A construction of Maximally Recoverable LRCs for small number of local groups

Manik Dhar* Sivakanth Gopi†

Abstract

Maximally Recoverable Local Reconstruction Codes (LRCs) are codes designed for distributed storage to provide maximum resilience to failures for a given amount of storage redundancy and locality. An (n, r, h, a, g)-MR LRC has n coordinates divided into g local groups of size $r = n/g$, where each local group has 'a' local parity checks and there are an additional 'h' global parity checks. Such a code can correct 'a' erasures in each local group and any h additional erasures. Constructions of MR LRCs over small fields is desirable since field size determines the encoding and decoding efficiency in practice. In this work, we give a new construction of (n, r, h, a, g)-MR-LRCs over fields of size $q = O(n)^{h+(g-1)a-\lceil h/g \rceil}$ which generalizes a construction of Hu and Yekhanin (ISIT 2016). This improves upon state of the art when there are a small number of local groups, which is true in practical deployments of MR LRCs.

1 Introduction

In modern distributed storage systems, data is split and stored in individual servers. A server crash can potentially lead to the loss of all data in a server. Even if a server becomes temporarily unavailable, for instance due to serving too many requests, that could lead to very slow access to data on that server. Replication of data is a potential solution but that is very inefficient in terms of storage. Error correcting codes offer a more efficient solution. For example, distributed storage systems such as RAID use Reed-Solomon codes. An (n, k)-Reed Solomon code will add $n-k$ redundant servers (parity checks) to k data servers and allows us to recover from an arbitrary $n-k$ erasures by reading the remaining k servers. But for large k, which is needed to get good storage efficiency, this would require us to read a lot of servers to recover lost data.

Local reconstruction codes (LRCs) were invented to deal with this problem. *Locality* means that when a small number of servers fail, any failed server can be recovered quickly by reading data from a small number of healthy servers. At the same time they can recover from catastrophic failures where a large number of servers fail (although recovery will necessarily be less efficient). Locality in distributed storage was first introduced in [HSX+12, CHL07], but LRCs were first formally defined and studied in [GHSY12] and [PD12].

Definition 1.1. *An (n, r, h, a, g)-LRC is a linear code C over $\mathbb{F}_q$ of length n (in other words a subspace $C \subseteq \mathbb{F}_q^n$), whose codeword symbols are partitioned into g local groups each of size $r = n/g$.*

*Department of Computer Science, Princeton University. Email: manikd@princeton.edu. Part of this work was done while this author was working at Microsoft research. Research supported by NSF grant DMS-1953807.

†Microsoft Research. Email: sigopi@microsoft.com.

The coordinates in each local group satisfy 'a' local parity checks (in other if the codewords are column matrices then the co-ordinates in a group lie in the kernel of an $a \times r$ matrix) and there are further h global parity checks that all the n coordinates satisfy (in other words the codewords lie in the kernel of an $h \times n$ matrix).

The local parity checks are used to recover from up to a erasures in a local group by reading at most $r - a$ symbols in that local group. The h global parities are used to correct more global erasure patterns which involve more than a erasures in each local group.

The above definition implies that the parity check matrix H of an (n, r, h, a, g)-LRC has the structure shown in Equation 1.

$$H = \left[\begin{array}{c|c|c|c} A_1 & 0 & \cdots & 0 \\ \hline 0 & A_2 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & A_g \\ \hline B_1 & B_2 & \cdots & B_g \end{array}\right] \tag{1}$$

Recall that g is the number of local groups and each local group has size $r = n/g$. $A_1, A_2, \ldots, A_g$ are $a \times r$ matrices over $\mathbb{F}_q$ which correspond to the local parity checks that each local group satisfies. $B_1, B_2, \ldots, B_g$ are $h \times r$ matrices over $\mathbb{F}_q$ and together they represent the h global parity checks that the codewords should satisfy.

We are interested in LRCs which can correct as many erasures as possible. The best one could hope for is a set of values for entries in H (as shown in (1)) which can correct any set of erasures a generic matrix of that form can (one way to define a generic H is to assume every entry is algebraically independent of each other). We know that a set of erasures E can be recovered from if and only if the subset of columns of H corresponding to E have full rank. Using this it can be shown that a generic H can correct an additional h erasures distributed across local groups on top of the 'a' erasures in each local group. LRCs which can correct all such erasure patterns are called *Maximally Recoverable (MR) LRCs*. Maximal recoverability was first introduced by [CHL07, HSX+12] and extended to more general settings in [GHJY14]. MR-LRCs were also studied earlier in [BHH12] where they are called *Partial-MDS (Maximum Distance Separable) codes*.

