Low Field Size constructions of Access Optimal Convertible Codes

Thesis Submitted

in Partial Fulfillment of the

Requirements for the Degree of

POST-GRADUATION M.Tech

in

COMPUTER SCIENCE AND ENGINEERING

By

VANGALAPUDI APURV

(ROLL NO. 2K23/CSE/17)

Under the Supervision of

Dr. Minni Jain

Assistant Professor, Department of Computer Science Engineering Delhi Technological University (DTU)

To the

Department of Computer Science Engineering

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India

May, 2025

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE’S DECLARATION

I **Vangalapudi Apurv** hereby certify that the work which is being presented in the thesis entitled **Low Field Size constructions of Access Optimal Convertible Codes** in partial fulfillment of requirements for the award of the Degree of Masters of Technology (MTECH.), submitted in the Department of Computer Science Engineering, Delhi Technological University is an authentic record of my own work under the supervision of **Dr. Minni Jain.**

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Candidate's Signature

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

Certified that **Vangalapudi Apurv (**2K23/CSE/17) has carried out their search work presented in this thesis entitled **"Low Field Size constructions of Access Optimal Convertible Codes"** for the award of Master of Technology from Department of Computer Science Engineering, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Dr. Minni Jain
Assistant Professor
(Signature)
Department of CSE, DTU

Date:

ABSTRACT

Large-scale storage systems typically use erasure coding to ensure data durability against disk failures. Recent research indicates that adjusting the level of redundancy based on varying disk failure rates can lead to significant storage efficiency improvements. This adjustment involves code conversion, where data originally encoded with a code needs to be re-encoded into a code, a process that can be resource- demanding. Convertible codes offer a way to facilitate this transformation efficiently while preserving other valuable properties. This project examines the access cost of conversion, defined as the total number of code symbols accessed during the process, and explores a specific type of conversion called the merge regime, which consolidates multiple initial codewords into one final codeword. Although systematic, access-optimal Maximum Distance Separable (MDS) convertible codes for all parameters in the merge regime have been established, the current method for a key subset of these parameters relies on Left Shift parity matrices, requiring a large field size and thus limiting its practical usability. In this work, we present (1) improved bounds on the minimum field size needed for such codes and (2) probabilistic constructions that support lower field sizes for a variety of parameter ranges, utilizing the Combinatorial Nullstellensatz theorem.

ACKNOWLEDGEMENT

I would like to express our heartfelt gratitude to all the individuals who have supported and assisted me throughout my M.Tech thesis. First and foremost, I would like to thank my supervisor, “Dr. Minni Jain”, Assistant Professor, Department of Computer Engineering, Delhi Technological University for his constant guidance, support, and encouragement throughout the project. I am indebted to him for sharing his knowledge, expertise, and valuable feedback that helped me in shaping the thesis.

I would like to extend my sincere thanks to the Vice Chancellor of Delhi Technological University and the faculty members of the Department of Computer Engineering for their support and encouragement throughout our academic journey.

Vangalapudi

Apurv

(2K23/CSE/17)

Table Of Content

List Of Tables

List Of Figures

CHAPTER 1

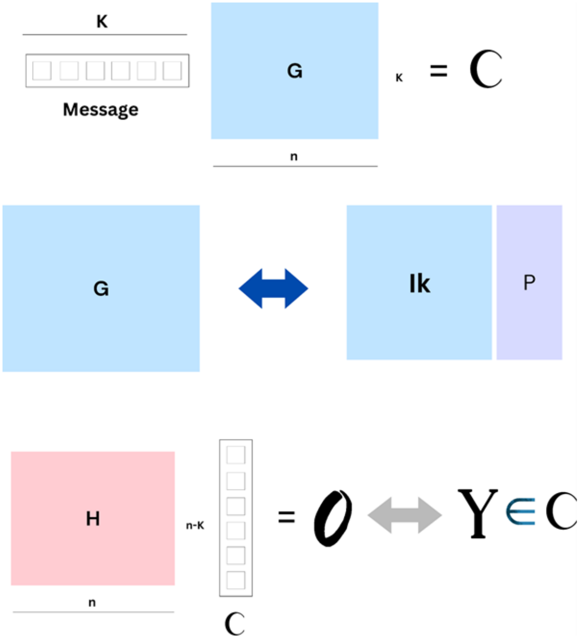
Introduction

Erasure canons are used considerably in current big scale allocated storehouse structures as a way to alleviate data loss in the event of fragment failures. in this environment, erasure coding includes dividing data into businesses of k gobbets which are every decoded into stripes of n gobbets using an (n, k) erasure law. these decoded gobbets are also saved across n distinct storehouse bumps within the machine. The law parameters ‘n’ and ‘k’ decide the volume of redundancy brought to the contrivance and the parchment of continuity assured.

There are colorful training of canons that are generally used in factual-transnational structures. as an illustration, methodical canons are those wherein the original communication symbols are bedded utmost of the law symbols. this is incredibly desirable in practice as inside the event that there are no determined fragment screw ups, there is no decoding system had to recover the original information. Methodical canons with Vandermonde equality matrices are indeed lesser superb as there are recognised effective algorithms using rapid-fire Fourier rework (FFT) for calculating the product between vectors and Vandermonde matrix (5, 12), speeding up the garbling procedure.

