Electronic Colloquium on Computational Complexity, Report No. 183 (2017)

On Maximally Recoverable Local Reconstruction Codes

Sivakanth Gopi*
Princeton University
sgopi@cs.princeton.edu

Venkatesan Guruswami†
Carnegie Mellon Univ.
venkatg@cs.cmu.edu

Sergey Yekhanin
Microsoft Research
yekhanin@microsoft.com

Abstract

In recent years the explosion in the volumes of data being stored online has resulted in distributed storage systems transitioning to erasure coding based schemes. Local Reconstruction Codes (LRCs) have emerged as the codes of choice for these applications. An (n, r, h, a, q)-LRC is a q-ary code, where encoding is as a two stage process. In the first stage, h redundant parity symbols are generated from k data symbols. In the second stage, the $k+h$ symbols are partitioned into sets of size $r-a$ and each set is extended with a redundant symbols using an MDS code to form a local group. Local groups ensure that when at most a coordinates are erased, any missing coordinate can be recovered by accessing at most $r-a$ symbols. Also, if a larger number of coordinates is erased, the missing symbols can be recovered by potentially accessing all remaining symbols.

An (n, r, h, a, q)-LRC code as above is Maximally Recoverable (MR), if it corrects all erasure patterns which are information theoretically correctable given the presence of local groups. Obtaining MR LRCs over finite fields of minimal size is important in practice and has been the goal of a line of work in coding theory. In this work we make progress towards this goal. In particular:

- We show that when a and h are constant and r may grow, for every maximally recoverable LRC, $q \geqslant \Omega_{a,h}\left(n \cdot r^{\min\{a, h-2\}}\right)$. Prior to our work, there was no super-linear lower bound known on the field size of MR LRCs for any setting of parameters.
- We obtain a family of MR $(n, r, h = 2, a, q)$-LRCs, where $q = O(n)$ for all settings of parameters. Prior to our work the best constructions required q to be quadratic in n for some regimes.
- We obtain a family of MR $(n, r, h = 3, a, q)$-LRCs, where $q = O(n^3)$ for all settings of parameters. Prior to our work the best constructions required q to be $n^{\Theta(a)}$ for some regimes.
- Our results in the first two bullets above suggest the setting of $r = 3, a = 1, h = 3$ as the first setting where existence of MR LRCs over fields of near linear size is an open question. We resolve this question in the positive by developing a new approach to LRC constructions based on elliptic curves and arithmetic progression free sets.

*Research supported by NSF CAREER award 1451191 and NSF grant CCF-1523816. Most of this work was done when the author was visiting Microsoft Research.

†Research supported in part by NSF grant CCF-1563742. Most of this work was done during a visit by the author to Microsoft Research, Redmond. The work was also partly done when the author was visiting the School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore.

ISSN 1433-8092

1 Introduction

The explosion in the volumes of data being stored online means that duplicating or triplicating data is not economically feasible. This has resulted in distributed storage systems employing erasure coding based schemes in order to ensure reliability with low storage overheads. In recent years Local Reconstruction Codes (LRCs) emerged as the codes of choice for many such scenarios and have been implemented in a number of large scale systems e.g., Microsoft Azure [HSX+12] and Hadoop [SAP+13].

Classical erasure correcting codes [MS77] guarantee that data can be recovered if a bounded number of codeword coordinates is erased. However recovering data typically involves accessing all surviving coordinates. By contrast, Local Reconstruction Codes* (LRCs) distinguish between the typical case when only a small number of codeword coordinates are erased (e.g., few machines in a datacenter fail) and a worst case when a larger number of coordinates might be unavailable, and guarantee that in the prior case recovery of individual coordinates can be accomplished in sub-linear time, without having to access all surviving symbols.

LRCs are systematic linear codes, where encoding is a two stage process. In the first stage, h redundant heavy parity symbols are generated from k data symbols. Each heavy parity is a linear combination of all k data symbols. During the second stage, the $k+h$ symbols are partitioned into $\frac{k+h}{r-a}$ sets of size $r-a$ and each set is extended with a local parity symbols using an MDS code to form a *local group*. Encoding as above ensures that when at most a coordinates are erased, any missing coordinate can be recovered by accessing at most $r-a$ symbols. However, if a larger number of coordinates (that depends on h) is erased; then all missing symbols can be recovered by potentially accessing all remaining symbols.

Our description of LRC codes above is not complete. To specify a concrete code we need to fix coefficients in linear combinations that define h heavy and $\frac{k+h}{r-a} \cdot a$ local parities. Different choices of coefficients could lead to codes with different erasure correcting capabilities. The best we could hope for is to have an optimal choice of coefficients which ensures that our code can correct every pattern of erasures that is correctable for some setting of coefficients. Such codes always exist and are called Maximally Recoverable (MR) [CHL07, HCL07] LRCs.† Combinatorially, an (n, r, h, a, q)-LRC is maximally recoverable it if corrects every pattern of erasures that can be obtained by erasing a coordinates in each local group and up to h additional coordinates elsewhere. Explicit constructions of MR LRCs are available (e.g., [CK17]) for all ranges of parameters. Unfortunately, all known constructions require finite fields of very large size.

Encoding a linear code and decoding it from erasures involve matrix vector multiplication and linear equation solving respectively. Both of these require performing numerous finite field arithmetic operations. Having small finite fields results in faster encoding and decoding and thus improves the overall throughput of the system [PGM13, Section 2]. It is also desirable in practice to work over finite fields of characteristic 2. Obtaining MR LRCs over finite fields of minimal size is one of the central problems in the area of codes for distributed storage.

1.1 State of the art and our results

We now summarize what is known about the minimal field size of maximally recoverable local reconstruction codes with parameters n, r, a and h and first cover the easy cases.

- When $a = 0$, LRCs are equivalent to classical erasure correcting codes. In this case Reed Solomon codes are maximally recoverable, and they have a field size of roughly n, which is known to be

*The term local reconstruction codes is from [HSX+12]. Essentially the same codes were called locally repairable codes in [PD14] and locally recoverable codes in [TB14]. Thankfully all names above abbreviate to LRCs.

†Maximally recoverable LRCs are called Partial MDS (PMDS) in [Bla13, BHH13] and many follow up works.

optimal up to constant factors [Bal12].

- When $h \leqslant 1$, there are constructions of maximally recoverable LRCs over fields of size $O(r)$ [BHH13] which is optimal.
- When $r = a + 1$, codes in the local groups are necessarily simple repetition codes. MR LRCs can be obtained by starting with a Reed Solomon code of length n/r and repeating every coordinate r times. Thus the optimal field size is $\Theta(n/r)$.

This leaves us with the main case, when $a \geqslant 1$, $r \geqslant a+2$, and $h \geqslant 2$. A number of constructions have been obtained [Bla13, BHH13, TPD16, GHJY14, HY16, GHK+17, CK17, BPSY16, GYBS17]. The best constructions for the case of $h = 2$ are from [BPSY16] and require a field of size $O(a \cdot n)$. For most other settings of parameters the best families of MR LRCs are from [GYBS17]. They require fields of size

$$O\left(r \cdot n^{(a+1)h-1}\right) \quad \text{and} \quad O\left(\max\left(n/r, r^{h+a}\right)^h\right). \tag{1}$$

The first bound is typically better when $r = \Omega(n)$. The second bound is better when $r \ll n$. Both bounds require q to grow rapidly with the codeword length. The results above exhibit code constructions but not any inherent limitations. In particular, up until our work it remained a possibility that codes over fields of size $O(n)$ could exist for all ranges of LRC parameters. We now summarize our results:

- We obtain the first lower bound on the field size of MR LRCs. In particular, we show that when a and h are constant and r may grow, for every maximally recoverable LRC, subject to $h \leqslant n/r$:

$$q \geqslant \Omega_{a,h}\left(n \cdot r^{\min\{a, h-2\}}\right). \tag{2}$$

- We obtain a family of MR $(n, r, h = 2, a, q)$-LRCs, where $q = O(n)$ for all settings of parameters. Prior to our work the best constructions [BPSY16] required q to be $O(a \cdot n)$ which in general may be up to quadratic in n. If we require that the field has characteristic two, we can get such codes with $q = n^{1+o(1)}$.
- We obtain a family of MR $(n, r, h = 3, a, q)$-LRCs, where $q = O(n^3)$ for all settings of parameters. Prior to our work the best constructions (1) required q to be up to $n^{\Theta(a)}$ for some regimes. If we require that the field has characteristic two, we can get such codes with $q = n^{3+o(1)}$.
- Our results in the first two bullets above suggest the setting of $r = 3, a = 1, h = 3$ as the first setting where existence of MR LRCs over fields of near linear size is an open question. We resolve this question in the positive by developing a new approach to LRC constructions based on elliptic curves and AP free sets.