For clarity, we formally define MR-LRCs.

Definition 1.2. *Let C be an arbitrary (n, r, h, a, g)-LRC where $r = n/g$. We say that C is maximally recoverable if:*

1. *Any set of 'a' erasures in a local group can be corrected by reading the rest of the $r - a$ symbols in that local group.*
2. *Any erasure pattern $E \subseteq [n]$, $|E| = ga + h$, where E is obtained by selecting a symbols from each of g local groups and h additional symbols arbitrarily, is correctable by the code C.*

Again, our discussion gives us the following characterization for the parity check matrices of MR-LRCs.

Proposition 1.3. *An (n, r, h, a, g)-LRC with parity check matrix given by H from Equation 1 is maximally recoverable iff:*

1. *Each of the local parity check matrices A_i are the parity check matrices of an MDS code, i.e., any a columns of A_i are linearly independent.*

2. *Any submatrix of H which can be formed by selecting a columns in each local group and additional h columns has full column rank.*

Practical deployments of MR LRCs typically have a small number of local groups (say $g = 2, 3, 4$) and a small number of local parities per local group (say $a = 1, 2$) [HSX+12]. Moreover, field size is the most important determinant of encoding and decoding efficiency of MR LRCs [HSX+12], since encoding and decoding requires several finite field operations. Therefore constructions of MR LRCs over small fields in this regime are important in practice.

In this work, we give a new construction of MR LRC tailored for this regime.

Theorem 1.4. *There exists an explicit (n, r, h, a, g)-MR-LRC over a field of size $O(n)^{h+(g-1)a-\lceil h/g \rceil}$.*

For the $a > 1$ and small g parameter regime, the general construction of [GG22] shown in (2) is the only one we need to compare to. Here we see that for small constant g, which means $r = \Omega(n)$, as long as $h - \lceil h/g \rceil + a(g-1) < \min\{h, r-a\}$ our construction gives us a better MR-LRC construction. In particular, when $g = 2$, h is even, and a is a constant our construction is over fields of size $O(n)^{h/2+a}$ as opposed to $O(n)^h$ from [GG22].

Our construction generalizes a construction from [HY16] which is specialized for $a = 1$ and requires a field size of $n^{h-\lceil h/g \rceil+1}$. While the proof in [HY16] can also be generalized, this paper analyses the generalization with a new and direct proof using simple linear algebra and basic properties of Vandermonde matrices and Gabidulin codes. This also gives a new proof for the original result of [HY16]. Our construction and its proof is inspired by the constructions of MR LRCs in [GG22, CMST21].

1.1 Prior Work

Upper Bounds: There are a several construction of MR LRCs since their introduction [BPSY16, BCT22, GHJY14, MPK19, GJX20, GYBS18, MP22, LX22, GGY20, CSYS15]. The current best constructions of MR-LRCs over most range of parameters is due to [GG22, CMST21] which require a field size of

$$\big(O\big(\max\{n/r, r\}\big)\big)^{\min\{h, r-a\}}. \tag{2}$$

These codes are constructed using the theory of skew-symmetric polynomials. As mentioned earlier our construction provides better field sizes in the case of constant g. For g constant, constant a, and h divisible by g we have $r = \Omega(n)$ which means our construction gives a field size of $O(n)^{h(1-1/g)+a}$ and 2 gives $O(n)^h$. Our construction does better as long as $a < h/g$. For more constrained settings much better bounds are known as shown in Table 1.

When g and r are powers of a prime p and $h \pmod p \neq 1$ and $\lceil h/g \rceil \pmod p \neq p-1$ then [HY16] can shave off a factor of n to give $(n, r, h, a = 1, g)$-MR LRCs over fields of size $O(n)^{h-\lceil h/g \rceil+g-2}$ (when $p = g = 2$ this gives the third row of our table).

Lower bounds: The best known lower bounds on the field size required for (n, r, h, a, g)-MR LRCs (with $gr = n$) is from [GGY20] who show that for $h \geq 2$,

$$q \geq \Omega_{h,a}\left(n \cdot r^{\alpha}\right) \text{ where } \alpha = \frac{\min\{a, h - 2\lceil h/g \rceil\}}{\lceil h/g \rceil}. \tag{3}$$

Table 1: Table showing the best known upper bounds on the field size of (n, r, h, a, g)-MR LRCs over several constrained settings.

