu Sudan. “On sums of locally testable affine invariant properties”. In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 2011, pp. 400–411.

[BU18] Karim Banawan and Sennur Ulukus. “The Capacity of Private Information Retrieval From Coded Databases”. In: *IEEE Transactions on Information Theory* 64.3 (2018), pp. 1945–1956. issn: 0018-9448.

[BU19] Karim Banawan and Sennur Ulukus. “The Capacity of Private Information Retrieval from Byzantine and Colluding Databases”. In: *IEEE Transactions on Information Theory* 65.2 (2019), pp. 1206–1219. issn: 0018-9448.

[Bu80] Tor Bu. “Partitions of a vector space”. In: *Discrete Mathematics* 31.1 (1980), pp. 79–83.

[CGKS95] Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan. "Private information retrieval". In: *Proceedings of IEEE 36th Annual Foundations of Computer Science*. IEEE. 1995, pp. 41–50.

[CH13] Henry Cohn and Nadia Heninger. "Approximate Common Divisors via Lattices". In: *The Open Book Series* 1.1 (2013), pp. 271–293.

[CHL07] Minghua Chen, Cheng Huang, and Jin Li. "On the maximally recoverable property for multi-protection group codes". In: *2007 IEEE International Symposium on Information Theory*. IEEE. 2007, pp. 486–490.

[CJM$^+$13] Viveck R. Cadambe, Syed Ali Jafar, Hamed Maleki, Kannan Ramchandran, and Changho Suh. "Asymptotic interference alignment for optimal repair of MDS codes in distributed storage". In: *IEEE Transactions on Information Theory* 59.5 (2013), pp. 2974–2987.

[CK16] Gokhan Calis and Onur Ozan Koyluoglu. "A general construction for PMDS codes". In: *IEEE Communications Letters* 21.3 (2016), pp. 452–455.

[CLS09] Hyojin Choi, Wei Liu, and Wonyong Sung. "VLSI implementation of BCH error correction for multilevel cell NAND flash memory". In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 18.5 (2009), pp. 843–847.

[CMCP17] Alain Couvreur, Irene Márquez-Corbella, and Ruud Pellikaan. "Cryptanalysis of McEliece Cryptosystem Based on Algebraic Geometry Codes and their Subcodes". In: *IEEE Transactions on Information Theory* 63.8 (2017), pp. 5404–5418.

[CMST20] Han Cai, Ying Miao, Moshe Schwartz, and Xiaohu Tang. "A Construction of Maximally Recoverable Codes with Order-Optimal Field Size". In: *arXiv preprint arXiv:2011.13606* (2020).

[CR20] Alain Couvreur and Hugues Randriambololona. "Algebraic Geometry Codes and some Applications". In: *arXiv preprint arXiv:2009.01281* (2020).

[CS03] Don Coppersmith and Madhu Sudan. "Reconstructing Curves in Three (and Higher) Dimensional Space from Noisy Data". In: *ACM Symposium on the Theory of Computing*. 2003.

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 1991. ISBN: 0-471-06259-6.

[CW90] Don Coppersmith and Shmuel Winograd. "Matrix Multiplication via Arithmetic Progressions". In: *Journal of Symbolic Computation* 9.3 (Mar. 1990), 251–280. ISSN: 0747–7171.

[DE19] Rafael G.L. D'Oliveira and Salim El Rouayheb. "One-shot PIR: Refinement and lifting". In: *IEEE Transactions on Information Theory* 66.4 (2019), pp. 2443–2455.

[Del75] Philippe Delsarte. "On subfield subcodes of modified Reed–Solomon codes (Corresp.)" In: *IEEE Transactions on Information Theory* 21.5 (Sept. 1975), pp. 575–576. ISSN: 1557-9654.

[Del78] Philippe Delsarte. "Bilinear Forms over a Finite Field with Applications to Coding Theory". In: *Journal of Combinatorial Theory, Series A* 25.3 (1978), pp. 226–241.

[DGMW70] Philippe Delsarte, Jean-Marie Goethals, and F Jessie Mac Williams. "On generalized Reed–Muller codes and their relatives". In: *Information and Control* 16.5 (1970), pp. 403–442.

[DGW$^+$10] Alexandros G. Dimakis, P. Brighten Godfrey, Yunnan Wu, Martin J. Wainwright, and Kannan Ramchandran. "Network coding for distributed storage systems". In: *IEEE Transactions on Information Theory* 56.9 (2010), pp. 4539–4551. ISSN: 0018-9448.

[Dum04] Ilya Dumer. "Recursive decoding and its performance for low-rate Reed–Muller codes". In: *IEEE Transactions on Information Theory* 50.5 (2004), pp. 811–823.

[EWZ18] Molka Elleuch, Antonia Wachter-Zeh, and Alexander Zeh. *A Public-Key Cryptosystem from Interleaved Goppa Codes.* 2018. arXiv: `1809.03024`. URL: `http://arxiv.org/abs/1809.03024`.

[FGHK17] Ragnar Freij-Hollanti, Oliver W. Gnilke, Camilla Hollanti, and David A. Karpuk. "Private Information Retrieval from Coded Databases with Colluding Servers". In: *SIAM Journal on Applied Algebra and Geometry* 1.1 (2017), pp. 647–664. ISSN: 2470-6566.

[FGH$^+$19] Ragnar Freij-Hollanti, Oliver W. Gnilke, Camilla Hollanti, Anna-Lena Horlemann-Trautmann, David Karpuk, and Ivo Kubjas. "t-Private Information Retrieval Schemes Using Transitive Codes". In: *IEEE Transactions on Information Theory* 65.4 (2019), pp. 2107–2118. ISSN: 0018-9448.

[Fik10] Andrew Fikes. *Storage Architecture and Challenges.* `https://cloud.google.com/files/storage_architecture_and_challenges.pdf`. Accessed: 28.04.2021. 2010.