1.2 Our techniques

Similarly to most earlier works in the area we represent LRC codes via their parity check matrices. Such matrices H have size $\left(a \cdot \frac{n}{r} + h\right) \times n$ and a simple block structure. Columns are partitioned into r-sized local groups. For each local group there is a corresponding collection of a rows that impose *MDS* constraints on coordinates in the group, and have no support outside the group. Remaining h rows of H correspond to heavy parity symbols and carry arbitrary values.

A LRC is MR if any subset of columns of H that can be obtained by selecting a columns from each local group and then h more has full rank. Suppose all h additional columns are selected from distinct local groups. In this case showing that some $ag + h$ columns are independent easily reduces to showing that a certain $(ah+h) \times (ah+h)$ determinant is non-zero. An important algebraic identity that underlies our first

three results reduces such determinants to much smaller $h \times h$ determinants of determinants in the entries of H. A special case of this identity when $h = 2$ and matrices are Vandermonde type appears in [BPSY16].

To establish our lower bound, we start with a parity check matrix of an arbitrary maximally recoverable local reconstruction code. We utilize the determinantal identity to obtain a family of large subsets $X_1, \ldots, X_{n/r}$ in the projective space $\mathbb{PF}_q^{h-1}$, such that any collection of vectors $x_1, \ldots, x_h$, where vectors $\{x_i\}_{i \in [h]}$ belong to different sets $\{X_j\}$ has full rank. We use vertex expansion properties of the hyperplane-point incidence graph of the projective space [Alo86] to bound the size of the families $\{X_j\}$ and translate this bound to the lower bound on the alphabet size.

Our code constructions both in the case of $h = 2$ and $h = 3$ also employ the determinantal identity. In addition to that we utilize various properties of finite fields such as the structure of multiplicative subgroups and field extensions. In the case of $h = 3$, we deviate from most existing constructions of MR LRCs in that we do not use linearized constraints (x, x^q, x^{q^2}) or Vandermonde constraints (x, x^2, x^3) and instead rely on Cauchy matrices [LN83] to specify heavy parities.

Our construction of MR $(n, r = 3, h = 3, a = 1, q)$-LRCs is technically disjoint from our other results. We observe that in this narrow case, MR LRCs are equivalent to subsets A of the projective plane $\mathbb{PF}_q^2$, where A is partitioned in triples $A = \sqcup_i \{a_i, b_i, c_i\}$ so that some three elements of A are collinear if and only if they constitute one of the triples $\{a_i, b_i, c_i\}$ in the partition. Moreover, minimizing the field size of maximally recoverable local reconstruction codes is in fact equivalent to maximizing the cardinality of such sets A. By considering all the $q + 1$ lines through an arbitrary point of A, it is easy to see that $|A| \leqslant q + 3$. We construct sets A with size $|A| \geqslant q^{1-o(1)}$. For our construction we start with an elliptic curve E over $\mathbb{F}_q$ such that the group of $\mathbb{F}_q$-rational points, $E(\mathbb{F}_q)$, is a cyclic group of size $\Omega(q)$. We observe that three points of $E(\mathbb{F}_q)$ are collinear if only and only if they sum to zero in the group. We then select a large AP-free set of points of $E(\mathbb{F}_q)$ using the classical construction of Behrend [Beh46] and complete these points to desired triples.

1.3 Related work

The first family of codes with locality for applications in storage comes from [HCL07, CHL07]. These papers also introduced the concept of maximal recoverability in a certain restricted setting. The work of [GHSY12] introduced a formal definition of local recovery and focused on codes that guarantee local recovery for a single failure. For this simple setting they were able to show that optimal codes must have a certain natural topology, e.g., codeword coordinates have to be arranged in groups where each group has a local parity. While [GHSY12] focused on systematic codes that provide local recovery for information symbols, [PD14] considered codes that provide locality for all symbols and defined local reconstruction codes. In parallel works maximally recoverable LRCs have been studied in [BHH13, Bla13]. Construction of local reconstruction codes with optimal distance over fields of linear size has been given in [TB14]. (Note that distance optimality is a much weaker property than maximal recoverability, e.g., when $a + h < r$ it only requires all patterns of size $a + h$ to be correctable, while MR property requires lots of very large patterns including some of size $(a + 1)h$ to be correctable.)

Maximal recoverability can be defined with respect to more general topologies then just local reconstruction codes [GHJY14]. The first lower bound for the field size of MR codes in any topology was recently given in [GHK+17]. This line of work was continued in [KLR17] where nearly matching upper and lower bounds were obtained. The topology considered in [GHK+17, KLR17] is a grid-like topology, where codewords form a codimension one subspace of tensor product codes, i.e., codewords are matrices, there is one heavy parity symbol, and each row / column constitutes a local group with one redundant symbol.

Finally, there are few other models of erasure correcting codes that provide efficient recovery in typical failure scenarios. These include regenerating codes [DGW+10, WTB17, YB17, GW16] that optimize band-

width consumed during repair rather than the number of coordinates (machines) accessed during repair; locally decodable codes [Yek12] that guarantee sub-linear time recovery of information coordinates even when a constant fraction of coordinates are erased; and SD codes [Bla13, BPSY16] that correct a certain subset of failure patterns correctable by MR LRCs.

1.4 Organization

In Section 2 we setup our notation, give formal definitions of local reconstruction codes and maximal recoverability, and establish some basic facts about MR LRCs. In Section 3 we present our determinantal identity and use it to obtain the lower bound on the alphabet size. In Section 4 we give a construction of MR LRCs with two heavy parity symbols over fields of nearly linear size. In Section 5 we generalize the determinantal identity and employ it together with some basic properties of finite fields to get explicit MR codes over fields of nearly cubic size. Finally, in Section 6 we focus on the narrow case of codes with three heavy parities, one parity per local group, and local groups of size three. We introduce the machinery of elliptic curves and AP free sets and employ it to obtain maximally recoverable codes over fields of nearly linear size. We conclude by listing some open problems in Section 7.

2 Preliminaries

We begin by summarizing few standard facts about erasure correcting codes [MS77].

- $[n, k, d]_q$ denotes a linear code (subspace) of dimension k, codeword length n, and Hamming distance d over a field $\mathbb{F}_q$. We often write $[n, k, d]$ or $[n, k]$ instead of $[n, k, d]_q$ when the left out parameters are not important.
- An $[n, k, d]$ code is called Maximum Distance Separable (MDS) if $d = n - k + 1$.
- A linear $[n, k, d]_q$ code C can be specified via its parity check matrix $H \in \mathbb{F}_q^{(n-k)\times n}$, where $C = \{x \in \mathbb{F}_q^n \mid H \cdot x = 0\}$. A code C is MDS iff every $(n-k) \times (n-k)$ minor of H is full rank.
- Let C be an $[n, k]$ code with a parity check matrix $H \in \mathbb{F}^{(n-k)\times n}$. Let E be a subset of the coordinates of C. If coordinates in E are erased; then they can be recovered (corrected) iff the matrix H restricted to coordinates in E has full rank.

We proceed to formally define local reconstruction codes.

Definition 2.1. *Let $r \mid n$, $a < r$, and h be integers and q be a prime power. Let $g = \frac{n}{r}$. Assume $h \leqslant n - ag$ and let $k = n - ga - h$. A linear $[n, k]$ code C over a field $\mathbb{F}_q$ is an (n, r, h, a, q)-LRC if for each $i \in [g]$, restricting C to coordinates in $\{r(i-1)+1, \ldots, ri\}$, yields a maximum distance separable code with parameters $[r, r-a, a+1]$.*

Let $[n] = \{1, \ldots, n\}$. In what follows we refer to subsets $\{r(i-1)+1, \ldots, ri\}$ of the set of code coordinates $[n]$ as local groups. There are g local groups and each such group has size r. It is immediate from the Definition 2.1 that every (n, r, h, a, q)-LRC admits a parity check matrix H of the following form

$$H = \left[\begin{array}{c|c|c|c} A_1 & 0 & \cdots & 0 \\ \hline 0 & A_2 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & A_g \\ \hline B_1 & B_2 & \cdots & B_g \end{array}\right]. \tag{3}$$

Here $A_1, A_2, \cdots, A_g$ are $a \times r$ matrices over $\mathbb{F}_q$, $B_1, B_2, \cdots, B_g$ are $h \times r$ matrices over $\mathbb{F}_q$. The rest of the matrix is filled with zeros. Every matrix $\{A_i\}_{i\in[g]}$ is a parity check matrix of an $[r, r-a, a+1]$ MDS code. The bottom h rows of H serve to increase the code co-dimension from ag to $ag+h$. Conversely, every matrix H as in (3), where $\mathrm{rank}(H) = ag + h$, and every $a \times a$ minor in each $\{A_i\}_{i\in[g]}$ has full rank, defines an (n, r, h, a, q)-LRC.