This trait is turning into an adding number of critical given the rearmost fashion to use wider (high k) and longer (high n) erasure canons (6, 10). also, maximum Distance Separable (MDS) canons are a subset of erasure canons that bear the least quantum of redundant garage as a way to meet a particular failure forbearance purpose. An (n, k) MDS law can tolerate lack of any $n - k$ out of the n law symbols. in this charge, the point of interest is on methodical MDS canons with Vandermonde equality matrices.

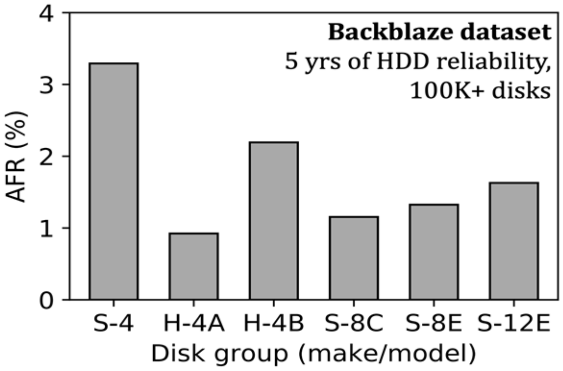
Recent findings via Kadekodi et al. cover the dynamic variability in fragment failure rates through the times. Their exploration highlights the capability for meaningful fiscal savings in garage and related functional freights via tuning law parameters to discovered failure charges. but, the aid above related to the dereliction approach forward-encoding all of the statistics with the intention to acclimate n and k is prohibitively acutely- priced.



(Fig.1.1)

The law conversion trouble formalizes the trouble of effectively revising statistics that has been decoded beneath an (n(I), k(I)) original law C(I) to its new illustration underneath an (n(F), k(F)) veritably last law C(F). one of the crucial measures of the price of conversion is the get entry to cost, which represents the overall number of law symbols penetrated (examine/ written) at some point of conversion. Convertible canons are an order of canons that permit green conversion indeed as keeping different respectable homes together with being MDS and methodical .
among different feathers of transformations, the merge governance, in which k(F)=λk(I) for any integer λ ≥ 2(i.e., combining multiple original devices right into a single final metaphor), is the maximum important bone . First, the merge governance requires the least resource application among all feathers of transformations and latterly are a rather favorable desire for sensible structures. 2d, structures for the merge governance are crucial structure blocks for the constructions for canons inside the standard governance which allows for any set of original parameters and any set of veritably last parameters. This project focuses on methodical MDS convertible canons inside the merge governance.
The authors hooked up drop bounds on the get admission to cost of conversion among dyads of direct MDS canons and furnished structures of get admission to-top-quality convertible canons for all parameters in the merge governance, which meet the hooked up drop bounds. let us denote r(I) = n(I) − k(I) and r(F) = n(F) − k(F),(which correspond to the wide variety of equality symbols inside the original and veritably last canons if the canons are methodical). For multitudinous cases wherein r(I)> r(F) (i.e., whilst the primary configuration has further equivalence than the final configuration), the authors offer unequivocal structures of methodical MDS access-foremost convertible canons over fields of size direct in n(F). For cases in which r(I)< r(F)(i.e., whilst redundant equivalence are wanted inside the final configuration than inside the primary), it has been shown that the get entry to price

of conversion for MDS erasure canons is lower bounded via that of the dereliction approach to decode forward-encode all of the data. as a result, it is not doable to comprehend any fiscal savings with specialised law structures.
but, within the case in which r(I) = r(F), the high- quality- regarded creation requires a minimal subject size of pD for any high p and a many D ∊ Θ((n(F))^3). This area size is far too inordinate for green sensible executions. maximum current coaching-set infrastructures are optimized to perform on bytes of data at a time. exercising erasure canons described over larger subject sizes can abate the encoding/ decoding pace. for this reason outside (if not each) sensible executions of garage canons use F(256) (which translates each subject symbol to a one- byte illustration). for that reason, the hassle of constructing low subject size access- gold standard convertible canons remains open for the case r(I) = r(F).Methodical canons with Cauchy equality matrices are indeed more tremendous due to the fact Cauchy matrices have robust fine parcels that permit effective garbling and decrypting algorithms to be designed. substantially, addition related to Cauchy matrices may be applied rightly through structured algorithms that avoid largely-priced matrix inversions or heavy calculations. This characteristic turns into decreasingly further vital given the fashion toward using wider (high k) and longer (high n) erasure canons. also, Cauchy- grounded completely structures constantly allow for small area sizes, that's critical for practical performance on present day tackle infrastructures which can be optimized for byte- degree operations. likewise, maximum Distance Separable (MDS) canons are a subset of erasure canons that bear the least volume of fresh storehouse to satisfy a particular failure forbearance thing. An (n, k) MDS law can tolerate the loss of any (n − k) out of the n law symbols. on this adventure, the focal point is on methodical MDS canons with Cauchy- grounded equality matrices.