[For65] George Forney. "On decoding BCH codes". In: *IEEE Transactions on Information Theory* 11.4 (1965), pp. 549–557.

[FT91] G.-L. Feng and Kenneth K. Tzeng. “A generalization of the Berlekamp-Massey algorithm for multisequence shift-register synthesis with applications to decoding cyclic codes”. In: *IEEE Transactions on Information Theory* 37.5 (1991), pp. 1274–1287.

[FVY15] Arman Fazeli, Alexander Vardy, and Eitan Yaakobi. “Codes for distributed PIR with low storage overhead”. In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2015, pp. 2852–2856.

[Gab85] Ernest Mukhamedovich Gabidulin. “Theory of Codes with Maximum Rank Distance”. In: *Problems of Information Transmission* 21.1 (1985), pp. 3–16.

[Gal11] François Le Gall. “Quantum private information retrieval with sublinear communication complexity”. In: *arXiv preprint arXiv:1107.5881* (2011).

[Gao03] Shuhong Gao. “A new algorithm for decoding Reed–Solomon codes”. In: *Communications, Information and Network Security*. Springer, 2003, pp. 55–68.

[GFV17] Sreechakra Goparaju, Arman Fazeli, and Alexander Vardy. “Minimum storage regenerating codes for all parameters”. In: *IEEE Transactions on Information Theory* 63.10 (2017), pp. 6318–6328.

[GG20] Sivakanth Gopi and Venkatesan Guruswami. “Improved Maximally Recoverable LRCs using Skew Polynomials”. In: *arXiv preprint arXiv:2012.07804* (2020).

[GH19] Craig Gentry and Shai Halevi. “Compressible FHE with applications to PIR”. In: *Theory of Cryptography Conference*. Springer. 2019, pp. 438–464.

[GHJY14] Parikshit Gopalan, Cheng Huang, Bob Jenkins, and Sergey Yekhanin. “Explicit Maximally Recoverable Codes With Locality”. In: *IEEE Transactions on Information Theory* 60.9 (2014), pp. 5245–5256. ISSN: 0018-9448.

[GHK$^+$17] Parikshit Gopalan, Guangda Hu, Swastik Kopparty, Shubhangi Saraf, Carol Wang, and Sergey Yekhanin. “Maximally recoverable codes for grid-like topologies”. In: *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2017, pp. 2092–2108.

[GHSY12] Parikshit Gopalan, Cheng Huang, Huseyin Simitci, and Sergey Yekhanin. “On the Locality of Codeword Symbols”. In: *IEEE Transactions on Information Theory* 58.11 (2012), pp. 6925–6934. ISSN: 0018-9448.

[GIKM00] Yael Gertner, Yuval Ishai, Eyal Kushilevitz, and Tal Malkin. "Protecting data privacy in private information retrieval schemes". In: *Journal of Computer and System Sciences* 60.3 (2000), pp. 592–629.

[GK16] Alan Guo and Swastik Kopparty. "List-decoding algorithms for lifted codes". In: *IEEE Transactions on Information Theory* 62.5 (2016), pp. 2719–2725.

[GKJS17] Danilo Gligoroski, Katina Kralevska, Rune E Jensen, and Per Simonsen. "Repair duality with locally repairable and locally regenerating codes". In: *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. IEEE. 2017, pp. 979–984.

[GKS13] Alan Guo, Swastik Kopparty, and Madhu Sudan. "New affine-invariant codes from lifting". In: *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*. 2013, pp. 529–540.

[GKZ08] Parikshit Gopalan, Adam R Klivans, and David Zuckerman. "List-decoding Reed–Muller codes over small fields". In: *Proceedings of the fortieth annual ACM symposium on Theory of computing*. 2008, pp. 265–274.

[GLM08] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Quantum private queries". In: *Physical review letters* 100.23 (2008), p. 230502.

[Gop70] Valerii Denisovich Goppa. "A New Class of Linear Error Correcting Codes". In: *Problems of Information Transmission* 6.3 (1970), pp. 207–212.

[Gop71] Valerii Denisovich Goppa. "Rational representation of codes and (L,g)-codes". In: *Problems of Information Transmission* 7.3 (1971), pp. 223–229.

[GPV13] Bernat Gastón, Jaume Pujol, and Merce Villanueva. "A Realistic Distributed Storage System That Minimizes Data Storage and Repair Bandwidth". In: *2013 Data Compression Conference*. 2013, pp. 491–491.

[GR08] Venkatesan Guruswami and Atri Rudra. "Explicit codes achieving list decoding capacity: Error-correction with optimal redundancy". In: *IEEE Transactions on Information Theory* 54.1 (2008), pp. 135–150.

[GRS00] Oded Goldreich, Ronitt Rubinfeld, and Madhu Sudan. "Learning polynomials with queries: The highly noisy case". In: *SIAM Journal on Discrete Mathematics* 13.4 (2000), pp. 535–570.

[Guo15] Alan Guo. “High-rate locally correctable codes via lifting”. In: *IEEE Transactions on Information Theory* 62.12 (2015), pp. 6672–6682.

[GX12] Venkatesan Guruswami and Chaoping Xing. “List decoding Reed–Solomon, algebraic-geometric, and Gabidulin subcodes up to the Singleton bound”. In: *Electronic Colloq. on Computational Complexity* 146 (2012). ISSN: 1433-8092.

[GYBS18] Ryan Gabrys, Eitan Yaakobi, Mario Blaum, and Paul H Siegel. “Constructions of partial MDS codes over small fields”. In: *IEEE Transactions on Information Theory* 65.6 (2018), pp. 3692–3701.

[HB16] Wentao Huang and Jehoshua Bruck. “Secure RAID schemes for distributed storage”. In: *IEEE International Symposium on Information Theory (ISIT).* 2016, pp. 1401–1405.