Definition 2.2. ‡ *Let C be an arbitrary (n, r, h, a, q)-local reconstruction code. We say that C is maximally recoverable if for any set $E \subseteq [n]$, $|E| = ga + h$, where E is obtained by selecting a coordinates from each of g local groups and then h more coordinates arbitrarily; E is correctable by the code C.*

The term maximally recoverable code is justified by the following observation (e.g., [GHJY14]): if an erasure pattern cannot be obtained via the process detailed in the Definition 2.2; then it cannot be corrected by any linear code whose parity check matrix has the shape (3). Thus MR codes provide the strongest possible reliability guarantees given the locality constraints defining the shape of the parity check matrix.

Existence of MR LRCs can be established non-explicitly [GHJY14] (i.e., by setting the non-zero entries in the matrix (3) at random in a large finite field and then analyzing the properties of the resulting code). There are also multiple explicit constructions available [CK17, GHJY14, GYBS17]. The key challenge in this line of work is to determine the minimal size of finite fields where such codes exist. In practice one is naturally mostly interested in fields of characteristic two.

Notation: We use $A \gtrsim B$ to denote $A = \Omega(B)$ and $A \lesssim B$ to denote $A = O(B)$. We use $A = O_\ell(B)$ and $A = \Omega_\ell(B)$ to denote that the hidden constants can depend on some parameter ℓ but independent of other parameters.

3 The lower bound

In this Section we establish our lower bound on the field size of maximally recoverable local reconstruction codes (Theorems 3.5 and 3.8). A code is MR if it corrects every erasure pattern that can be obtained by erasing a symbols per local group, and then h more. Note that if some local group carries at most a erasures; then it can be immediately corrected using only the properties of the local MDS code. Thus we never need to consider erasure patterns spread across more than h groups. Our lower bound does not use all the properties of MR LRCs, but only relies on code's ability to correct all patterns obtained by erasing $a + h$ elements in a single group as well as all patterns obtained by erasing exactly $a + 1$ coordinates in some h local groups.

The actual proof of the lower bound appears in Section 3.3, and in Sections 3.1 and 3.2 we prepare the necessary machinery. In particular, in Section 3.1 we setup notation, introduce some linear algebra tools, and establish an identity that allows us to reduce $(ah+h)\times(ah+h)$ determinants that arise when h distinct local groups each experience $a + 1$ erasures, to $h \times h$ determinants of determinants in the entries of the parity check matrix of the code. That identity is used in Section 3.3 to turn a parity check matrix of an MR LRC over a small field $\mathbb{F}_q$ into a family of large subsets $X_1, \ldots, X_g$ in the projective space $\mathbb{PF}_q^{h-1}$, where no hyperplane can simultaneously contain points from h different sets $\{X_j\}_{j\in[g]}$. In Section 3.2 we use vertex expansion properties of the hyperplane-point incidence graph of the projective space to bound the size of such families $\{X_j\}_{j\in[g]}$ and later translate this bound to the lower bound on the alphabet size.

‡Alternatively, one could define MR LRCs is as follows. Consider a matrix (3). Each way of fixing non-zero entries in (3) gives rise to (instantiates) a linear code. An instantiation is MR if it corrects all erasure patterns that are correctable for some other instantiation. It can be shown that under such definition and the minor technical assumption of $h \leqslant \frac{n}{r} \cdot (r-a) - \max\left\{\frac{n}{r}, r-a\right\}$ local codes have to be MDS [GHK+17, Proposition 4] as required in Definition 2.1.

3.1 The determinantal identity

Given a partitioned matrix of the form

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

where A is a square matrix of full rank, we can do column operations to remove U using linear combinations of columns of A as follows:

$$\left[\begin{array}{c|c} A & U \\ \hline B & V \end{array}\right] \cdot \left[\begin{array}{c|c} I & -A^{-1}U \\ \hline 0 & I \end{array}\right] = \left[\begin{array}{c|c} A & 0 \\ \hline B & V - BA^{-1}U \end{array}\right].$$

The matrix $V - BA^{-1}U$ is called the Schur complement of A and is denoted by M/A. Since column operations do not change the determinant, we have

$$\det(M) = \det(A)\det(M/A).$$

In what follows we sometimes abuse notation and employ the same variable to denote a number x, a 1×1 matrix containing x, or a single element set containing x. When A is an $a \times a$ matrix and M is $(a+1) \times (a+1)$ matrix, the Schur complement is

$$M/A = \frac{\det(M)}{\det(A)}.$$

For an $m \times n$ matrix A, we denote by $A^{(S)}(T)$, the submatrix of A formed by choosing the subset $S \subseteq [m]$ of its rows and the subset $T \subseteq [n]$ of its columns. When $S = [m]$ or $T = [n]$, we abbreviate $A^{(S)}(T)$ to respectively $A(T)$ and $A^{(S)}$.

The following lemma allows us to reduce $(ah + h) \times (ah + h)$ determinants that can be obtained by selecting $a + 1$ columns from some h different local groups in a matrix (3) to simpler determinants of size $h \times h$. A special case of this lemma when $h = 2$ and matrices are Vandermonde type appears in [BPSY16]. We also prove and use a more general form of this lemma in our MR code construction for $h = 3$ in Section 5, but we include the simpler form below for a self-contained treatment of our lower bound proof.

Lemma 3.1. *Let $C_1, \cdots, C_h$ be $a \times (a+1)$ dimensional matrices and $D_1, \cdots, D_h$ be $h \times (a+1)$ dimensional matrices over a field and let $D_i^{(j)}$ be the j^{th} row of D_i. Then,*

$$\det\left[\begin{array}{c|c|c|c} C_1 & 0 & \cdots & 0 \\ \hline 0 & C_2 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & C_h \\ \hline D_1 & D_2 & \cdots & D_h \end{array}\right] = (-1)^{\frac{ah(h-1)}{2}} \det\begin{bmatrix} \det\begin{pmatrix} C_1 \\ D_1^{(1)} \end{pmatrix} & \cdots & \det\begin{pmatrix} C_h \\ D_h^{(1)} \end{pmatrix} \\ \vdots & \ddots & \vdots \\ \det\begin{pmatrix} C_1 \\ D_1^{(h)} \end{pmatrix} & \cdots & \det\begin{pmatrix} C_h \\ D_h^{(h)} \end{pmatrix} \end{bmatrix}.$$

Proof. As we are showing an algebraic identity, without loss of generality, we can assume that the first a columns of C_i are independent for each $i \in [h]$. We can add a linear combination of the first a columns of C_i to the last column to make it zero and use the fact that the Schur complement of the $(a+1) \times (a+1)$ matrix $\begin{pmatrix} C_i \\ D_i^{(j)} \end{pmatrix}$ with respect to leading principal $a \times a$ minor $C_i([a])$ is $\det\begin{pmatrix} C_i \\ D_i^{(j)} \end{pmatrix} / \det(C_i([a]))$. The corresponding

column operation is shown below:

$$\begin{bmatrix} C_i \\ D_i \end{bmatrix} \to \begin{bmatrix} C_i([a]) & 0 \\ D_i^{(1)}([a]) & \det\begin{pmatrix} C_i \\ D_i^{(1)} \end{pmatrix} / \det(C_i([a])) \\ \vdots & \vdots \\ D_i^{(h)}([a]) & \det\begin{pmatrix} C_i \\ D_i^{(h)} \end{pmatrix} / \det(C_i([a])) \end{bmatrix}.$$

After such columns operations, we can permute the columns to get the following matrix:

$$\left[\begin{array}{c|c|c|c|c} C_1([a]) & 0 & \cdots & 0 & 0 \\ \hline 0 & C_2([a]) & \cdots & 0 & 0 \\ \hline \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline 0 & 0 & \cdots & C_h([a]) & \\ \hline D_1([a]) & D_2([a]) & \cdots & D_h([a]) & \begin{matrix} \det\begin{pmatrix} C_1 \\ D_1^{(1)} \end{pmatrix} / \det(C_1([a])) & \cdots & \det\begin{pmatrix} C_h \\ D_h^{(1)} \end{pmatrix} / \det(C_h([a])) \\ \vdots & \ddots & \vdots \\ \det\begin{pmatrix} C_1 \\ D_1^{(h)} \end{pmatrix} / \det(C_1([a])) & \cdots & \det\begin{pmatrix} C_h \\ D_h^{(h)} \end{pmatrix} / \det(C_h([a])) \end{matrix} \end{array}\right]$$

The above matrix is block triangular and so its determinant is the product of determinants of the diagonal blocks. Since column operations do not change the determinant and permuting the columns changes the sign of the determinant by $(-1)^{ah(h-1)/2}$, we get the desired result. □

Next we establish the following basic linear algebra fact, that will be used in the proof of our field size lower bound.