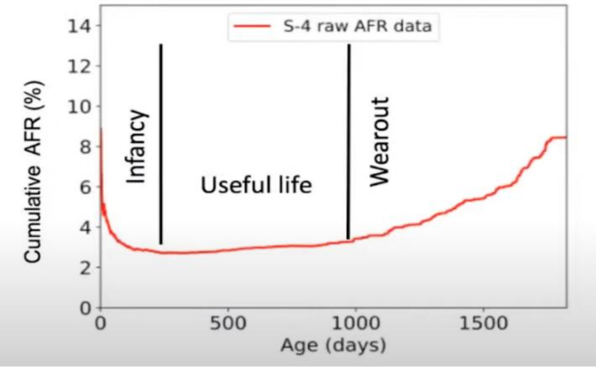
(Fig.1.2)

current findings through Kadekodi et al. reveal the dynamic variability in fragment failure prices over time. Their exploration highlights the capacity for meaningful fiscal savings in garage and associated functional prices through tuning law parameters to located failure quotations. but, the resource above related to the dereliction approach forward-encoding all of the information with a purpose to modify 'n' and 'k' is prohibitively luxurious. The law conversion problem formalizes the assignment of effectively transubstantiating records that has been decoded underneath an primary law (n(I), k(I)) into its new representation below a veritably last law (n(F), k(F)). one of the crucial measures of the price of conversion is the get right of entry to price, representing the total wide variety of law symbols penetrated (study/ written) throughout conversion. Convertible canons are a class of canons that allow green conversion indeed as retaining respectable houses which include being MDS and methodical among colorful kinds of transformations, the merge governance, where k(F) = λk(I) for any integer λ ≥ 2 (i.e., combining multiple original devices into a unattached veritably last metaphor), is the most critical. The merge governance calls for the least resource operation amongst all types of transformations and therefore is a rather favorable preference for realistic systems. also, structures for the merge governance serve as essential structure blocks for lesser star structures that permit arbitrary original and veritably last law parameters. thus, this thesis specializes in methodical MDS convertible canons inside the merge governance.

The authors established drop bounds on the get right of entry to cost of conversion among dyads of direct MDS canons and supplied unequivocal structures of access- premier convertible canons for all parameters in the merge governance, negotiating those lower bounds. let us denote r(I) = n(I) − k(I) and r(F) = n(F) − k(F), which correspond to the range of equality symbols inside the primary and veritably last canons(assuming methodical shape).For multitudinous cases where r(I)> r(F)(i.e., whilst the primary law has redundant equivalence than the final), specific constructions of methodical MDS get admission to- stylish convertible canons over fields of length direct in n(F) are supplied. still, for cases in which r(I) redundant equivalence are wanted in the veritably last configuration), it's been proven that the access figure of conversion for MDS erasure canons is drop bounded by the figure of absolutely decoding forward-encoding all the information. consequently, no specialised law constructions can outperform the naive system in that case.

still, while r(I) = r(F), the satisfactory- conceded constructions bear large area sizes, substantially fields of size pD for some high p and D ∊ Θ((n(F))^3). similar area sizes are impracticable for green real- world executions due to the fact current processors are optimized for operations over small fields, specifically over F(256) (in which every field element can match within one byte). hence, the hassle of constructing low- subject- length get right of entry to- gold standard convertible canons stays open for the case r(I) = r(F).

(Fig.1.3)



CHAPTER 2

Background and related work

Let us begin with an overview of important concepts and notation referred to throughout this project, along with a literature review of previous related work.

2.1 Systematic MDS codes and matrices

An (n, k) direct erasure law 'C' with creator matrix G ∊ M.F(k × n) over a finite area' F' is said to be methodical , or in fashionable shape, if G = (I(k)| P) wherein I(k) is the k × k identification matrix and ' P' is a k×(n − k) matrix also appertained to as the equality matrix. permit m be a communication and' c' be its corresponding metaphor underneath ' C', wherein m = (m(i)) k(i) = 1 and c = (c(i)) n(i) = 1 are vectors of communication and law symbols, independently. As' m' is decoded underneath' C' through the addition c = (mT) G, it follows that c(i) = m(i) for all i ≤ k if 'C' is methodical.

An (n, k) direct erasure law' C' is most Distance Separable (MDS) if and stylish if each 'k' columns of its creator matrix 'G' are linearly unprejudiced; in different expressions, each k× k submatrix of' G' is non-singular. As a result, statistics decoded via an (n, k) MDS law can face up to any erasure sample of (n − k) out symbols in any metaphor and nevertheless efficiently recover the original facts. However, this is original to the things that every square submatrix of' P' is non-singular, If 'C' is likewise methodical with equality matrix' P'. such a matrix is likewise called outstanding- everyday. It's useful to word that any submatrix of a extremely good-regular matrix is also splendid-ordinary.
we are running with an (n, k) direct erasure law C described over a finite subject F, in which the creator matrix G is a k × n matrix. within the methodical form, G is grounded as (I(k) | P), where I(k) is the k × k identification matrix and P is a k ×(n − k) equality matrix. This shape guarantees that when a communication vector m = (m1,m2,…, mk) is decoded into a metaphor c = (c1, c2,…, cn) through the addition c =
(mT) G, the primary ok symbols of the metaphor are exactly the communication symbols; this is, c(i) = m(i) for all i ≤ k.
For one of these law C to be maximum Distance Separable (MDS), it need to fulfill the things that any selection of k columns of G are linearly unprejudiced. in the case of a scientific law, this demand simplifies it suffices to insure that every square submatrix of the equality matrix P is invertible(non-singular). while P has this things, we name it awful-ordinary, and it ensures that the law can get over any sample of

(n – k) erasures.