[HB17] Wentao Huang and Jehoshua Bruck. “Secure RAID schemes from EVENODD and STAR codes”. In: (2017), pp. 609–613.

[HCL13] Cheng Huang, Minghua Chen, and Jin Li. “Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems”. In: *ACM Transactions on Storage (TOS)* 9.1 (2013), pp. 1–28.

[HGK$^+$18] Anoosheh Heidarzadeh, Brenden Garcia, Swanand Kadhe, Salim El Rouayheb, and Alex Sprintson. “On the Capacity of Single-Server Multi-Message Private Information Retrieval with Side Information”. In: *2018 56th Annual Allerton Conference on Communication, Control, and Computing.* 2018, pp. 180–187.

[HJ91] Roger A. Horn and Charles R. Johnson. *Matrix Analysis.* Cambridge University Press, 1991. ISBN: 9780521548236.

[HKS18] Anoosheh Heidarzadeh, Fatemeh Kazemi, and Alex Sprintson. “Capacity of Single-Server Single-Message Private Information Retrieval with Coded Side Information”. In: *IEEE Information Theory Workshop (ITW).* 2018, pp. 1–5.

[HLH20] Hanxu Hou, Patrick PC Lee, and Yunghsiang S Han. “Minimum Storage Rack-Aware Regenerating Codes with Exact Repair and Small Sub-Packetization”. In: *IEEE International Symposium on Information Theory (ISIT).* IEEE. 2020, pp. 554–559.

[HLSH19] Hanxu Hou, Patrick PC Lee, Kenneth W Shum, and Yuchong Hu. “Rack-aware regenerating codes for data centers”. In: *IEEE Transactions on Information Theory* 65.8 (2019), pp. 4730–4745.

[HLZ16] Yuchong Hu, Patrick PC Lee, and Xiaoyang Zhang. “Double regenerating codes for hierarchical data centers”. In: *IEEE International Symposium on Information Theory (ISIT).* IEEE. 2016, pp. 245–249.

[HN20] Anna-Lena Horlemann-Trautmann and Alessandro Neri. "A Complete Classification of Partial-MDS (Maximally Recoverable) Codes with One Global Parity". In: *Advances in Mathematics of Communications* 14.1 (2020), pp. 69–88.

[Hoc59] Alexis Hocquenghem. "Codes correcteurs d'erreurs". In: *Chiffres* 2.2 (1959), pp. 147–56.

[Hoe94] Wassily Hoeffding. "Probability inequalities for sums of bounded random variables". In: *The Collected Works of Wassily Hoeffding*. Springer, 1994, pp. 409–426.

[Hol14] Henk D. L. Hollmann. "On the minimum storage overhead of distributed storage codes with a given repair locality". In: *2014 IEEE International Symposium on Information Theory*. IEEE. 2014, pp. 1041–1045.

[HSX$^+$12] Cheng Huang, Huseyin Simitci, Yikang Xu, Aaron Ogus, Brad Calder, Parikshit Gopalan, Jin Li, and Sergey Yekhanin. "Erasure Coding in Windows Azure Storage". In: *Proceedings of the 2012 USENIX Conference on Annual Technical Conference*. USENIX ATC'12. Boston, MA: USENIX Association, 2012, pp. 2–2.

[Hua17] Wentao Huang. "Coding for Security and Reliability in Distributed Systems". PhD thesis. California Institute of Technology, 2017.

[HV00] Christoph Haslach and A. J. Han Vinck. "Efficient Decoding of Interleaved Linear Block Codes". In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2000, p. 149.

[HV99] Christoph Haslach and A. J. Han Vinck. "A Decoding Algorithm With Restrictions for Array Codes". In: *IEEE Transactions on Information Theory* 45.7 (1999), pp. 2339–2344.

[HY16] Guangda Hu and Sergey Yekhanin. "New constructions of SD and MR codes over small finite fields". In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2016, pp. 1591–1595.

[IKOS04] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. "Batch codes and their applications". In: *Proceedings of the Thirty-Sixth Annual ACM Symposium on Theory of Computing*. 2004, pp. 262–271.

[Joh62] Selmer Johnson. "A New Upper Bound for Error-Correcting Codes". In: *IRE Transactions on Information Theory* 8.3 (1962), pp. 203–207.

[JTH04] Jørn Justesen, Christian Thommesen, and Tom Høholdt. "Decoding of Concatenated Codes with Interleaved Outer Codes". In: *IEEE International Symposium on Information Theory (ISIT)*. 2004, pp. 328–328.

[Kam14] Sabine Kampf. “Bounds on Collaborative Decoding of Interleaved Hermitian Codes and Virtual Extension”. In: *Designs, Codes and Cryptography* 70.1-2 (2014), pp. 9–25.

[KB16] Margreta Kuijper and Martin Bossert. “On (partial) unit memory codes based on Reed–Solomon codes for streaming”. In: *IEEE International Symposium on Information Theory (ISIT).* 2016, pp. 920–924.

[KDLM05] Zoran Kadelburg, Dusan Dukic, Milivoje Lukic, and Ivan Matic. “Inequalities of Karamata, Schur and Muirhead, and some applications”. In: *The Teaching of Mathematics* 8.1 (2005), pp. 31–45.

[KDW04] Iordanis Kerenidis and Ronald De Wolf. “Quantum symmetrically-private information retrieval”. In: *Information Processing Letters* 90.3 (2004), pp. 109–114.

[KGJØ17] Katina Kralevska, Danilo Gligoroski, Rune E. Jensen, and Harald Øverby. “Hashtag erasure codes: From theory to practice”. In: *IEEE Transactions on Big Data* 4.4 (2017), pp. 516–529.

[KGØ16] Katina Kralevska, Danilo Gligoroski, and Harald Øverby. “General sub-packetized access-optimal regenerating codes”. In: *IEEE Communications Letters* 20.7 (2016), pp. 1281–1284.