Lemma 3.2. *Let V be a $(d-1) \times d$ dimensional matrix over $\mathbb{F}$ which is full rank. Then the one dimensional subspace orthogonal to the row space of V is spanned by*

$$V^{\perp} = (-1)^{d+1}\left(\det(V([d] \setminus \{1\})), -\det(V([d] \setminus \{2\})), \cdots, (-1)^{d-1}\det(V([d] \setminus \{d\}))\right). \tag{4}$$

Moreover, for any vector $u \in \mathbb{F}^d$, we have:

$$\left\langle V^{\perp}, u \right\rangle = \det\begin{pmatrix} V \\ u \end{pmatrix}.$$

Proof. By Laplace expansion of the determinant, it is clear that

$$\left\langle V^{\perp}, u \right\rangle = \det\begin{pmatrix} V \\ u \end{pmatrix}$$

where $V^{\perp}$ is defined as above. Now $u \in \mathbb{F}^d$ is spanned by the rows of V iff the $\det\begin{pmatrix} V \\ u \end{pmatrix}$ vanishes, so $V^{\perp}$ is the vector orthogonal to the subspace spanned by the rows of V. □

3.2 Families of subsets of $\mathbb{PF}_q^d$ in general position

In this section we prepare the last ingredient (Lemma 3.4) of our field size lower bound. Our lemma follows from the following result regarding vertex expansion properties of the the point-hyperplane incidence graph in $\mathbb{PF}_q^d$.

Lemma 3.3. *(Theorem 2.3 in [Alo86]) Let $U = \mathbb{PF}_q^d$ be the set of points and $V = \mathbb{PF}_q^d$ be the set of hyperplanes in the d-dimensional projective space over $\mathbb{F}_q$. Let $G(U \cup V, E)$ be the incidence bipartite graph where a point $p \in U$ is connected to a hyperplane $H \in V$ iff p lies on H. Then for every $X \subseteq U$,*

$$|N(X)| \geqslant |V| - \frac{|V|^{1+\frac{1}{d}}}{|X|}$$

where $N(X)$ is the neighborhood of X in G.

We are now ready to prove the Lemma.

Lemma 3.4. *Let $X_1, \cdots, X_M \subseteq \mathbb{PF}_q^d$ be mutually disjoint subsets of size t. If*

$$\left\lfloor \frac{M}{d+1} \right\rfloor t > (q+1)(d+1) \tag{5}$$

then there exists a hyperplane $H \in \mathbb{PF}_q^d$ which contains points from $d+1$ distinct subsets among $X_1, \cdots, X_M$.

Proof. Let G be the point-hyperplane incidence graph in $\mathbb{PF}_q^d$ and for $X \subseteq \mathbb{PF}_q^d$, let $N(X)$ be the hyperplanes incident to at least one point in X. Let $n = |\mathbb{PF}_q^d| = (q^{d+1}-1)/(q-1)$ be the number of points in $\mathbb{PF}_q^d$. Set $m = \left\lfloor \frac{M}{d+1} \right\rfloor$ and for $i \in [d+1]$, define

$$S_i = X_{(i-1)m+1} \cup X_{(i-1)m+2} \cup \cdots \cup X_{im}.$$

By Lemma 3.3 and (5),

$$|\overline{N(S_i)}| \leqslant \frac{n^{1+1/d}}{mt} \leqslant \frac{n(q+1)}{mt} < \frac{n}{d+1}.$$

By union bound there exists a hyperplane H such that $H \in N(S_1) \cap N(S_2) \cap \cdots \cap N(S_{d+1})$ and therefore contains $d+1$ points from distinct subsets among $X_1, \cdots, X_{m(d+1)}$. □

3.3 Proof of the lower bound

We first establish the following theorem that relies on the condition $a + 2 \leqslant h$. Later in Theorem 3.8 we generalize our argument to take care of the case when $h < a + 2$.

Theorem 3.5. *When $a + 2 \leqslant h \leqslant n/r$, any maximally recoverable (n, r, h, a, q)-local reconstruction code must have*

$$q \geqslant \left\lfloor \frac{n}{rh^2} \right\rfloor \cdot \binom{r}{a+1} - 1. \tag{6}$$

Proof. Consider an arbitrary maximally recoverable (n, r, h, a, q)-LRC C with $g = \frac{n}{r}$ local groups. According to the discussion in Section 2 the code C admits a parity check matrix of the shape

$$\left[\begin{array}{c|c|c|c} A_1 & 0 & \cdots & 0 \\ \hline 0 & A_2 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & A_g \\ \hline B_1 & B_2 & \cdots & B_g \end{array}\right]. \tag{7}$$

Here $A_1, A_2, \cdots, A_g$ are $a \times r$ matrices over $\mathbb{F}_q$, $B_1, B_2, \cdots, B_g$ are $h \times r$ matrices over $\mathbb{F}_q$. The rest of the matrix is filled with zeros. Every $a \times a$ minor in each matrix $\{A_i\}_{i \in [g]}$ has full rank. For $i \in [g]$ and each subset $S \subseteq [r]$ of size $|S| = a + 1$, define $p_{i,S} \in \mathbb{F}_q^h$ as

$$p_{i,S} = \begin{bmatrix} \det \begin{pmatrix} A_i(S) \\ B_i^{(1)}(S) \end{pmatrix} \\ \det \begin{pmatrix} A_i(S) \\ B_i^{(2)}(S) \end{pmatrix} \\ \vdots \\ \det \begin{pmatrix} A_i(S) \\ B_i^{(h)}(S) \end{pmatrix} \end{bmatrix}.$$

The MR property implies that any subset of columns of the parity check matrix (7) which can be obtained by picking a columns in each local group and h arbitrary additional columns is full rank. We will use this property to make two claims about the vectors $\{p_{i,S}\}$.

Claim 3.6. *For every distinct $\ell_1, \cdots, \ell_h \in [g]$ and subsets $S_1, \cdots, S_h \subseteq [r]$ of size $a+1$ each, the $h \times h$ matrix $\left[p_{\ell_1,S_1}, \cdots, p_{\ell_h,S_h}\right]$ is full rank.*

Proof. Suppose we erase the coordinates corresponding to $S_1, \cdots, S_h$ in groups $\ell_1, \cdots, \ell_h$ respectively. MR property implies that the following matrix is full rank:

$$\left[\begin{array}{c|c|c|c} A_{\ell_1}(S_1) & 0 & \cdots & 0 \\ \hline 0 & A_{\ell_2}(S_2) & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & A_{\ell_h}(S_h) \\ \hline B_{\ell_1}(S_1) & B_{\ell_2}(S_2) & \cdots & B_{\ell_h}(S_h) \end{array}\right] \tag{8}$$

By Lemma 3.1, the above matrix is full rank iff the $h \times h$ matrix $\left[p_{\ell_1,S_1}, \cdots, p_{\ell_h,S_h}\right]$ is full rank. □

In particular the vectors $p_{i,S}$ are nonzero for every $i \in [g]$ and $S \in \binom{[r]}{a+1}$. We can also conclude that across different local groups, $p_{i,S}$ and $p_{j,T}$ are never multiples of each other when $i \neq j$. In fact, we will now show that even in the same local group, $p_{i,S}$ and $p_{i,T}$ are not multiples of each other unless $S = T$.

Claim 3.7. *For every $i \in [g]$, no two vectors in $\{p_{i,S} : S \subseteq \binom{[r]}{a+1}\}$ are multiples of each other.*

Proof. Suppose $p_{i,S} = \lambda \cdot p_{i,T}$ for some distinct sets $S, T \subset [r]$ of size $a + 1$ each and some nonzero $\lambda \in \mathbb{F}_q$.