To assemble a great-regular equality matrix P, one important approach we can use is to employ a Cauchy matrix. A Cauchy matrix is described by using entries of the form

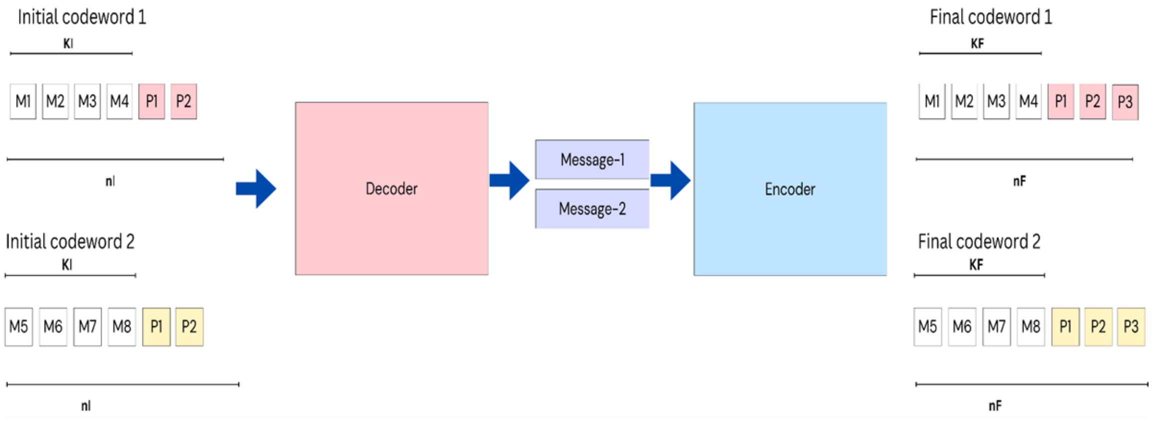
$P(ij)$ = 1/($x(\ i)$ − $y(\ j)$), in which { $x1,, xk$} and { $y1,, y(\ n - k)$} are two disjoint units of factors from the sphere F, and $xi \neq yj$. This construction ensures that no longer simplest the entire matrix still also each square submatrix of P is non-singular, for this reason easily furnishing the needed super-regularity.

Cauchy matrices are especially effective because their shape guarantees the MDS means without taking us to corroborate the invertibility of every submatrix in my opinion. likewise, they retain specific expression for their antitheses, making them computationally appealing for garbling and decrypting operations. we're suitable to construct Cauchy matrices over any finite subject massive enough to deal with the necessary stupendous x and y factors, making them considerably applicable.

for that reason, by using deciding on suitable awful x and y sets, erecting P using the Cauchy formula, and forming G = (I(k) | P), we gain a scientific MDS law prepared to render any k- image communication with strong ensures of erasure mending. This fashion affords an swish and effective pathway to achieving methodical MDS canons, fending off the complications related to vindicating matrix homes for redundant arbitrary structures.

2.2 Code conversion

This introduces the motivation behind the research, highlighting that disk failure rates in storage systems are highly variable. To address this, code conversion is proposed, which involves adjusting the parameters of erasure coding from an initial code ($[n(I),k(I)]$)($C(I)$) to a final code ($[n(F),k(F)]$)($C(F)$). The goal of this adjustment is to optimize storage efficiency and redundancy dynamically as the system requirements change. The slide references research by Kadekodi et al., emphasizing that dynamic tuning of erasure codes is crucial for minimizing storage demands and reducing operational costs. Additionally, the concept of Access Cost is defined, referring to the total number of symbols that must be accessed (both read and written) during the code conversion process.



(Fig.2.2.1)

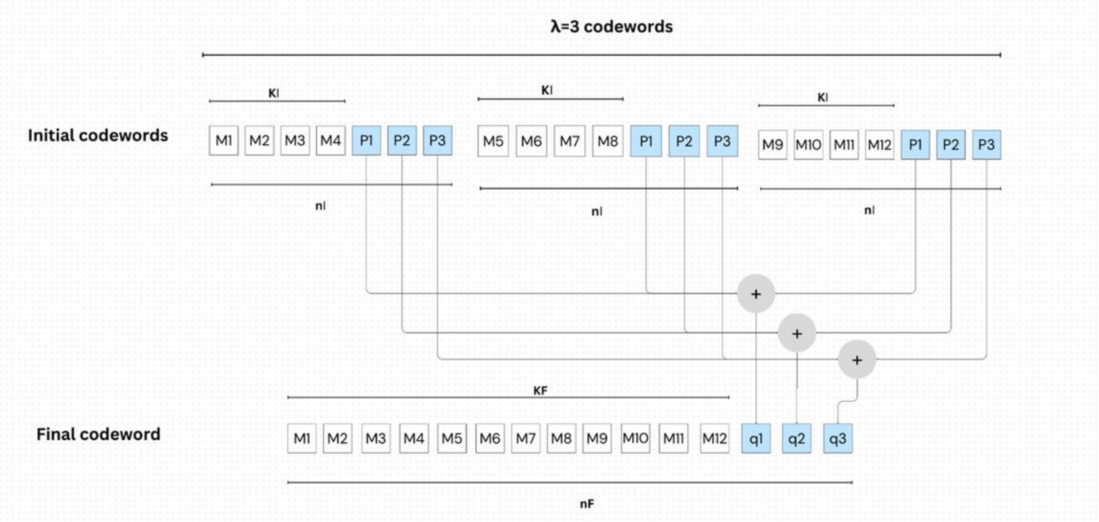
This explains the traditional method for converting codes: the straightforward decode and re- encode approach. It involves fully decoding the existing coded data and then re-encoding it according to the new parameters. An example is provided where a code of structure [2(K+2),2K] is converted to [2(K+3),2K], representing a transition from dual parity check codes to tri-parity check codes. The visual shows initial codewords consisting of messages and parity bits being fed into a decoder, producing intermediate messages, which are then input to an encoder to form the new codewords. This method, although simple, can be costly in terms of access operations because it requires reading and writing a large amount of data.