[KK16a] John Y. Kim and Swastik Kopparty. “Decoding Reed–Muller Codes Over Product Sets”. In: *31st Conference on Computational Complexity.* 2016.

[KK16b] M. Nikhil Krishnan and P. Vijay Kumar. “On MBR codes with replication”. In: *IEEE International Symposium on Information Theory (ISIT).* IEEE. 2016, pp. 71–75.

[KL97] V. Yu Krachkovsky and Yuan Xing Lee. “Decoding for Iterative Reed–Solomon Coding Schemes”. In: *IEEE Transactions on Magnetics* 33.5 (1997), pp. 2740–2742.

[KL98] V. Yu Krachkovsky and Yuan Xing Lee. “Decoding of Parallel Reed–Solomon Codes with Applications to Product and Concatenated Codes”. In: *IEEE International Symposium on Information Theory.* 1998, p. 55.

[KLL$^+$15] Aggelos Kiayias, Nikos Leonardos, Helger Lipmaa, Kateryna Pavlyk, and Qiang Tang. “Optimal rate private information retrieval from homomorphic encryption”. In: *Proceedings on Privacy Enhancing Technologies* 2015.2 (2015), pp. 222–243.

[KLP68] Tadao Kasami, Shu Lin, and W Peterson. “New generalizations of the Reed–Muller codes–I: Primitive codes”. In: *IEEE Transactions on Information Theory* 14.2 (1968), pp. 189–199.

[KLR19] Daniel Kane, Shachar Lovett, and Sankeerth Rao. “The independence number of the Birkhoff polytope graph, and applications to maximally recoverable codes”. In: *SIAM Journal on Computing* 48.4 (2019), pp. 1425–1435.

[KLRA19] Siddhartha Kumar, Hsuan-Yin Lin, Eirik Rosnes, and Alexandre Graell i Amat. “Achieving Maximum Distance Separable Private Information Retrieval Capacity With Linear Codes”. In: *IEEE Transactions on Information Theory* (2019), pp. 1–1. ISSN: 0018-9448.

[KMG19] Xiangliang Kong, Jingxue Ma, and Gennian Ge. “New Bounds on the Field Size for Maximally Recoverable Codes Instantiating Grid-like Topologies”. In: *arXiv preprint arXiv:1901.06915* (2019).

[KNK18] M. Nikhil Krishnan, Anantha Narayanan R., and P. Vijay Kumar. “Codes with Combined Locality and Regeneration Having Optimal Rate, $d_{\min}$ and Linear Field Size”. In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 1196–1200. ISBN: 978-1-5386-4781-3.

[KO97] Eyal Kushilevitz and Rafail Ostrovsky. “Replication is not needed: Single database, computationally-private information retrieval”. In: *Proceedings 38th Annual Symposium on Foundations of Computer Science*. IEEE. 1997, pp. 364–373.

[KPLK14] Govinda M. Kamath, N. Prakash, V. Lalitha, and P. Vijay Kumar. “Codes With Local Regeneration and Erasure Correction”. In: *IEEE Transactions on Information Theory* 60.8 (2014), pp. 4637–4660. ISSN: 0018-9448.

[KR68] C. G. Khatri and C. Radhakrishna Rao. “Solutions to some functional equations and their applications to characterization of probability distributions”. In: *Sankhyā: The Indian Journal of Statistics, Series A* (1968), pp. 167–180.

[Kri70] R. E. Krichevskiy. “On the number of Reed–Muller code correctable errors”. In: *Dokl. Sov. Acad. Sci.* Vol. 191. 1970, pp. 541–547.

[KS08] Tali Kaufman and Madhu Sudan. “Algebraic property testing: the role of invariance”. In: *Proceedings of the fortieth annual ACM symposium on Theory of computing*. 2008, pp. 403–412.

[KSP$^+$13] Govinda M. Kamath, Natalia Silberstein, N. Prakash, Ankit S. Rawat, V. Lalitha, O. Ozan Koyluoglu, P. Vijay Kumar, and Sriram Vishwanath. “Explicit MBR all-symbol locality codes”. In: *2013 IEEE International Symposium on Information Theory*. IEEE, 2013, pp. 504–508. ISBN: 978-1-4799-0446-4.

[KSY14] Swastik Kopparty, Shubhangi Saraf, and Sergey Yekhanin. “High-rate codes with sublinear-time decoding”. In: *J. Assoc. Comput. Mach.* 61.5 (2014), p. 28.

[LB16] Jiayang Liu and Jingguo Bi. “Cryptanalysis of a Fast Private Information Retrieval Protocol”. In: *Proceedings of the 3rd ACM International Workshop on ASIA Public-Key Cryptography.* ACM. 2016, pp. 56–60.

[LC14] Sian-Jheng Lin and Wei-Ho Chung. “Novel repair-by-transfer codes and systematic exact-MBR codes with lower complexities and smaller field sizes”. In: *IEEE Transactions on Parallel and Distributed Systems* 25.12 (2014), pp. 3232–3241.

[LG14] François Le Gall. “Powers of Tensors and Fast Matrix Multiplication”. In: *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation.* ISSAC ’14. Kobe, Japan: Association for Computing Machinery, 2014, 296–303. ISBN: 9781450325011.

[Lip05] Helger Lipmaa. “An oblivious transfer protocol with log-squared communication”. In: *International Conference on Information Security.* Springer. 2005, pp. 314–328.

[LKH20] Jie Li, David Karpuk, and Camilla Hollanti. “Towards Practical Private Information Retrieval from MDS Array Codes”. In: *IEEE Transactions on Communications* 68.6 (2020), pp. 3415–3425.

[LLL16] Mingqiang Li, Runhui Li, and Patrick PC Lee. “Relieving both storage and recovery burdens in big data clusters with R-STAIR codes”. In: *IEEE Internet Computing* (2016).