$O(r)$ when $h = 0$ or $h = 1$	[BHH12]
$O(n)^{h-\lceil h/g \rceil + g - 1}$ when $a = 1$	[HY16]
$O(n)^{h/2}$ when $a = 1, g = 2$ and $h \pmod 4 = 0$	[HY16]
$O(n)$ when $h = 2$	[GGY20]
$O(n^3)$ when $h = 3$	[GGY20]
$\widetilde{O}(n)$ when $h = 3, a = 1, r = 3$	[GGY20]
$\big(O(n)\big)^{\lceil \min\{h, r-1\}(1-1/q_0) \rceil}$ when $a = 1$ and $q_0 \geq g + 1$ is a prime power	[GG22]

When $g \geq h$, this simplifies to $q \geq \Omega_{h,a} n \cdot r^{\min\{a, h-2\}}$. If $g < h$, in particular when g divides h, then the lower bound simplifies to,

$$q \geq \Omega_{h,a} \left(n r^{\min\{ag/h, g-2\}} \right) \tag{4}$$

If in particular we look at $g = 2$, we see that the current lower bound is linear while the current best constructions for $a = 1$ and h divisible by 4 is $n^{h/2}$ [HY16]. This shows that there is a large scope of improvement in either direction. Any progress in this question is very interesting because as mentioned earlier the regime of small number of groups is important in practice for distributed storage [HSX+12].

2 Proof of Theorem 1.4

We are going to use a simple property of Moore matrices which are used to construct Gabidulin codes [KG05, GS21].

Lemma 2.1. *Let $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n \in \mathbb{F}_{q^m}$ be linearly independent over $\mathbb{F}_q$ (this requires $n \leq m$). Then the following Moore matrix M has full column rank.*

$$M = \begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_n \\ \boldsymbol{\beta}_1^q & \boldsymbol{\beta}_2^q & \cdots & \boldsymbol{\beta}_n^q \\ \boldsymbol{\beta}_1^{q^2} & \boldsymbol{\beta}_2^{q^2} & \cdots & \boldsymbol{\beta}_n^{q^2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\beta}_1^{q^{m-1}} & \boldsymbol{\beta}_2^{q^{m-1}} & \cdots & \boldsymbol{\beta}_n^{q^{m-1}} \end{bmatrix}$$

We will also need the concept of Schur complement.

Lemma 2.2. *Consider the square matrix*

$$M = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

where A, D are square matrices and A is invertible. Then doing column operations to remove the columns of B using the columns of A will result in the following matrix[*]

$$M' = \left[\begin{array}{c|c} A & 0 \\ \hline C & D - CA^{-1}B \end{array}\right].$$

In the remainder of this section, we present the proof of Theorem 1.4 using simple linear algebra and the above lemma. Let $t = a + \lceil h/g \rceil$ and $m = h + ga - t$. Let $q \geq n$ be a prime power. We will construct a parity check matrix over the field $\mathbb{F}_{q^m}$. This means our field size will be $q^m = q^{h+ga-t} = q^{h+(g-1)a-\lceil h/g \rceil}$ as desired.

We partition $\mathbb{F}_q$ into g sets $\{x_{i,1}, \ldots, x_{i,r}\}$ of size r and some left over elements which we ignore.

Recall that we want to construct a parity check matrix of the form

$$H = \left[\begin{array}{c|c|c|c} A_1 & 0 & \cdots & 0 \\ \hline 0 & A_2 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & A_g \\ \hline B_1 & B_2 & \cdots & B_g \end{array}\right] \tag{5}$$

where each A_i is an $a \times r$ matrix and each B_i is an $h \times r$ matrix. Define A_i as

$$A_i = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_{i,1} & x_{i,2} & \ldots & x_{i,r} \\ x_{i,1}^2 & x_{i,2}^2 & \ldots & x_{i,r}^2 \\ \vdots & \vdots & & \vdots \\ x_{i,1}^{a-1} & x_{i,2}^{a-1} & \ldots & x_{i,r}^{a-1} \end{bmatrix}.$$