2.2(a) Optimized Code Conversion with Lower Access Cost

Here, an improved method for code conversion is detailed, showing a more efficient way to generate the final codewords from the initial codewords without full decoding. Instead of decoding and re-encoding the entire data, a selected number of blocks are read and processed with minimal addition operations (indicated by the "+" symbols). This reduces the read access to only four blocks and write access to three new parity blocks (q1, q2, q3). The visualization on the left shows how specific blocks are accessed and combined to create the new parity symbols, while the one on the right shows the pathways between the initial and final codewords. In contrast, the default re-encoding method would have required reading 2n(I) blocks, demonstrating a clear efficiency advantage with this optimized approach.

2.2(b) Merge regime

The merge regime in a coding setup, particularly when merging multiple codewords into fewer ones to optimize storage or transmission. Initially, we have blocks of codewords grouped under Initial codewords. Each block contains a set of message symbols $M(i)$ and associated parity checks $P1$, $P2$, and $P3$. These parity checks are linear combinations of the messages within each group, for example: $P1(M1,M2,M3,M4)=M1+M2+M4$, and similarly for the other groups. Here, λ=3 indicates that three initial codewords are combined to form one final codeword. The merging process results in the Final codeword row, where message symbols $M1$ to $M12$ are retained alongside new parity symbols $q1$, q 2, and $q3$ .These qi values are carefully constructed as specific sums and weighted combinations of the earlier parities and messages to preserve code properties. Specifically, $q1$ is a sum of the first parities across three groups, $q2$involves weighted sums where coefficients increase as powers of 2 (i.e., 8, 16), and $q3$uses even higher weights (16, 32), ensuring distinctiveness and reliability. Mathematically, the relationship between parameters is maintained such that k F=λk (I)(final dimension is λ times the initial dimension), and similarly for n F and nI, with adjustments in redundancy r. The access cost, representing the effort to reconstruct or access the original data, is guaranteed to be at least r F+λmin(k I,r F), ensuring efficient and robust recovery even after merging. This elegant structure reduces redundancy while maintaining reliability, a typical goal in advanced coding techniques like distributed storage and network coding.



(Fig.2.2.2)

2.3 Construction of vandermonde matrix in merge regime

the construction of a Vandermonde matrix in the merge regime, typically relevant for coding theory, network coding, or distributed storage.First, the setting involves initial parameters N(i)=6, K(i) =3, implying r(i)=3, and final parameters N(f)=9, K(f)=6, giving r(f)=3 as well.

Here, N and K represent the total and the dimension (or redundancy parameters) of the system, while r denotes the rank (or repair degree).Two generator matrices are defined:

*$G(I)$=[$I|P(I)$]

*G (F)=[I|P(F)]

where I is an identity matrix and P(I), P(F) are specific Vandermonde matrices.The matrices P (I) and P(F) are based on powers of a primitive element θ. The initial Vandermonde matrix P(I) is relatively simple: powers of θ increment uniformly across rows and columns. In contrast, P(F) has a more complex structure: powers are functions of multiples of i and k, reflecting the expansion from N(i) to N(f).

The diagram on the right explains how these matrices merge practically:

*Initially, three sets of matrices M1−M3, M4−M6, and M7−M9 are grouped with corresponding query vectors q1, q2, q3 in the N(i) domain.

*Each group performs operations indicated as 1X, θ^3X, and θ^6X — meaning multiplication of the query vectors by powers of θ.

*Then, the outputs are summed together (denoted by the circle with a '+' symbol).

*The final merged system is of size N(f) where all nine matrices M1 to M9 are available, maintaining the correct aligned query vectors q1, q2, and q3.

This design ensures that the merged system preserves the key Vandermonde structure across the expansion, supporting fault tolerance, efficient merging, and retrieval while maintaining coding redundancy.

2.4 My idea

2.4(a) Right Cyclic Shift Extended Matrices

This introduces a study around right cyclic shift matrices, particularly comparing them to Vandermonde matrices — which are classically used in coding theory and interpolation. Theoretical Comparison: Here, the idea is to understand how large a field size must be to construct matrices that are suitable for coding applications when using right cyclic shift matrices versus traditional Vandermonde matrices. Vandermonde matrices have nice properties (like full-rank for distinct elements), but they often require relatively large field sizes. Right cyclic shift matrices might offer an advantage by needing smaller fields.
Low Field Size Constructions: The goal here is practical: to build matrices that exhibit super- regularity (i.e., all square submatrices are full-rank) while keeping the field size as small as possible. Small fields are desirable in practice because they make implementation more efficient (less memory, faster computation).
This matrix undergoes cyclic shifts to generate different rows. A cyclic shift moves the elements of a row or column by one position in a cyclic manner (wrapping around at the ends).

*Lagrange Interpolation in Encoding

The lower part explains how polynomial-based encoding is performed:
Polynomial Representation:

A message is represented as a
polynomial
P(x)=a1x+a2x^2+⋯+akx^(k−1)+ak
x^k.
The coefficients a1,a2,…,ak represent the information symbols.