[LP17] Helger Lipmaa and Kateryna Pavlyk. “A simpler rate-optimal CPIR protocol”. In: *International Conference on Financial Cryptography and Data Security.* Springer. 2017, pp. 621–638.

[LRS06] Wei Liu, Junrye Rho, and Wonyong Sung. “Low-Power High-Throughput BCH Error Correction VLSI Design for Multi-Level Cell NAND Flash Memories”. In: *2006 IEEE Workshop on Signal Processing Systems Design and Implementation.* 2006, pp. 303–308.

[LTT18] Jie Li, Xiaohu Tang, and Chao Tian. “A generic transformation to enable optimal repair in MDS codes for distributed storage systems”. In: *IEEE Transactions on Information Theory* 64.9 (2018), pp. 6257–6267.

[LW19] Ray Li and Mary Wootters. “Lifted Multiplicity Codes and the Disjoint Repair Group Property”. In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019).* Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2019.

[McE78] Robert J McEliece. "A Public-Key Cryptosystem Based On Algebraic Coding Theory". In: *The Deep Space Network Progress Report* 44 (1978), pp. 114–116.

[MCJ73] James L. Massey, Daniel J. Costello, and Jorn Justesen. "Polynomial weights and code constructions". In: *IEEE Transactions on Information Theory* 19.1 (1973), pp. 101–110.

[MK90] John J. Metzner and Edward J. Kapturowski. "A General Decoding Technique Applicable to Replicated File Disagreement Location and Concatenated Code Decoding". In: *IEEE Transactions on Information Theory* 36.4 (1990), pp. 911–917.

[MLR$^+$14] Subramanian Muralidhar, Wyatt Lloyd, Sabyasachi Roy, Cory Hill, Ernest Lin, Weiwen Liu, Satadru Pan, Shiva Shankar, Viswanath Sivakumar, Linpeng Tang, et al. "f4: Facebook's Warm {BLOB} Storage System". In: *11th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 14)*. 2014, pp. 383–398.

[MP18] Umberto Martínez-Peñas. "Skew and Linearized Reed–Solomon Codes and Maximum Sum Rank Distance Codes Over Any Division Ring". In: *Journal of Algebra* 504 (2018), pp. 587–612.

[MP20] Umberto Martínez-Peñas. "A general family of MSRD codes and PMDS codes with smaller field sizes from extended Moore matrices". In: *arXiv preprint arXiv:2011.14109* (2020).

[MPK19] Umberto Martínez-Peñas and Frank R. Kschischang. "Universal and dynamic locally repairable codes with maximal recoverability via sum-rank codes". In: *IEEE Transactions on Information Theory* (2019).

[MS77] Florence Jessie MacWilliams and Neil James Alexander Sloane. *The Theory of Error Correcting Codes*. Vol. 16. Elsevier, 1977.

[NH20] Alessandro Neri and Anna-Lena Horlemann-Trautmann. "Random construction of partial MDS codes". In: *Designs, Codes and Cryptography* 88.4 (2020), pp. 711–725.

[Nie13] Johan S. R. Nielsen. "Generalised Multi-Sequence Shift-Register Synthesis Using Module Minimisation". In: *IEEE International Symposium on Information Theory (ISIT)*. 2013, pp. 882–886.

[Ove07] Raphael Overbeck. "Public Key Cryptography based on Coding Theory". PhD thesis. TU Darmstadt, Darmstadt, Germany, 2007.

[PAM18] N. Prakash, Vitaly Abdrashitov, and Muriel Médard. "The storage versus repair-bandwidth trade-off for clustered storage systems". In: *IEEE Transactions on Information Theory* 64.8 (2018), pp. 5783–5805.

[Par07] Farzad Parvaresh. “Algebraic List-Decoding of Error-Correcting Codes”. PhD thesis. University of California, San Diego, 2007.

[PD14] Dimitris S. Papailiopoulos and Alexandros G. Dimakis. “Locally Repairable Codes”. In: *IEEE Transactions on Information Theory* 60.10 (2014), pp. 5843–5855. ISSN: 0018-9448.

[Pet60] Wesley Peterson. “Encoding and error-correction procedures for the Bose-Chaudhuri codes”. In: *IRE Transactions on Information Theory* 6.4 (1960), pp. 459–470. ISSN: 0096-1000.

[PLD$^+$12] Dimitris S Papailiopoulos, Jianqiang Luo, Alexandros G Dimakis, Cheng Huang, and Jin Li. “Simple regenerating codes: Network coding for cloud storage”. In: *2012 Proceedings IEEE INFOCOM*. IEEE. 2012, pp. 2801–2805.

[PR17] Sven Puchinger and Johan Rosenkilde né Nielsen. “Decoding of Interleaved Reed–Solomon Codes Using Improved Power Decoding”. In: *IEEE International Symposium on Information Theory (ISIT)*. 2017.

[PRB19] Sven Puchinger, Johan Rosenkilde, and Irene Bouw. “Improved Power Decoding of Interleaved One-Point Hermitian Codes”. In: *Designs, Codes and Cryptography* 87.2-3 (2019), pp. 589–607.

[PV04] Farzad Parvaresh and Alexander Vardy. “Multivariate Interpolation Decoding Beyond the Guruswami–Sudan Radius”. In: *Allerton Conference on Communication, Control and Computing*. 2004.

[PV05] Farzad Parvaresh and Alexander Vardy. “Correcting errors beyond the Guruswami–Sudan radius in polynomial time”. In: *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS’05)*. IEEE. 2005, pp. 285–294.

[PV19] Nikita Polyanskii and Ilya Vorobyev. “Trivariate Lifted Codes with Disjoint Repair Groups”. In: *XVI International Symposium "Problems of Redundancy in Information and Control Systems" (REDUNDANCY)*. 2019, pp. 64–68.