Define $\boldsymbol{\beta}_{i,j} \in \mathbb{F}_{q^m}$ for $i \in [g], j \in [r]$ as

$$\boldsymbol{\beta}_{i,j} = \begin{bmatrix} x_{i,1}^t \\ x_{i,1}^{t+1} \\ \vdots \\ x_{i,1}^{t+m-1} \end{bmatrix},$$

where we are expressing $\boldsymbol{\beta}_{i,j}$ in some basis for $\mathbb{F}_{q^m}$ (which is a $\mathbb{F}_q$-vector space of dimension m). We now define B_i as

$$B_i = \begin{bmatrix} x_{i,1}^a & x_{i,2}^a & \ldots & x_{i,r}^a \\ x_{i,1}^{a+1} & x_{i,2}^{a+1} & \ldots & x_{i,r}^{a+1} \\ \vdots & \vdots & & \vdots \\ x_{i,1}^{t-1} & x_{i,2}^{t-1} & \ldots & x_{i,r}^{t-1} \\ \boldsymbol{\beta}_{i,1} & \boldsymbol{\beta}_{i,2} & \ldots & \boldsymbol{\beta}_{i,r} \\ \boldsymbol{\beta}_{i,1}^q & \boldsymbol{\beta}_{i,2}^q & \ldots & \boldsymbol{\beta}_{i,r}^q \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\beta}_{i,1}^{q^{h+a-t-1}} & \boldsymbol{\beta}_{i,2}^{q^{h+a-t-1}} & \ldots & \boldsymbol{\beta}_{i,r}^{q^{h+a-t-1}} \end{bmatrix}.$$

[*]The matrix $D - CA^{-1}B$ is called Schur complement.

For convenience we collect the first $t-a$ rows of B_i into V_i and the remaining rows into G_i so that

$$B_i = \left[\frac{V_i}{G_i}\right]. \tag{6}$$

We see that V_i contains the 'Vandermonde'-like rows and G_i contains the 'Gabidulin'-like rows. Also note that the powers of $x_{i,j}$ are increasing steadily from 0 to $t+m-1$ (which is equal to $ag+h-1$); first along A_i, then V_i and then the first row of G_i (when expressed in $\mathbb{F}_q$ basis). This is the most crucial part of the construction as we will shortly see. The MR LRC constructions from [GG22, CMST21] also use a similar trick.

Given a matrix M and a subset I of its columns, we let $M(I)$ refer to the sub-matrix of M corresponding to the columns I. We also use $\boldsymbol{\beta}_i$ to denote the $m \times r$ matrix formed by $\boldsymbol{\beta}_{i,1}, \ldots, \boldsymbol{\beta}_{i,r}$,

$$\boldsymbol{\beta}_i = \begin{bmatrix}\boldsymbol{\beta}_{i,1} & \boldsymbol{\beta}_{i,2} & \ldots & \boldsymbol{\beta}_{i,r}\end{bmatrix}.$$

Let E be an erasure pattern of size $ag+h$ formed by selecting a columns in each of the local groups and additional h columns from anywhere in H as shown in (5). We want to show that $H(E)$ (which is an $(ag+h)\times(ag+h)$ matrix) is full rank. Showing $H(E)$ is full rank for every correctable erasure pattern E will prove our theorem.

One of the local groups will contain at least $t = a + \lceil h/g \rceil$ many columns. Without loss of generality let us say that is group g. We arbitrarily split the columns in group g into X which is of size t and Y which contains the remaining columns. For groups $1, \ldots, g-1$ we arbitrarily split the columns selected in each group to a set S_i of size a and T_i which contains the remaining elements chosen. So we can write $H(E)$ as

$$H(E) = \left[\begin{array}{c|c|c|c} A_1(S_1 \cup T_1) & \cdots & 0 & 0 \\ \hline \vdots & \ddots & \vdots & \vdots \\ \hline 0 & \cdots & A_{g-1}(S_{g-1} \cup T_{g-1}) & 0 \\ \hline 0 & \cdots & 0 & A_g(X \cup Y) \\ \hline V_1(S_1 \cup T_1) & \cdots & V_{g-1}(S_{g-1} \cup T_{g-1}) & V_g(X \cup Y) \\ \hline G_1(S_1 \cup T_1) & \cdots & G_{g-1}(S_{g-1} \cup T_{g-1}) & G_g(X \cup Y) \end{array}\right].$$