Codeword Generation:

To generate the codeword, evaluate P(x) at n distinct points (say α1,α2,…,αn) in a finite field
GF(2^m).The resulting codeword is:

C=[P(α1),P(α2),…,P(αn)] (2.4.1)

This is similar to Reed-Solomon encoding. The combination of right cyclic shift matrices and polynomial evaluations aims to produce highly reliable codes over smaller fields — making them attractive for communication and storage systems.
This goes deeper into the actual matrix construction. The large blue matrix shows a

block structure with an identity-like diagonal of important elements like α1,α2,α3,....

Here:

*kf (vertical axis) represents the number of rows (likely proportional to the number of information symbols).

*rf (horizontal axis) represents the number of parity-check symbols.

Construction Details:

Powers of α are assigned to different matrix entries. These powers are chosen carefully to ensure the matrix's desired properties like super-regularity and full-rank

conditions. Specific formula examples:

For instance, x 13 is expanded as a product of differences involving powers of a (primitive element of GF(2^m)).

Similarly, x14 is obtained by taking the product over terms like (x−a^k) for appropriate k.

Substitution:

*When x=0 is substituted, the evaluation simplifies to a product of constants (negated powers of a).

Color-coded blocks on the right side:

*Different colored blocks seem to represent how different components are constructed and related through cyclic shifts.

*Arrows indicate the direction of shift or assignment.

CHAPTER 3

Fundamental limits on field size

The exceptional- honored product of methodical MDS get entry to most dependable convertible canons for the merge governance wherein r(I) = r(F), results in a fully inordinate area size demand. in this chapter, we take into account a conception of this preliminarily satisfactory-given product. the brand new creation is still grounded on canons with Vandermonde equality matrices, but we allow the scalars to attack any stupendous nonzero values, as opposed to being limited to successive powers of a primitive detail inside the subject. via distinctive point of the original and final equality matrices being Vandermonde matrices over the same set of scalars, the new creation of convertible canons remain success-most effective. It follows that life of any k × r notable-normal Vandermonde matrix over the sector F(q) yields(n(I), k(I); n(F), k(F) = λk(I)) methodical MDS get entry to most effective convertible canons over F(q) for any λ≥ 2, kF ≤ k, and r(I) = r(F) ≤ r. therefore, in this ruin, we will observe the abecedarian limits on the sphere sizes that make sure the cultures of ok × r tremendous-ordinary Vandermonde matrices. we're suitable to establish an life situation, a drop bound for fields of function, and a general upper sure on the minimum subject length that guarantees the cultures of k × r first rate-regular Vandermonde matrices.

We start with a result which offers a demand on the sector sizes over which exquisite- everyday Vandermonde matrices live. This end result attracts upon instinct that an most applicable choice of scalars for the Vandermonde matrix would keep down from opting rudiments with lower order to avoid reiteration alongside the corresponding columns.

Theorem 3.1: As presented, outlines a condition for the existence of a k × r atrocious-ordinary Vandermonde matrix over the field F(q).It states that such a matrix can only exist if a certain condition holds for each divisor ‘m’ of (q-1), where m < k

And q ≥ r ^ m.

Proof: We aim to show that if m<k, and there exists a multiplicative subgroup G of order m, then:

- Any Vandermonde matrix built using elements from G will have rank at most m,
- Therefore, a full-rank k × r requires **more than** m linearly independent rows, hence m ≥ k,
- Therefore, if m < k, to avoid such low-rank constructions (which would violate the matrix's desired properties), the number of elements in F(q) must be large enough to avoid these dependencies.

Let’s proceed:

Step 1: Structure of Subgroups
Let m divide q−1, so there is a unique multiplicative subgroup G of order m in $F^*(q)$.
If all α(i) lie in G, then the matrix entries $\alpha\,{}^{i}_{j}$ are limited to at most ‘m’ distinct values, and the row space of the matrix is constrained. That means any matrix formed using such α(i) has **rank at most ‘m’**.
So to construct a full-rank Vandermonde matrix of **more than ‘m’ rows**, not all the α(i) can lie in a subgroup of order ‘m’.

Step 2: Combinatorial Bound
We consider how many **distinct sets** of ‘k’ linearly independent elements can be used to construct the matrix. For safety, to avoid all α(i) lying in any small subgroup (i.e., those of order m < k), we must ensure that the total number of such "safe" elements exceeds the total number of ways to choose ‘r’ columns.
That is, the field must be large enough to offer **sufficiently diverse elements**.
In particular, if a subgroup ‘G’ of order ‘m’ exists, then the number of possible distinct k-tuples over ‘G’ is at most m ^ k, and since k > m, these configurations cannot yield full-rank matrices.
Therefore, the total number of distinct columns ‘r’ (i.e., monomials up to r−1) must satisfy:

$$q \geq r \wedge m \qquad (3.1)$$

to ensure enough distinct values when evaluated at ‘m’ elements.

CHAPTER 4

Conclusion

Code conversion provides a theoretical frame to model the trouble of redundancy edition, a large undertaking to utmost massive- scale cluster storehouse systems. Convertible canons are a order of especially designed canons that enable effective conversion indeed as maintaining favored decodability constraints. The access cost of conversion represents the entire volume of symbols examine or written during the conversion procedure, which corresponds to the wide variety of disks penetrated in the device for the conversion fashion. also, the merge governance is an essential class of transformations which involve incorporating multiple devices below an (n(I), k(I)) primary law C(I) into a veritably last metaphor below an (n(F), k(F)) veritably last law C(F).