Let $A_i(S)^\perp$ and $A_i(T)^\perp$ be column vectors defined as in (4). By Lemma 3.2,

$$\begin{bmatrix} A_i(S) \\ B_i(S) \end{bmatrix} \cdot A_i(S)^\perp - \lambda \cdot \begin{bmatrix} A_i(T) \\ B_i(T) \end{bmatrix} \cdot A_i(T)^\perp = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \det \begin{pmatrix} A_i(S) \\ B_i^{(1)}(S) \end{pmatrix} \\ \det \begin{pmatrix} A_i(S) \\ B_i^{(2)}(S) \end{pmatrix} \\ \vdots \\ \det \begin{pmatrix} A_i(S) \\ B_i^{(h)}(S) \end{pmatrix} \end{bmatrix} - \lambda \cdot \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \det \begin{pmatrix} A_i(T) \\ B_i^{(1)}(T) \end{pmatrix} \\ \det \begin{pmatrix} A_i(T) \\ B_i^{(2)}(T) \end{pmatrix} \\ \vdots \\ \det \begin{pmatrix} A_i(T) \\ B_i^{(h)}(T) \end{pmatrix} \end{bmatrix}$$

$$= \begin{pmatrix} 0 \\ p_{i,S} \end{pmatrix} - \lambda \cdot \begin{pmatrix} 0 \\ p_{i,T} \end{pmatrix} = 0$$

Note that every coordinate of $A_i(S)^\perp$ and $A_i(T)^\perp$ is nonzero as it carries the value of the determinant of an $a \times a$ minor of the matrix A_i. Thus we have a linear combination of the columns of $\begin{pmatrix} A_i(S \cup T) \\ B_i(S \cup T) \end{pmatrix}$ which is zero. Moreover the combination is non-trivial because there is some $j \in S \setminus T$ and the column $A_i(j)$ has a nonzero coefficient. However

$$|S \cup T| \leqslant 2a + 2 \leqslant a + h. \tag{9}$$

Observe however that any set of columns of the matrix $\begin{pmatrix} A_i \\ B_i \end{pmatrix}$ of cardinality at most $a+h$ has to be full rank by the MR property, as this set can be obtained by selecting (a subset of) a and then h more columns from the matrix (7). Thus we arrive at a contradiction that completes the proof of the claim. □

By Claim 3.7 and the discussion above the claim, we can think of $\left\{p_{i,S} : i \in [g], S \in \binom{[r]}{a+1}\right\}$ as distinct points in $\mathbb{PF}_q^{h-1}$. For brevity, from here on we assume that $p_{i,S}$ refers to the corresponding point in $\mathbb{PF}_q^{h-1}$. Define sets $X_1, \cdots, X_g \subseteq \mathbb{PF}_q^{h-1}$ as $X_i = \left\{p_{i,S} : S \in \binom{[r]}{a+1}\right\}$, we have $|X_1| = |X_2| = \cdots = |X_g| = \binom{r}{a+1}$ and they are mutually disjoint. Also $g \geqslant h$ by the hypothesis. By Claim 3.6, there is no hyperplane in $\mathbb{PF}_q^{h-1}$ which contains h points from distinct subsets of $X_1, \cdots, X_g$. So applying Lemma 3.4,

$$(q+1)h \geqslant \left\lfloor \frac{g}{h} \right\rfloor \cdot \binom{r}{a+1}.$$

Thus

$$q \geqslant \left\lfloor \frac{n}{rh^2} \right\rfloor \cdot \binom{r}{a+1} - 1,$$

which concludes the proof. □

In the argument above we used vectors $\{p_{i,S}\}$, where i varies across indices of g local groups and S varies across all $\binom{r}{a+1}$ subsets of $[r]$ of size $a+1$. In the proof we relied on the condition $a+2 \leqslant h$ to ensure that the union of any two such sets S has size at most $a+h$.

Parikshit Gopalan [Gop17] has recently observed (and kindly allowed us to include his observation here) that we can generalize Theorem 3.5 to the case when $h < a+2$. To do this, in cases when $h < a+2$ we only consider sets S that have size $a+1$ but are constrained to contain the set $\{1, 2, \ldots, a+2-h\}$, as this ensures that pairwise unions still have size at most $a+h$. Clearly, the total number of such sets is $\binom{r-a+h-2}{h-1}$. The rest of the proof remains the same and yields the following

Theorem 3.8. *Assume $h < a + 2$ and $h \leqslant n/r$; then any maximally recoverable (n, r, h, a, q)-local reconstruction code must have*

$$q \geqslant \left\lfloor \frac{n}{rh^2} \right\rfloor \cdot \binom{r - a + h - 2}{h - 1} - 1. \tag{10}$$

The following corollary follows immediately from Theorems 3.5 and 3.8 and presents the asymptotic form of our field size lower bound.

Corollary 3.9. *Suppose that a and h are arbitrary constants, but r may grow with n. Further suppose that $h \leqslant n/r$. In every maximally recoverable (n, r, h, a, q)-LRC, we have:*

$$q \geqslant \Omega_{a,h}\left(n \cdot r^{\min\{a, h-2\}}\right). \tag{11}$$

4 Maximally recoverable LRCs with $h = 2$

In this section we present our construction of maximally recoverable local reconstruction codes with two heavy parity symbols. Our construction relies on the determinantal identity (Lemma 3.1) and properties of finite fields. Let $\mathbb{F}_q^*$ denote the multiplicative group of the field $\mathbb{F}_q$.

Lemma 4.1. *Let $r \mid n$, $a < r$ be integers. Let $g = \frac{n}{r}$. Assume that $n - ga - 2$ is positive. Suppose q is a prime power such that there exists a subgroup of $\mathbb{F}_q^*$ of size at least r and with at least n/r cosets; then there exists an explicit maximally recoverable $(n, r, h = 2, a, q)$-local reconstruction code.*

Proof. Let $G \subset \mathbb{F}_q^*$ be the multiplicative subgroup from the statement of the theorem. Let $\alpha_1, \alpha_2, \cdots, \alpha_r \in G$ be distinct elements from G and let $\lambda_1, \lambda_2, \cdots, \lambda_g \in \mathbb{F}_q^*$ be elements from distinct cosets of G. We specify our code via a parity check matrix of the form (3). For $i \in [g]$, we choose matrices $\{A_i\}$ and $\{B_i\}$ as:

$$A_i = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_r \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_r^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^a & \alpha_2^a & \cdots & \alpha_r^a \end{bmatrix}; \; B_i = \begin{bmatrix} \lambda_i & \lambda_i & \cdots & \lambda_i \\ \alpha_1^{a+1} & \alpha_2^{a+1} & \cdots & \alpha_r^{a+1} \end{bmatrix}. \tag{12}$$

Suppose that we have a erasures per local group and two more. We can easily correct the coordinates corresponding to local groups which have at most a erasures in them. This is because every matrix A_i is a Vandermonde matrix and all its $a \times a$ minors are nonzero. Now we are left with two cases:

Case 1: Both the extra erasures occured in the same local group. Say, the i^{th} local group. In this case, we can correct the erased coordinates because any $(a + 2) \times (a + 2)$ minor of $\begin{bmatrix} A_i \\ B_i \end{bmatrix}$ (which is a Vandermonde matrix) is non degenerate.

Case 2: The two extra erasures occur in different groups say groups ℓ and ℓ', so we are left with two groups with $a + 1$ erasures in each. Let S be the columns erased in group ℓ and let S' be the columns erased in group ℓ'. We want to argue that the following $(2a + 2) \times (2a + 2)$ submatrix is full rank:

$$M = \left[\begin{array}{c|c} A_\ell(S) & 0 \\ \hline 0 & A_{\ell'}(S') \\ \hline B_\ell(S) & B_{\ell'}(S') \end{array}\right]. \tag{13}$$

Let $S = \{\gamma_1, \gamma_2, \cdots, \gamma_{a+1}\}$ and $S' = \{\gamma'_1, \gamma'_2, \cdots, \gamma'_{a+1}\}$, then by Lemma 3.1,

$$\det(M) = 0 \iff \det \begin{bmatrix} \det \begin{pmatrix} A_\ell(S) \\ B_\ell(S)^{(1)} \end{pmatrix} & \det \begin{pmatrix} A_{\ell'}(S') \\ B_{\ell'}(S')^{(1)} \end{pmatrix} \\ \det \begin{pmatrix} A_\ell(S) \\ B_\ell(S)^{(2)} \end{pmatrix} & \det \begin{pmatrix} A_{\ell'}(S') \\ B_{\ell'}(S')^{(2)} \end{pmatrix} \end{bmatrix} = 0 \tag{14}$$