Now we note that $A_1(S_1), A_2(S_2), \cdots, A_{g-1}(S_{g-1})$ are $a \times a$ matrices of full rank. We now do column operations on $H(E)$, where in the first $g-1$ local groups we use the columns of $A_i(S_i)$ for $i \le g-1$ to remove the columns of $A_i(T_i)$. This results in the lower block $\left[\frac{V_i(T_i)}{G_i(T_i)}\right]$ to change into a Schur complement as follows:

$$\left[\begin{array}{c|c} A_i(S_i) & A_i(T_i) \\ \hline V_i(S_i) & V_i(T_i) \\ \hline G_i(S_i) & G_i(T_i) \end{array}\right] \to \left[\begin{array}{c|c} A_i(S_i) & 0 \\ \hline V_i(S_i) & V_i(T_i) - V_i(S_i)A_i(S_i)^{-1}A_i(T_i) \\ \hline G_i(S_i) & G_i(T_i) - G_i(S_i)A_i(S_i)^{-1}A_i(T_i) \end{array}\right],$$

for all $i \le g-1$. For convenience we let $G_i'(T_i) = G_i(T_i) - G_i(S_i)A_i(S_i)^{-1}A_i(T_i)$ and $V_i'(T_i) = V_i(T_i) - V_i(S_i)A_i(S_i)^{-1}A_i(T_i)$ for $i \le g-1$.

At the end of this $H(E)$ transforms into the following:

$$M_1 = \left[\begin{array}{c|c|c|c|c|c} A_1(S_1) & 0 & \cdots & 0 & 0 & 0 \\ \hline \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \hline 0 & 0 & \cdots & A_{g-1}(S_{g-1}) & 0 & 0 \\ \hline 0 & 0 & \cdots & 0 & 0 & A_g(X \cup Y) \\ \hline V_1(S_1) & V_1'(T_1) & \cdots & V_{g-1}(S_{g-1}) & V_{g-1}'(T_{g-1}) & V_g(X \cup Y) \\ \hline G_1(S_1) & G_1'(T_1) & \cdots & G_{g-1}(S_{g-1}) & G_{g-1}'(T_{g-1}) & G_g(X \cup Y) \end{array}\right].$$

It is clear that it suffices to show the following $(h+a) \times (h+a)$ sub-matrix of M_1 is full rank.

$$M_2 = \left[\begin{array}{c|c|c|c|c} 0 & \cdots & 0 & A_g(Y) & A_g(X) \\ \hline V_1'(T_1) & \cdots & V_{g-1}'(T_{g-1}) & V_g(Y) & V_g(X) \\ \hline G_1'(T_1) & \cdots & G_{g-1}'(T_{g-1}) & G_g(Y) & G_g(X) \end{array}\right].$$

In M_2 we look at the sub-matrix $W(X) = \left[\frac{A_g(X)}{V_g(X)}\right]$ (we also let $W(Y) = \left[\frac{A_g(Y)}{V_g(Y)}\right]$). By construction $W(X)$ is a $t \times t$ Vandermonde matrix of full rank. We use $W(X)$ to remove all the remaining columns in M_3 corresponding to the rows of $W(X)$ by doing column column operations giving us the following matrix:

$$M_3 = \left[\begin{array}{c|c|c|c|c} 0 & \cdots & 0 & 0 & A_g(X) \\ \hline 0 & \cdots & 0 & 0 & V_g(X) \\ \hline G_1''(T_1) & \cdots & G_{g-1}''(T_{g-1}) & G_g(Y) - G_g(X)W(X)^{-1}W(Y) & G_g(X) \end{array}\right],$$

where $G_i''(T_i) = G_i'(T_i) - G_g(X)W(X)^{-1}\left[\frac{0}{V_i'(T_i)}\right]$ for $i \leq g-1$. For convenience, we also let $V_i''(T_i) = \left[\frac{0}{V_i'(T_i)}\right]$.