In this oils, we take a look at methodical MDS access- optimum convertible canons within the merge governance, wherein the form situations before than and after incorporating are identical (r(I) = r(F)). before structures for similar canons wished extraordinarily big discipline sizes. We ameliorate in this with the aid of conducting the excellent- recognised upper bounds on the sphere length needed, using constructions grounded on brilliant-regular Vandermonde matrices. We first set up conditions for the actuality of those matrices, along with lower and upper bounds on the field length. This drastically reduces the sphere length wanted as compared to former oils. We also present, for the primary time, express constructions of methodical MDS get entry to-most useful convertible canons in the merge governance with nearly usable area sizes. especially, for any high strength field F(q), we give express constructions while k ^ F, the operation of area automorphisms to pick the matrix scalars. whilst the field has characteristic 2, we expand these issues to the case k ^ F methodical MDS get right of entry to-most excellent convertible canons inside the merge governance, where the restore ranges earlier than and after incorporating are same (r(I) = r(F)). former structures for similar canons needed extraordinarily massive discipline sizes.

We ameliorate this by means of supplying the first- rate- regarded upper bounds on the sector length needed, the use of structures grounded completely on extraordinary-everyday Cauchy matrices. We first set up conditions for the life of these matrices, at the side of drop and advanced bounds at the minimum discipline length demanded. This permits for a enormous reduction in field length compared to in advance results. We also gift, for the primary time, specific constructions of methodical MDS get admission to- premiere convertible canons in the merge governance with nearly usable discipline sizes. especially, for any top power discipline F q, we give specific structures when k ^ F, using the shape of Cauchy matrices. when the sphere has point 2, we increase those structures to the case in which k ^ F.

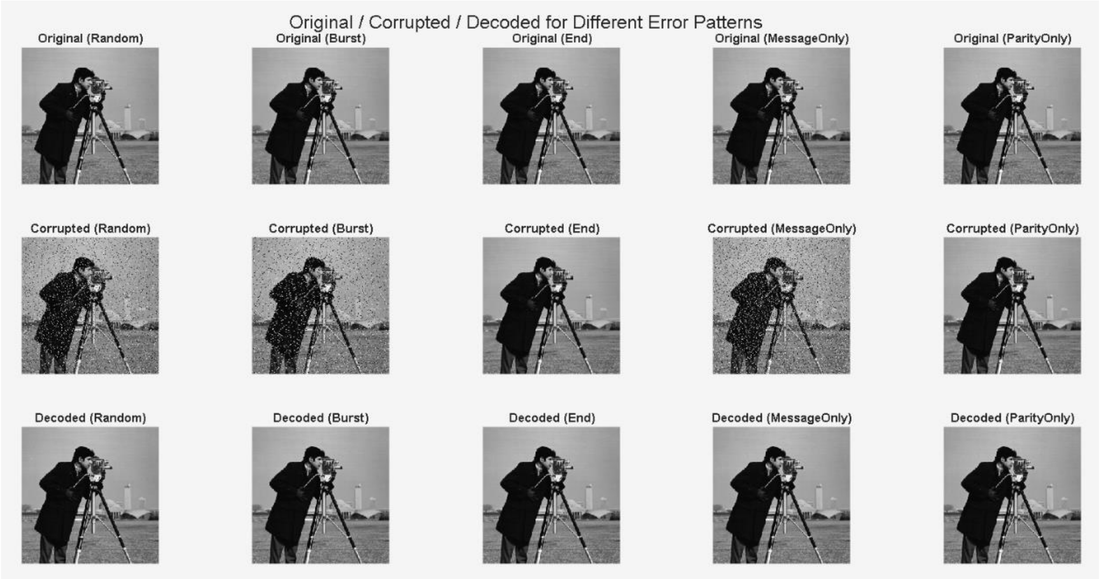


Fig. (4.1)

Reed-Solomon (RS) decoding using the PGZ decoder, along with image recovery under varying noise levels. Let's expand on the ideas and connect them with **cyclic shifts** and **super-regularity in the parity part of the matrix**, which are crucial concepts in designing efficient error correction codes and improving decoding performance.

Visual Demonstration of Error Correction

- **Images** are shown in three columns: Original, Corrupted (with varying probabilities p), and Reconstructed.
- As p increases, the image gets increasingly noisy.
- **Reconstruction** via error correction (using RS decoding) effectively recovers the original image, even at moderate noise levels.

PGZ Reed-Solomon Decoder – Mathematical Formulation

- **Syndrome Computation**: Computes error syndromes from received codeword.
- **Error Locator Polynomial**: Solved via a system of linear equations using a syndrome matrix.

Error Magnitude Evaluation

- Determines the **error positions** and computes their magnitudes.
- Final recovery is done using $r^{*}(j) = r(j) - Y(j)$, correcting the corrupted codeword.

2. Right Cyclic Shifts in Matrix Construction

A **right cyclic shift** refers to shifting each row or column of a matrix to the right, wrapping around the end. When used in constructing generator or parity-check matrices in Reed-Solomon or LDPC-type codes, it offers:

Advantages:

- **Cyclic codes** (a subclass of linear block codes) are invariant under cyclic shifts — they simplify encoder/decoder design.
- Facilitates **FFT-based** encoding/decoding due to circulant structures.
- Right cyclic shifts help generate **structured matrices** with desirable properties.

3. Super-Regularity in the Parity Part of the Matrix

A **super-regular matrix** is one where **all square submatrices are non-singular** (i.e., invertible). This property is highly desirable for:

Error Correction:

- Guarantees that the decoding system of equations (like in the PGZ method) is always solvable.
- Prevents linear dependency among rows/columns, improving **error localization**.