$$\iff \det \begin{bmatrix} \det \begin{pmatrix} \gamma_1 & \cdots & \gamma_{a+1} \\ \gamma_1^2 & \cdots & \gamma_{a+1}^2 \\ \vdots & \ddots & \vdots \\ \gamma_1^a & \cdots & \gamma_{a+1}^a \\ \lambda_\ell & \cdots & \lambda_\ell \end{pmatrix} & \det \begin{pmatrix} \gamma'_1 & \cdots & \gamma'_{a+1} \\ (\gamma'_1)^2 & \cdots & (\gamma'_{a+1})^2 \\ \vdots & \ddots & \vdots \\ (\gamma'_1)^a & \cdots & (\gamma'_{a+1})^a \\ \lambda_{\ell'} & \cdots & \lambda_{\ell'} \end{pmatrix} \\ \det \begin{pmatrix} \gamma_1 & \cdots & \gamma_{a+1} \\ \gamma_1^2 & \cdots & \gamma_{a+1}^2 \\ \vdots & \ddots & \vdots \\ \gamma_1^a & \cdots & \gamma_{a+1}^a \\ \gamma_1^{a+1} & \cdots & \gamma_{a+1}^{a+1} \end{pmatrix} & \det \begin{pmatrix} \gamma'_1 & \cdots & \gamma'_{a+1} \\ \gamma_1'^2 & \cdots & (\gamma'_{a+1})^2 \\ \vdots & \ddots & \vdots \\ \gamma_1'^a & \cdots & (\gamma'_{a+1})^a \\ \gamma_1'^{a+1} & \cdots & (\gamma'_{a+1})^{a+1} \end{pmatrix} \end{bmatrix} = 0 \tag{15}$$

$$\iff \det \begin{bmatrix} \lambda_\ell & \lambda_{\ell'} \\ \prod_{i\in[a+1]} \gamma_i & \prod_{i\in[a+1]} \gamma'_i \end{bmatrix} = 0 \tag{16}$$

where we factored out the (nonzero) Vandermonde determinant from each column. Since $\gamma_i, \gamma'_i \in G$ and $\lambda_\ell, \lambda_{\ell'}$ are in different cosets of G, the last determinant is not zero. □

In Lemma 4.1, given n and r such that $r \mid n$, we want to find a small field $\mathbb{F}_q$ such that $\mathbb{F}_q^*$ contains a subgroup of size at least r and with at least n/r cosets. For example, if $n+1$ is a prime power, then we can take $q = n+1$. The following lemma shows that one can always find such a field of size $q = O(n)$. We defer the proof to the Appendix.

Lemma 4.2. *Let r, n be some positive integers with $r \leqslant n$. Then there exists a finite field $\mathbb{F}_q$ with $q = O(n)$ such that the multiplicative group $\mathbb{F}_q^*$ contains a subgroup of size at least r and with at least n/r cosets. If additionally we require that the field has characteristic two, then such a field exists with $q = n \cdot \exp(O(\sqrt{\log n}))$.*

Combining Lemma 4.2 with Lemma 4.1 gives the following theorem.

Theorem 4.3. *Let $r \mid n$, $a < r$ be integers. Let $g = \frac{n}{r}$. Assume that $n - ga - 2$ is positive. Then there exists an explicit maximally recoverable $(n, r, h = 2, a, q)$-local reconstruction code with $q = O(n)$. If we require the field to be of characteristic 2, such a code exists with $q \leqslant n \cdot \exp(O(\sqrt{\log n}))$.*

5 Maximally recoverable LRCs with $h = 3$

In this section, we present our construction of maximally recoverable local reconstruction codes with three heavy parity symbols. Our construction extends the ideas in the construction of Section 4 using field extensions.

5.1 A determinantal identity

For our construction and analysis, we will need a generalization of the determinantal identity in Lemma 3.1. To prove the generalization, we need the following expansion of determinant of a column partitioned matrix.

Lemma 5.1. *For $i \in [\ell]$, let F_i be an $h \times t_i$ matrix with $\sum_{i=1}^{\ell} t_i = h$. Then,*

$$\det[F_1|F_2|\cdots|F_\ell] = \sum_{S_1 \sqcup \cdots \sqcup S_\ell = [h],\, |S_i| = t_i} \operatorname{sgn}(S_1, \cdots, S_\ell) \prod_{i \in [\ell]} \det F_i^{(S_i)}$$

where $S_1 \sqcup \cdots \sqcup S_\ell$ ranges over partitions of $[h]$ such that $|S_i| = t_i$. Here $\operatorname{sgn}(S_1, \cdots, S_\ell)$ is the sign of the permutation taking $(1, 2, \cdots, h)$ to $(\tilde{S}_1, \tilde{S}_2, \cdots, \tilde{S}_\ell)$ where $\tilde{S}_i$ is the tuple formed by ordering the elements of S_i in increasing order.

Proof. Given distinct integers $a_1, \cdots, a_n$, define $\operatorname{sgn}(a_1, a_2, \cdots, a_n) := (-1)^t$ where t is number of transpositions needed to sort the elements $a_1, a_2, \cdots, a_n$ in increasing order. Thus for a permutation $\pi \in S_h$, $\operatorname{sgn}(\pi) = \operatorname{sgn}(\pi(1), \pi(2), \cdots, \pi(h))$. Let $F = [F_1|F_2|\cdots|F_\ell]$ and for $i \in [\ell]$, let $T_i = \{t_{i-1} + 1, \cdots, t_i\}$ where $t_0 = 0$. We can expand $\det(F)$ as:

$$\begin{aligned}
\det(F) &= \sum_{\pi \in S_h} \operatorname{sgn}(\pi) \prod_{i=1}^{h} F_{\pi(i)i} \\
&= \sum_{S_1 \sqcup \cdots \sqcup S_\ell = [h],\, |S_i| = t_i} \;\; \sum_{\pi:\ \pi(T_i) = S_i} \operatorname{sgn}(\pi) \prod_{i=1}^{h} F_{\pi(i)i}
\end{aligned}$$

Note that if $\pi(T_i) = S_i$, then for $i \in [\ell]$,

$$\operatorname{sgn}(\pi) = \operatorname{sgn}(\tilde{S}_1, \cdots, \tilde{S}_\ell) \prod_{i=1}^{\ell} \operatorname{sgn}(\pi(t_{i-1} + 1), \cdots, \pi(t_i))$$

because we can sort $(\pi(1), \cdots, \pi(h))$ first within each group to get $(\tilde{S}_1, \cdots, \tilde{S}_\ell)$ and then sort it to get $(1, 2, \cdots, h)$. Therefore,

$$\begin{aligned}
&\sum_{\pi:\ \pi(T_i) = S_i} \operatorname{sgn}(\pi) \prod_{i=1}^{h} F_{\pi(i)i} \\
&= \sum_{\sigma_1 : T_1 \to S_1, \ldots,\ \sigma_\ell : T_\ell \to S_\ell} \operatorname{sgn}(\tilde{S}_1, \cdots, \tilde{S}_\ell) \prod_{i=1}^{\ell} \left(\operatorname{sgn}(\sigma_i(t_{i-1} + 1), \cdots, \sigma_i(t_i)) \prod_{j=t_{i-1}+1}^{t_i} F_{\sigma_i(j)j} \right) \\
&\qquad\qquad \text{(where the summation is over all bijections } \sigma_i : T_i \to S_i) \\
&= \operatorname{sgn}(\tilde{S}_1, \cdots, \tilde{S}_\ell) \prod_{i=1}^{\ell} \left(\sum_{\sigma_i : T_i \to S_i} \operatorname{sgn}(\sigma_i(t_{i-1} + 1), \cdots, \sigma_i(t_i)) \prod_{j=t_{i-1}+1}^{t_i} F_{\sigma_i(j)j} \right) \\
&= \operatorname{sgn}(\tilde{S}_1, \cdots, \tilde{S}_\ell) \prod_{i=1}^{\ell} \det F_i^{(S_i)}.
\end{aligned}$$

□

The following is a generalization of Lemma 3.1.

Lemma 5.2. *For $i \in [\ell]$, let C_i be an $a \times (a + t_i)$ matrix and D_i be an $h \times (a + t_i)$ matrix for some $t_1 + t_2 + \cdots + t_\ell = h$ where $t_i \geqslant 1$. Then,*

$$\det \left[\begin{array}{c|c|c|c} C_1 & 0 & \cdots & 0 \\ \hline 0 & C_2 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & C_\ell \\ \hline D_1 & D_2 & \cdots & D_\ell \end{array}\right] = (-1)^{a\left(\sum_{i=1}^{\ell} t_i (\ell - i)\right)} \sum_{S_1 \sqcup \cdots \sqcup S_\ell = [h],\, |S_i| = t_i} \operatorname{sgn}(S_1, \cdots, S_\ell) \prod_{i \in [\ell]} \det \begin{pmatrix} C_i \\ D_i^{(S_i)} \end{pmatrix}$$

where $S_1 \sqcup \cdots \sqcup S_\ell$ ranges over partitions of $[h]$ such that $|S_i| = t_i$ and $\operatorname{sgn}(S_1, \cdots, S_\ell)$ is defined as in Lemma 5.1.