It is now clear that it suffices to show the following $(h+a-t) \times (h+a-t)$ sub-matrix of M_3 is full rank.

$$M_4 = \left[\begin{array}{c|c|c|c} G_1''(T_1) & \cdots & G_{g-1}''(T_{g-1}) & G_g(Y) - G_g(X)W(X)^{-1}W(Y) \end{array}\right],$$

Note that all the entries in $W(X), W(Y), V'(T_i), A_i(S_i)$, and $A_i(T_i)$ are in the base field $\mathbb{F}_q$. Column operations on G_i with $\mathbb{F}_q$ coefficients retain its structure with $\boldsymbol{\beta}$'s replaced by their corresponding $\mathbb{F}_q$-linear combinations. Therefore by Lemma 2.1, it is enough to show that the following $m \times (h+a-t)$ matrix has full column rank[†]

$$M_5 = \left[\begin{array}{c|c|c|c} \boldsymbol{\beta}_1''(T_1) & \cdots & \boldsymbol{\beta}_{g-1}''(T_{g-1}) & \boldsymbol{\beta}_g(Y) - \boldsymbol{\beta}_g(X)W(X)^{-1}W(Y) \end{array}\right],$$

where $\boldsymbol{\beta}_i''(T_i) = \boldsymbol{\beta}_1(T_i) - \boldsymbol{\beta}_1(S_i)A_i(S_i)^{-1}A_i(T_i) - \boldsymbol{\beta}_g(X)W(X)^{-1}V_i''(T_i)$ for $i \leq g-1$. Now consider the following $(h+ga) \times (h+ga)$ matrix F,

$$F = \left[\begin{array}{c|c|c|c|c|c|c} A_1(S_1) & A_1(T_1) & \cdots & A_{g-1}(S_{g-1}) & A_{g-1}(T_{g-1}) & A_g(Y) & A_g(X) \\ \hline V_1(S_1) & V_1(T_1) & \cdots & V_{g-1}(S_{g-1}) & V_{g-1}(T_{g-1}) & V_g(Y) & V_g(X) \\ \hline \boldsymbol{\beta}_1(S_1) & \boldsymbol{\beta}_1(T_1) & \cdots & \boldsymbol{\beta}_{g-1}(S_{g-1}) & \boldsymbol{\beta}_{g-1}(T_{g-1}) & \boldsymbol{\beta}_g(Y) & \boldsymbol{\beta}_g(X) \end{array}\right].$$

[†]Note that $\sum_{i=1}^{g-1} |T_i| + |Y| = h + a - t$.

By construction, F is a Vandermonde matrix over $\mathbb{F}_q$ and therefore is of full rank. We now repeat column operations on F analogous to the ones we did on $H(E)$. Concretely, we first use the $a \times a$ invertible matrix $A_i(S_i)$ to clear out the columns $A_i(T_i)$ for $i \leq g-1$. This gives us the matrix,

$$F_1 = \left[\begin{array}{c|c|c|c|c|c|c} A_1(S_1) & 0 & \cdots & A_{g-1}(S_{g-1}) & 0 & A_g(Y) & A_g(X) \\ \hline V_1(S_1) & V_1'(T_1) & \cdots & V_{g-1}(S_{g-1}) & V_{g-1}'(T_{g-1}) & V_g(Y) & V_g(X) \\ \hline \boldsymbol{\beta}_1(S_1) & \boldsymbol{\beta}_1'(T_1) & \cdots & \boldsymbol{\beta}_{g-1}(S_{g-1}) & \boldsymbol{\beta}_{g-1}'(T_{g-1}) & \boldsymbol{\beta}_g(Y) & \boldsymbol{\beta}_g(X) \end{array}\right]$$

where $\boldsymbol{\beta}_i'(T_i) = \boldsymbol{\beta}_1(T_i) - \boldsymbol{\beta}_1(S_i)A_i(S_i)^{-1}A_i(T_i)$ for $i \leq g-1$. Next we use the sub-matrix $W(X) = \left[\frac{A_g(X)}{V_g(X)}\right]$ to clear out the columns $W(Y) = \left[\frac{A_g(Y)}{V_g(Y)}\right]$ and $V_i''(T_i) = \left[\frac{0}{V_i'(T_i)}\right]$. This gives us the matrix,

$$F_2 = \left[\begin{array}{c|c|c|c|c|c|c} A_1(S_1) & 0 & \cdots & A_{g-1}(S_{g-1}) & 0 & 0 & A_g(X) \\ \hline V_1(S_1) & 0 & \cdots & V_{g-1}(S_{g-1}) & 0 & 0 & V_g(X) \\ \hline \boldsymbol{\beta}_1(S_1) & \boldsymbol{\beta}_1''(T_1) & \cdots & \boldsymbol{\beta}_{g-1}(S_{g-1}) & \boldsymbol{\beta}_{g-1}''(T_{g-1}) & \boldsymbol{\beta}_g(Y) - \boldsymbol{\beta}_g(X)W(X)^{-1}W(Y) & \boldsymbol{\beta}_g(X) \end{array}\right].$$

As F_2 is a full rank square matrix, M_5 should have full column rank which implies $H(E)$ has full rank completing the proof.