Implementation:

- Ensures uniqueness and stability in solving for error locator polynomials and magnitudes.
- Often used in **convolutional codes**, **network coding**, and **space-time coding**.

4. Combining Cyclic Shifts & Super-Regularity in Reed-Solomon

How They Work Together:

- **Parity Matrix (P)** can be constructed using **cyclic right shifts** of a base row, forming a **circulant** or **Toeplitz-like** structure.
- By carefully choosing this base row (e.g., derived from a primitive polynomial over F(2m)), the matrix can be **super-regular**.

Example:

Let G=[Ik|P] be the generator matrix, where:

- I(k) is the identity matrix (systematic part).
- P is a parity matrix built by right cyclic shifts of a primitive sequence over F(2m).

This ensures:

- **Structured and compact representation**.

- **Efficient encoding/decoding using matrix algebra**.
- **Resilience to burst errors and erasures**, especially useful in image transmission.

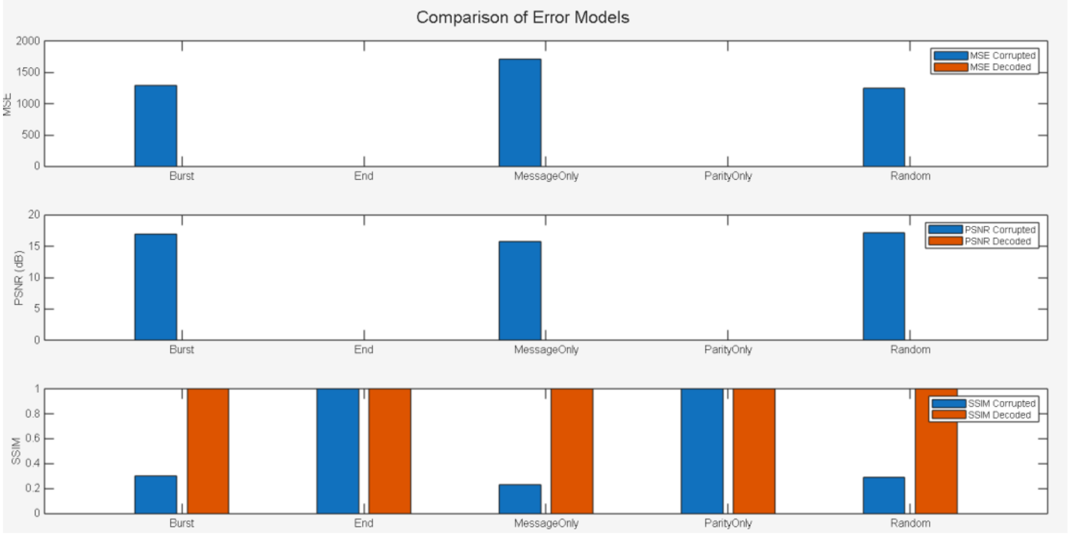
Application to Image Transmission

Using RS codes with parity matrices constructed from cyclic shifts and satisfying super-regularity:

- Makes the decoder more robust against localized (burst) errors typical in images.
- Enhances image reconstruction quality, as demonstrated in your corrupted/recovered image pairs.
- Efficient for **hardware implementations** due to regular structure.
- The PGZ decoder allows for efficient decoding of RS codes by solving a linear system based on syndromes.
- **Right cyclic shifts** create a **structured parity matrix**, making encoding and

decoding faster and more hardware-friendly.

- Enforcing **super-regularity** in the matrix ensures robust error correction even with high noise (as in image recovery).
- The synergy of these techniques yields powerful error-correcting capabilities, ideal for multimedia and communication systems.



REFERENCES

[1]Francisco Maturana and K. V. Rashmi. Bandwidth Cost of Code Conversions in Distributed Storage: Fundamental Limits and Optimal Constructions. IEEE Transactions on Information Theory, 69(8):4993–5008, 2023. doi: 10.1109/TIT.2023.3265512. 1

[2]Francisco Maturana and K. V. Rashmi. Locally Repairable Convertible Codes: Erasure Codes for Efficient Repair and Conversion. In 2023 IEEE International Symposium on Information Theory (ISIT), pages 2033–2038, 2023. doi: 10.1109/ISIT54713.2023.10206604. 2.5

[3]Francisco Maturana and K. V. Rashmi. Convertible codes: enabling efficient conversion of coded data in distributed storage. IEEE Transactions on Information Theory, 68:4392–4407, 2022. ISSN 1557-9654. doi: 10.1109/TIT.2022.3155972. 1, 2.2, 2.3, 2.1, 2.2, 2.3, 2.3, 2.4, 2.3, 2.4, 2.5, 3, 3

[4]Francisco Maturana and K. V. Rashmi. Bandwidth Cost of Code Conversions in the Split Regime. In 2022 IEEE International Symposium on Information Theory (ISIT), pages 3262–3267, 2022. doi: 10.1109/ISIT50566.2022.9834604. 1, 2.5

[5]Saurabh Kadekodi, K. V. Rashmi, and Gregory R. Ganger. Cluster storage systems gotta have HeART: improving storage efficiency by exploiting disk-reliability heterogeneity. In Arif Merchant and Hakim Weatherspoon, editors, 17th USENIX Conference on File and Storage Technologies, FAST 2019, Boston, MA, February 25-28, 2019, pages 345–358. USENIX Association, 2019. 1