Proof. Let

$$F = [F_1|F_2|\cdots|F_\ell] = \left[\begin{array}{c|c|c|c} C_1 & 0 & \cdots & 0 \\ \hline 0 & C_2 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & C_\ell \\ \hline D_1 & D_2 & \cdots & D_\ell \end{array}\right].$$

Let $[p, q]$ be the integers between p and q, i.e., $[p, q] = \{i : p \leqslant i \leqslant q\}$. By Lemma 5.1,

$$\det F = \det[F_1|F_2|\cdots|F_\ell] = \sum_{T_1 \sqcup \cdots \sqcup T_\ell = [a\ell+h],\, |T_i| = a+t_i} \operatorname{sgn}(T_1, \cdots, T_\ell) \prod_{i\in[\ell]} \det F_i^{(T_i)}$$

Note that the only terms which survive correspond to partitions $T_1 \sqcup T_2 \sqcup \cdots \sqcup T_\ell$ of rows of F such that for every $i \in [\ell]$, T_i contains the rows of C_i (i.e. $[(i-1)a+1, ia]$). In the other terms, there exists some $i \in [\ell]$ such that $F_i^{(T_i)}$ contains a zero row and thus $\det F_i^{(T_i)} = 0$. Such partitions are given by $T_i = [(i-1)a+1, ia] \cup S_i$ where $S_1 \sqcup S_2 \cdots \sqcup S_\ell$ is some partition of rows of $[D_1|D_2|\cdots|D_\ell]$ such that $|S_i| = t_i$. So the expansion for $\det F$ can be written as:

$$\begin{aligned}
\det F &= \sum_{S_1 \sqcup \cdots \sqcup S_\ell = [a\ell+1,\, a\ell+h],\, |S_i| = t_i} \operatorname{sgn}([1,a] \cup S_1, \cdots, [(\ell-1)a+1, \ell a] \cup S_\ell) \prod_{i\in[\ell]} \det F_i^{([(i-1)a, ia] \cup S_i)} \\
&= (-1)^{a(\sum_{i=1}^{\ell} t_i(\ell-i))} \sum_{S_1 \sqcup \cdots \sqcup S_\ell = [a\ell+1,\, a\ell+h],\, |S_i| = t_i} \operatorname{sgn}([1, \ell a], S_1, S_2, \cdots S_\ell) \prod_{i\in[\ell]} \det F_i^{([(i-1)a, ia] \cup S_i)} \\
&= (-1)^{a(\sum_{i=1}^{\ell} t_i(\ell-i))} \sum_{S_1 \sqcup \cdots \sqcup S_\ell = [h],\, |S_i| = t_i} \operatorname{sgn}(S_1, S_2, \cdots S_\ell) \prod_{i\in[\ell]} \det \begin{pmatrix} C_i \\ D_i^{(S_i)} \end{pmatrix}.
\end{aligned}$$

□

Applying the above lemma to a special case, we get the following corollary:

Corollary 5.3. *Let C_1 be an $a \times (a+1)$ matrix, C_2 be an $a \times (a+2)$ matrix, D_1 be a $3 \times (a+1)$ matrix and D_2 be a $3 \times (a+2)$ matrix and let $D_i^{(j)}$ be the j^{th} row of D_i. Then,*

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

5.2 Code construction and analysis

Our construction is based on Cauchy matrices, and we will also need the the following lemma about the determinants of such matrices.

Lemma 5.4. *([LN83]) Let $\alpha_1, \cdots, \alpha_m, \beta_1, \cdots, \beta_m \in \mathbb{F}_q$ be all distinct; then*

$$\det \begin{bmatrix} \frac{1}{\alpha_1-\beta_1} & \frac{1}{\alpha_2-\beta_1} & \cdots & \frac{1}{\alpha_m-\beta_1} \\ \frac{1}{\alpha_1-\beta_2} & \frac{1}{\alpha_2-\beta_2} & \cdots & \frac{1}{\alpha_m-\beta_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\alpha_1-\beta_m} & \frac{1}{\alpha_2-\beta_m} & \cdots & \frac{1}{\alpha_m-\beta_m} \end{bmatrix} = \frac{\prod_{i>j}(\alpha_i-\alpha_j)(\beta_j-\beta_i)}{\prod_{i,j}(\alpha_i-\beta_j)}$$

Matrices of the above form are called Cauchy matrices. Every minor of a Cauchy matrix is nonzero because the minors themselves look like a Cauchy matrix. We are now ready to present the construction for three global parities.

Lemma 5.5. *Let $r \mid n$, $a < r$ be integers. Let $g = \frac{n}{r}$. Assume that $n - ga - 3$ is positive. Suppose $q_0 \geqslant 2r+3$ is a prime power such that there exists a subgroup of $\mathbb{F}_{q_0}^*$ of size at least $r+2$ and with at least n/r cosets. Then there exists an explicit maximally recoverable $(n, r, h = 3, a, q = q_0^3)$-local reconstruction code.*

Proof. Let $G \subset \mathbb{F}_{q_0}^*$ be the multiplicative subgroup from the statement of the theorem. Choose distinct $\beta_{a+1}, \beta_{a+2}, \beta_{a+3} \in \mathbb{F}_{q_0}$ and let

$$\Omega = \left\{ \alpha \in \mathbb{F}_{q_0} : \frac{\alpha - \beta_{a+2}}{\alpha - \beta_{a+3}} \in G \right\}.$$

Clearly $|\Omega| = |G| - 1 \geqslant r+1$, so we can choose distinct $\alpha_1, \cdots, \alpha_r \in \Omega \setminus \{\beta_{a+1}\}$. Finally, since $q_0 \geqslant 2r+3 \geqslant r+a+3$, we can choose distinct $\beta_1, \cdots, \beta_a \in \mathbb{F}_{q_0} \setminus \{\alpha_1, \cdots, \alpha_r, \beta_{a+1}, \beta_{a+2}, \beta_{a+3}\}$. Let $\mu_1, \cdots, \mu_g \in \mathbb{F}_{q_0}$ be elements from distinct cosets of G.

Now let $\mathbb{F}_q$ be a degree 3 extension of $\mathbb{F}_{q_0}$, so we have $q = q_0^3$. As $\mathbb{F}_q$ is a 3-dimensional vector space over $\mathbb{F}_{q_0}$, choose a basis $v_0, v_1, v_2 \in \mathbb{F}_q$ for this space and choose distinct $\gamma_1, \cdots, \gamma_g \in \mathbb{F}_{q_0}$. Define $\lambda_i = v_0 + \gamma_i v_1 + \gamma_i^2 v_2$. Then any three of the elements $\lambda_1, \cdots, \lambda_g \in \mathbb{F}_q$ are linearly independent over $\mathbb{F}_{q_0}$; we call this property 3-wise independence over $\mathbb{F}_{q_0}$. Define the matrices A_i and B_i as follows:

$$A_i = \begin{bmatrix} \frac{1}{\alpha_1 - \beta_1} & \cdots & \frac{1}{\alpha_r - \beta_1} \\ \vdots & \ddots & \vdots \\ \frac{1}{\alpha_1 - \beta_a} & \cdots & \frac{1}{\alpha_r - \beta_a} \end{bmatrix}; \; B_i = \begin{bmatrix} \frac{\lambda_i}{\alpha_1 - \beta_{a+1}} & \cdots & \frac{\lambda_i}{\alpha_r - \beta_{a+1}} \\ \frac{\mu_i}{\alpha_1 - \beta_{a+2}} & \cdots & \frac{\mu_i}{\alpha_r - \beta_{a+2}} \\ \frac{1}{\alpha_1 - \beta_{a+3}} & \cdots & \frac{1}{\alpha_r - \beta_{a+3}} \end{bmatrix} \tag{17}$$

Now we will show that the above construction satisfies the MR property. We have a erasures per local group and 3 more. We can easily correct groups with only a erasures because A_i are Cauchy matrices where every $a \times a$ minor is non-degenerate. So we only need to worry about local groups with more than a erasures. There are three cases.

Case 1: All three extra erasures in the same group.

Say we have $a+3$ erasures in local group i, then we can correct these errors because the matrix $\begin{pmatrix} A_i \\ B_i \end{pmatrix}$ is a Cauchy matrix (except for some scaling factors in the rows), and therefore each of its $(a+3) \times (a+3)$ minors is nonzero by Lemma 5.4.