[PW04] Ruud Pellikaan and Xin-Wen Wu. “List decoding of q-ary Reed–Muller codes”. In: *IEEE Transactions on Information Theory* 50.4 (2004), pp. 679–682.

[PW72] W. Wesley Peterson and Edward J. Weldon. *Error-correcting codes*. second. The MIT Press, 1972. ISBN: 0-262-16039-0.

[PYGP13] Jaume Pernas, Chau Yuen, Bernat Gastón, and Jaume Pujol. “Non-homogeneous two-rack model for distributed storage systems”. In: *2013 IEEE International Symposium on Information Theory*. IEEE. 2013, pp. 1237–1241.

[QLZ$^+$18] Shan Qu, Yu Liu, Jinbei Zhang, Haiwen Cao, and Xinbing Wang. “Multi-rack regenerating codes for hierarchical distributed storage systems”. In: *2018 IEEE International Conference on Communications (ICC)*. IEEE. 2018, pp. 1–6.

[Ran13] Hugues Randriambololona. “An upper bound of Singleton type for componentwise products of linear codes”. In: *IEEE Transactions on Information Theory* 59.12 (2013), pp. 7936–7939.

[RKSV13] Ankit Singh Rawat, Onur Ozan Koyluoglu, Natalia Silberstein, and Sriram Vishwanath. “Optimal locally repairable and secure codes for distributed storage systems”. In: *IEEE Transactions on Information Theory* 60.1 (2013), pp. 212–236.

[Ros18] Johan Rosenkilde. “Power Decoding Reed–Solomon Codes up to the Johnson Radius”. In: *Advances in Mathematics of Communications* 12.1 (2018), pp. 81–106.

[Rot06] Ron M. Roth. *Introduction to Coding Theory*. Cambridge University Press, 2006. DOI: 10.1017/CBO9780511808968.

[Rot91] Ron M. Roth. “Maximum-Rank Array Codes and their Application to Crisscross Error Correction”. In: *IEEE Transactions on Information Theory* 37.2 (Mar. 1991), pp. 328–336.

[RS21] Johan Rosenkilde and Arne Storjohann. “Algorithms for simultaneous Hermite–Padé approximations”. In: *Journal of Symbolic Computation* 102 (2021), pp. 279 –303. ISSN: 0747-7171.

[RS60] Irving S. Reed and Gustave Solomon. “Polynomial codes over certain finite fields”. In: *Journal of the Society for Industrial and Applied Mathematics* 8.2 (1960), pp. 300–304. ISSN: 0368-4245.

[RSG$^+$13] Korlakai Vinayak Rashmi, Nihar B. Shah, Dikang Gu, Hairong Kuang, Dhruba Borthakur, and Kannan Ramchandran. “A solution to the network challenges of data recovery in erasure-coded distributed storage systems: A study on the Facebook warehouse cluster”. In: *5th {USENIX} Workshop on Hot Topics in Storage and File Systems (HotStorage 13)*. 2013.

[RSG$^+$14] Korlakai Vinayak Rashmi, Nihar B. Shah, Dikang Gu, Hairong Kuang, Dhruba Borthakur, and Kannan Ramchandran. “A "hitchhiker’s" guide to fast and efficient data reconstruction in erasure-coded data centers”. In: *Proceedings of the 2014 ACM conference on SIGCOMM*. 2014, pp. 331–342.

[RSK11] Korlakai Vinayak Rashmi, Nihar B. Shah, and P. Vijay Kumar. "Optimal exact-regenerating codes for distributed storage at the MSR and MBR points via a product-matrix construction". In: *IEEE Transactions on Information Theory* 57.8 (2011), pp. 5227–5239.

[RV14] Ron M. Roth and Pascal O. Vontobel. "Coding for Combined Block–Symbol Error Correction". In: *IEEE Transactions on Information Theory* 60.5 (2014), pp. 2697–2713.

[SAP$^+$13] Maheswaran Sathiamoorthy, Megasthenis Asteris, Dimitris Papailiopoulos, Alexandros G. Dimakis, Ramkumar Vadali, Scott Chen, and Dhruba Borthakur. "XORing elephants: novel erasure codes for big data". In: *Proceedings of the 39th international conference on Very Large Data Bases*. PVLDB'13. Trento, Italy: VLDB Endowment, 2013, pp. 325–336.

[SC07] Radu Sion and Bogdan Carbunar. "On the computational practicality of private information retrieval". In: *Proceedings of the Network and Distributed Systems Security Symposium*. Internet Society. 2007, pp. 2006–06.

[SCM18] Jy-yong Sohn, Beongjun Choi, and Jaekyun Moon. "A class of MSR codes for clustered distributed storage". In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2018, pp. 2366–2370.

[SCYM18] Jy-yong Sohn, Beongjun Choi, Sung Whan Yoon, and Jaekyun Moon. "Capacity of clustered distributed storage". In: *IEEE Transactions on Information Theory* 65.1 (2018), pp. 81–107.

[Seg55] Beniamino Segre. "Curve razionali normali ek-archi negli spazi finiti". In: *Annali di Matematica Pura ed Applicata* 39.1 (1955), pp. 357–379.

[SH19] Seunghoan Song and Masahito Hayashi. "Capacity of quantum private information retrieval with collusion of all but one of servers". In: *IEEE Information Theory Workshop (ITW)*. IEEE. 2019, pp. 1–5.

[SH20a] Seunghoan Song and Masahito Hayashi. "Capacity of quantum private information retrieval with colluding servers". In: *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2020, pp. 1077–1082.

[SH20b] Seunghoan Song and Masahito Hayashi. "Capacity of quantum private information retrieval with multiple servers". In: *IEEE Transactions on Information Theory* 67.1 (2020), pp. 452–463.

[Sha48] C. E. Shannon. "A mathematical theory of communication". In: *Bell System Technical Journal* 27 (1948), pp. 379–423, 623–656.