References

[BCT22] Alexander Barg, Zitan Chen, and Itzhak Tamo. A construction of maximally recoverable codes. *Designs, Codes and Cryptography*, 90(4):939–945, 2022. Publisher: Springer.

[BHH12] Mario Blaum, James Lee Hafner, and Steven Hetzler. Partial-mds codes and their application to raid type of architectures. *IEEE Transactions on Information Theory*, 59:4510–4519, 2012.

[BPSY16] Mario Blaum, James S Plank, Moshe Schwartz, and Eitan Yaakobi. Construction of partial MDS and sector-disk codes with two global parity symbols. *IEEE Transactions on Information Theory*, 62(5):2673–2681, 2016. Publisher: IEEE.

[CHL07] Minghua Chen, Cheng Huang, and Jin Li. On the maximally recoverable property for multi-protection group codes. *2007 IEEE International Symposium on Information Theory*, pages 486–490, 2007.

[CMST21] Han Cai, Ying Miao, Moshe Schwartz, and Xiaohu Tang. A construction of maximally recoverable codes with order-optimal field size. *IEEE Transactions on Information Theory*, 68(1):204–212, 2021.

[CSYS15] Junyu Chen, Kenneth W Shum, Quan Yu, and Chi Wan Sung. Sector-disk codes and partial MDS codes with up to three global parities. pages 1876–1880. IEEE, 2015.

[GG22] Sivakanth Gopi and Venkatesan Guruswami. Improved maximally recoverable LRCs using skew polynomials. *IEEE Transactions on Information Theory*, 2022.

[GGY20] Sivakanth Gopi, Venkatesan Guruswami, and Sergey Yekhanin. Maximally recoverable LRCs: A field size lower bound and constructions for few heavy parities. *IEEE Transactions on Information Theory*, 66(10):6066–6083, 2020.

[GHJY14] Parikshit Gopalan, Cheng Huang, Bob Jenkins, and Sergey Yekhanin. Explicit maximally recoverable codes with locality. *IEEE Transactions on Information Theory*, 60(9):5245–5256, 2014.

[GHSY12] Parikshit Gopalan, Cheng Huang, Huseyin Simitci, and Sergey Yekhanin. On the locality of codeword symbols. *IEEE Transactions on Information theory*, 58(11):6925–6934, 2012.

[GJX20] Venkatesan Guruswami, Lingfei Jin, and Chaoping Xing. Constructions of maximally recoverable local reconstruction codes via function fields. *IEEE Transactions on Information Theory*, 66(10):6133–6143, 2020.

[GS21] Ernst M Gabidulin and Vladimir Sidorenko. *Rank codes*. 2021.

[GYBS18] Ryan Gabrys, Eitan Yaakobi, Mario Blaum, and Paul H Siegel. Constructions of partial MDS codes over small fields. *IEEE Transactions on Information Theory*, 65(6):3692–3701, 2018. Publisher: IEEE.

[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. pages 15–26, 2012.

[HY16] Guangda Hu and Sergey Yekhanin. New constructions of SD and MR codes over small finite fields. *2016 IEEE International Symposium on Information Theory*, pages 1591–1595, 2016.

[KG05] A. Kshevetskiy and E. Gabidulin. The new construction of rank codes. In *Proceedings. International Symposium on Information Theory, 2005*, pages 2105–2108, 2005.

[LX22] Shu Liu and Chaoping Xing. Maximally Recoverable Local Repairable Codes from Subspace Direct Sum Systems. *IEEE Transactions on Information Theory*, 2022. Publisher: IEEE.

[MP22] Umberto Martínez-Peñas. A general family of MSRD codes and PMDS codes with smaller field sizes from extended Moore matrices. *SIAM Journal on Discrete Mathematics*, 36(3):1868–1886, 2022.

[MPK19] Umberto Martínez-Peñas and Frank R Kschischang. Universal and dynamic locally repairable codes with maximal recoverability via sum-rank codes. *IEEE Transactions on Information Theory*, 65(12):7790–7805, 2019. Publisher: IEEE.

[PD12] Dimitris Papailiopoulos and Alexandros G. Dimakis. Locally repairable codes. *IEEE Transactions on Information Theory*, 60:5843–5855, 2012.