[Sha79] Adi Shamir. “How to Share a Secret”. In: *Commun. ACM* 22.11 (Nov. 1979), pp. 612–613. ISSN: 0001-0782.

[SJ17] Hua Sun and Syed Ali Jafar. “The Capacity of Private Information Retrieval”. In: *IEEE Transactions on Information Theory* 63.7 (2017), pp. 4075–4088. ISSN: 0018-9448.

[SJ18a] Hua Sun and Syed Ali Jafar. “Private Information Retrieval from MDS Coded Data With Colluding Servers: Settling a Conjecture by Freij-Hollanti et al.” In: *IEEE Transactions on Information Theory* 64.2 (2018), pp. 1000–1022. ISSN: 0018-9448.

[SJ18b] Hua Sun and Syed Ali Jafar. “The Capacity of Robust Private Information Retrieval With Colluding Databases”. In: *IEEE Transactions on Information Theory* 64.4 (2018), pp. 2361–2370. ISSN: 0018-9448.

[SJ18c] Hua Sun and Syed Ali Jafar. “The capacity of symmetric private information retrieval”. In: *IEEE Transactions on Information Theory* 65.1 (2018), pp. 322–329.

[SKHN76] Yasuo Sugiyama, Masao Kasahara, Shigeichi Hirasawa, and Toshihiko Namekawa. “Further Results on Goppa Codes and Their Applications to Constructing Efficient Binary Codes”. In: *IEEE Transactions on Information Theory* 22.5 (1976), pp. 518–526. ISSN: 0018-9448.

[Sly97] Vadim Slyusar. “New operations of matrices product for applications of radars”. In: *Proc. Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory.* 1997, pp. 73–74.

[SR10] Changho Suh and Kannan Ramchandran. “Exact-repair MDS codes for distributed storage using interference alignment”. In: *2010 IEEE International Symposium on Information Theory.* IEEE. 2010, pp. 161–165.

[SRKR11] Nihar B Shah, KV Rashmi, P Vijay Kumar, and Kannan Ramchandran. “Interference alignment in regenerating codes for distributed storage: Necessity and code constructions”. In: *IEEE Transactions on Information Theory* 58.4 (2011), pp. 2134–2158.

[SRLS18] D. Shivakrishna, V. Arvind Rameshwar, V. Lalitha, and Birenjith Sasidharan. “On Maximally Recoverable Codes for Product Topologies”. In: *2018 Twenty Fourth National Conference on Communications (NCC).* IEEE. 2018, pp. 1–6.

[SRV15] Natalia Silberstein, Ankit Singh Rawat, and Sriram Vishwanath. “Error-Correcting Regenerating and Locally Repairable Codes via Rank-Metric Codes”. In: *IEEE Transactions on Information Theory* 61.11 (2015), pp. 5765–5778. ISSN: 0018-9448.

[SRZ06] Fei Sun, Ken Rose, and Tong Zhang. "On the Use of Strong BCH Codes for Improving Multilevel NAND Flash Memory Storage Capacity". In: *IEEE Workshop on Signal Processing Systems (SiPS)*. 2006.

[SS11] Vladimir Sidorenko and Georg Schmidt. "A Linear Algebraic Approach to Multisequence Shift-Register Synthesis". In: *Problems of Information Transmission* 47 (June 2011), pp. 149–165.

[SSB05] Georg Schmidt, Vladimir R Sidorenko, and Martin Bossert. "Interleaved Reed–Solomon Codes in Concatenated Code Designs". In: *IEEE Information Theory Workshop*. 2005, 5–pp.

[SSB07] Georg Schmidt, Vladimir Sidorenko, and Martin Bossert. "Enhancing the Correcting Radius of Interleaved Reed–Solomon Decoding Using Syndrome Extension Techniques". In: *IEEE International Symposium on Information Theory (ISIT)*. 2007, pp. 1341–1345.

[SSB08] Vladimir Sidorenko, Georg Schmidt, and Martin Bossert. "Decoding Punctured Reed–Solomon Codes up to the Singleton Bound". In: *International ITG Conference on Source and Channel Coding*. VDE. 2008.

[SSB09a] G. Schmidt, V. R. Sidorenko, and M. Bossert. "Collaborative Decoding of Interleaved Reed–Solomon Codes and Concatenated Code Designs". In: *IEEE Transactions on Information Theory* 55.7 (2009), pp. 2991–3012. ISSN: 0018-9448.

[SSB09b] Georg Schmidt, Vladimir R Sidorenko, and Martin Bossert. "Collaborative Decoding of Interleaved Reed–Solomon Codes and Concatenated Code Designs". In: *IEEE Transactions on Information Theory* 55.7 (2009), pp. 2991–3012.

[SSB10] Georg Schmidt, Vladimir R Sidorenko, and Martin Bossert. "Syndrome Decoding of Reed–Solomon Codes Beyond Half the Minimum Distance Based on Shift-Register Synthesis". In: *IEEE Transactions on Information Theory* 56.10 (2010), pp. 5245–5252.

[STV01] Madhu Sudan, Luca Trevisan, and Salil Vadhan. "Pseudorandom generators without the XOR lemma". In: *Journal of Computer and System Sciences* 62.2 (2001), pp. 236–266.

[TB14a] Itzhak Tamo and Alexander Barg. "A Family of Optimal Locally Recoverable Codes". In: *IEEE Transactions on Information Theory* 60.8 (2014), pp. 4661–4676.

[TB14b] Itzhak Tamo and Alexander Barg. "A family of optimal locally recoverable codes". In: *IEEE Transactions on Information Theory* 60.8 (2014), pp. 4661–4676.

[TBGC15] Itzhak Tamo, Alexander Barg, Sreechakra Goparaju, and Robert Calderbank. "Cyclic LRC codes and their subfield subcodes". In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2015, pp. 1262–1266.

[TCS14] M. Ali Tebbi, Terence H. Chan, and Chi Wan Sung. "A code design framework for multi-rack distributed storage". In: *IEEE Information Theory Workshop (ITW)*. IEEE. 2014, pp. 55–59.

[TGE18] Razane Tajeddine, Oliver W. Gnilke, and Salim El Rouayheb. "Private information retrieval from MDS coded data in distributed storage systems". In: *IEEE Transactions on Information Theory* 64.11 (2018), pp. 7081–7093.

[TGK$^+$19] Razane Tajeddine, Oliver W Gnilke, David Karpuk, Ragnar Freij-Hollanti, and Camilla Hollanti. "Private information retrieval from coded storage systems with colluding, Byzantine, and unresponsive servers". In: *IEEE Transactions on Information Theory* 65.6 (2019), pp. 3898–3906.

[TPD16] Itzhak Tamo, Dimitris S Papailiopoulos, and Alexandros G Dimakis. "Optimal locally repairable codes and connections to matroid theory". In: *IEEE Transactions on Information Theory* 62.12 (2016), pp. 6661–6671.

[TSC19] Chao Tian, Hua Sun, and Jun Chen. "Capacity-achieving private information retrieval codes with optimal message size and upload cost". In: *IEEE Transactions on Information Theory* 65.11 (2019), pp. 7613–7627.

[TWB12] Itzhak Tamo, Zhiying Wang, and Jehoshua Bruck. "Zigzag codes: MDS array codes with optimal rebuilding". In: *IEEE Transactions on Information Theory* 59.3 (2012), pp. 1597–1616.

[WDPZ11] Xueqiang Wang, Guiqiang Dong, Liyang Pan, and Runde Zhou. "Error Correction Codes and Signal Processing in Flash Memory". In: InTech, 2011, pp. 57–82. ISBN: 978-953-307-272-2.

[Wir88] Michael Wirtz. "On the Parameters of Goppa Codes". In: *IEEE Transactions on Information Theory* 34.5 (1988), pp. 1341–1343. ISSN: 0018-9448.

[Wol65] Jack Wolf. "On codes derivable from the tensor product of check matrices". In: *IEEE Transactions on Information Theory* 11.2 (1965), pp. 281–284.

[WS17a] Qiwen Wang and Mikael Skoglund. "Linear symmetric private information retrieval for MDS coded distributed storage with colluding servers". In: *IEEE Information Theory Workshop (ITW)*. 2017, pp. 71–75.

[WS17b] Qiwen Wang and Mikael Skoglund. “Secure symmetric private information retrieval from colluding databases with adversaries”. In: *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. 2017, pp. 1083–1090.

[WS17c] Qiwen Wang and Mikael Skoglund. “Symmetric private information retrieval for MDS coded distributed storage”. In: *2017 IEEE International Conference on Communications (ICC)*. 2017, pp. 1–6.

[WS19] Qiwen Wang and Mikael Skoglund. “Symmetric Private Information Retrieval from MDS Coded Distributed Storage With Non-Colluding and Colluding Servers”. In: *IEEE Transactions on Information Theory* 65.8 (2019), pp. 5160–5175. ISSN: 1557-9654.

[WSS19] Qiwen Wang, Hua Sun, and Mikael Skoglund. “Symmetric Private Information Retrieval with Mismatched Coded Messages and Randomness”. In: *IEEE International Symposium on Information Theory (ISIT)*. 2019.

[Wu15] Liyasi Wu. “Revisiting the multiplicity codes: A new class of high-rate locally correctable codes”. In: *53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. 2015, pp. 509–513.

[WZB14] Antonia Wachter-Zeh, Alexander Zeh, and Martin Bossert. “Decoding Interleaved Reed–Solomon Codes Beyond Their Joint Error-Correcting Capability”. In: *Designs, Codes and Cryptography* 71.2 (2014), pp. 261–281.

[YB17a] Min Ye and Alexander Barg. “Explicit constructions of high-rate MDS array codes with optimal repair bandwidth”. In: *IEEE Transactions on Information Theory* 63.4 (2017), pp. 2001–2014.

[YB17b] Min Ye and Alexander Barg. “Explicit constructions of optimal-access MDS codes with nearly optimal sub-packetization”. In: *IEEE Transactions on Information Theory* 63.10 (2017), pp. 6307–6317.

[YEC12] Chengen Yang, Yunus Emre, and Chaitali Chakrabarti. “Product Code Schemes for Error Correction in MLC NAND Flash Memories”. In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 20.12 (2012), pp. 2302–2314.

[YKPB12] Xun Yi, Mohammed Golam Kaosar, Russell Paulet, and Elisa Bertino. “Single-database private information retrieval from fully homomorphic encryption”. In: *IEEE Transactions on Knowledge and Data Engineering* 25.5 (2012), pp. 1125–1134.

[YL18] Jiun-Hung Yu and Hans-Andrea Loeliger. “Simultaneous Partial Inverses and Decoding Interleaved Reed–Solomon Codes”. In: *IEEE Transactions on Information Theory* 64.12 (2018), pp. 7511–7528.

[ZTSL20] Ruida Zhou, Chao Tian, Hua Sun, and Tie Liu. “Capacity-achieving private information retrieval codes from MDS-coded databases with minimum message size”. In: *IEEE Transactions on Information Theory* 66.8 (2020), pp. 4904–4916.

[ZX18] Zhifang Zhang and Jingke Xu. “The optimal sub-packetization of linear capacity-achieving PIR schemes with colluding servers”. In: *IEEE Transactions on Information Theory* 65.5 (2018), pp. 2723–2735.

[ZYQT19] Jinbao Zhu, Qifa Yan, Chao Qi, and Xiaohu Tang. “A new capacity-achieving private information retrieval scheme with (almost) optimal file length for coded servers”. In: *IEEE Transactions on Information Forensics and Security* 15 (2019), pp. 1248–1260.