Case 2: The three extra erasures are distributed across two groups.
Suppose the extra erasures occur in groups ℓ, ℓ' with $(a+1)$ erasures in group ℓ corresponding to a subset $S \subseteq [r]$ of its columns and $(a+2)$ erasures in group ℓ' corresponding to a subset $S' \subseteq [r]$ of its columns. To correct these erasures we need to show the following matrix is full rank:

$$\left[\begin{array}{c|c} A_\ell(S) & 0 \\ \hline 0 & A_{\ell'}(S') \\ \hline B_\ell(S) & B_{\ell'}(S') \end{array}\right]. \tag{18}$$

By Corollary 5.3, the above matrix fails to be full rank iff

$$\det\begin{pmatrix} A_\ell(S) \\ B_\ell(S)^{(1)} \end{pmatrix} \cdot \det\begin{pmatrix} A_{\ell'}(S') \\ B_{\ell'}(S')^{(2)} \\ B_{\ell'}(S')^{(3)} \end{pmatrix} - \det\begin{pmatrix} A_\ell(S) \\ B_\ell(S)^{(2)} \end{pmatrix} \cdot \det\begin{pmatrix} A_{\ell'}(S') \\ B_{\ell'}(S')^{(1)} \\ B_{\ell'}(S')^{(3)} \end{pmatrix}$$
$$+ \det\begin{pmatrix} A_\ell(S) \\ B_\ell(S)^{(3)} \end{pmatrix} \cdot \det\begin{pmatrix} A_{\ell'}(S') \\ B_{\ell'}(S')^{(1)} \\ B_{\ell'}(S')^{(2)} \end{pmatrix} = 0$$

The above determinant is a $\mathbb{F}_q$-linear combination of λ_ℓ and $\lambda_{\ell'}$ and the coefficient of λ_ℓ, which arises from the first term, is nonzero because $\begin{pmatrix} A_\ell \\ B_\ell \end{pmatrix}$ and $\begin{pmatrix} A_{\ell'} \\ B_{\ell'} \end{pmatrix}$ are Cauchy matrices. By 3-wise independence of λ's, this linear combination cannot be zero, and therefore the matrix (18) has full rank.

Case 3: The three extra erasures occur in distinct groups.
Suppose the three extra erasures occur in groups $\ell_1, \ell_2, \ell_3 \in [g]$ and let $S_1, S_2, S_3 \subseteq [r]$ be sets of size $a+1$ corresponding to the erasures in the groups ℓ_1, ℓ_2, ℓ_3 respectively. To correct these erasures we need to show the following matrix is full rank:

$$\left[\begin{array}{c|c|c} A_{\ell_1}(S_1) & 0 & 0 \\ \hline 0 & A_{\ell_2}(S_2) & 0 \\ \hline 0 & 0 & A_{\ell_3}(S_3) \\ \hline B_{\ell_1}(S_1) & B_{\ell_2}(S_2) & B_{\ell_3}(S_3) \end{array}\right]$$

By Lemma 3.1, if the above matrix is not full rank then

$$\det\begin{bmatrix} \det\begin{pmatrix} A_{\ell_1}(S_1) \\ B^{(1)}_{\ell_1}(S_1) \end{pmatrix} & \det\begin{pmatrix} A_{\ell_2}(S_2) \\ B^{(1)}_{\ell_2}(S_2) \end{pmatrix} & \det\begin{pmatrix} A_{\ell_3}(S_3) \\ B^{(1)}_{\ell_3}(S_3) \end{pmatrix} \\ \det\begin{pmatrix} A_{\ell_1}(S_1) \\ B^{(2)}_{\ell_1}(S_1) \end{pmatrix} & \det\begin{pmatrix} A_{\ell_2}(S_2) \\ B^{(2)}_{\ell_2}(S_2) \end{pmatrix} & \det\begin{pmatrix} A_{\ell_3}(S_3) \\ B^{(2)}_{\ell_3}(S_3) \end{pmatrix} \\ \det\begin{pmatrix} A_{\ell_1}(S_1) \\ B^{(3)}_{\ell_1}(S_1) \end{pmatrix} & \det\begin{pmatrix} A_{\ell_2}(S_2) \\ B^{(3)}_{\ell_2}(S_2) \end{pmatrix} & \det\begin{pmatrix} A_{\ell_3}(S_3) \\ B^{(3)}_{\ell_3}(S_3) \end{pmatrix} \end{bmatrix} = 0.$$

For $k \in \{1,2,3\}$, let $c_k = \prod_{i>j, i,j\in S_k}(\alpha_i - \alpha_j), d = \prod_{i>j,i,j\in[a]}(\beta_j - \beta_i), e_k = \prod_{i\in S_k, j\in[a]}(\alpha_i - \beta_j)$. By Lemma 5.4, we can write down explicit expressions for the entries in the above determinant to get:

$$\det\begin{bmatrix} \lambda_{\ell_1}\frac{c_1 d\prod_{i\in[a]}(\beta_i-\beta_{a+1})}{e_1\prod_{i\in S_1}(\alpha_i-\beta_{a+1})} & \lambda_{\ell_2}\frac{c_2 d\prod_{i\in[a]}(\beta_i-\beta_{a+1})}{e_2\prod_{i\in S_2}(\alpha_i-\beta_{a+1})} & \lambda_{\ell_3}\frac{c_3 d\prod_{i\in[a]}(\beta_i-\beta_{a+1})}{e_3\prod_{i\in S_3}(\alpha_i-\beta_{a+1})} \\ \mu_{\ell_1}\frac{c_1 d\prod_{i\in[a]}(\beta_i-\beta_{a+2})}{e_1\prod_{i\in S_1}(\alpha_i-\beta_{a+2})} & \mu_{\ell_2}\frac{c_2 d\prod_{i\in[a]}(\beta_i-\beta_{a+2})}{e_2\prod_{i\in S_2}(\alpha_i-\beta_{a+2})} & \mu_{\ell_3}\frac{c_3 d\prod_{i\in[a]}(\beta_i-\beta_{a+2})}{e_3\prod_{i\in S_3}(\alpha_i-\beta_{a+2})} \\ \frac{c_1 d\prod_{i\in[a]}(\beta_i-\beta_{a+3})}{e_1\prod_{i\in S_1}(\alpha_i-\beta_{a+3})} & \frac{c_2 d\prod_{i\in[a]}(\beta_i-\beta_{a+3})}{e_2\prod_{i\in S_2}(\alpha_i-\beta_{a+3})} & \frac{c_3 d\prod_{i\in[a]}(\beta_i-\beta_{a+3})}{e_3\prod_{i\in S_3}(\alpha_i-\beta_{a+3})} \end{bmatrix} = 0.$$

We can scale rows and columns to conclude that

$$\det\begin{bmatrix} \lambda_{\ell_1}\prod_{i\in S_1}\left(\frac{\alpha_i-\beta_{a+3}}{\alpha_i-\beta_{a+1}}\right) & \lambda_{\ell_2}\prod_{i\in S_2}\left(\frac{\alpha_i-\beta_{a+3}}{\alpha_i-\beta_{a+1}}\right) & \lambda_{\ell_3}\prod_{i\in S_3}\left(\frac{\alpha_i-\beta_{a+3}}{\alpha_i-\beta_{a+1}}\right) \\ \mu_{\ell_1}\prod_{i\in S_1}\left(\frac{\alpha_i-\beta_{a+3}}{\alpha_i-\beta_{a+2}}\right) & \mu_{\ell_2}\prod_{i\in S_2}\left(\frac{\alpha_i-\beta_{a+3}}{\alpha_i-\beta_{a+2}}\right) & \mu_{\ell_3}\prod_{i\in S_3}\left(\frac{\alpha_i-\beta_{a+3}}{\alpha_i-\beta_{a+2}}\right) \\ 1 & 1 & 1 \end{bmatrix} = 0.$$

By the choice of α's, $\prod_{i\in S_j}\left(\frac{\alpha_i-\beta_{a+3}}{\alpha_i-\beta_{a+2}}\right) \in G$ for $j = 1, 2, 3$. By writing the Laplace expansion of the determinant over the first row, the above determinant is a linear combination in $\lambda_{\ell_1}, \lambda_{\ell_2}, \lambda_{\ell_3}$ with coefficients from $\mathbb{F}_{q_0}$. The coefficients of λ's in this linear combination are nonzero because $\mu_{\ell_1}, \mu_{\ell_2}, \mu_{\ell_3}$ belong to distinct cosets of G in $\mathbb{F}_{q_0}^*$. Because λ's are 3-wise independent over $\mathbb{F}_{q_0}$, we get a contradiction. □

Combining Lemma 5.5 with Lemma 4.2 gives the following theorem.

Theorem 5.6. *Let $r \mid n, a < r$ be integers. Let $g = \frac{n}{r} \geqslant 2$. Assume that $n - ga - 3$ is positive. Then there exists an explicit maximally recoverable $(n, r, h = 3, a, q)$-local reconstruction code with $q = O(n^3)$. If we require the field to be of characteristic 2, such a code exists with $q = n^3 \cdot \exp(O(\sqrt{\log n}))$.*

6 Maximally recoverable LRCs from elliptic curves

Our construction of MR $(n, r = 3, h = 3, a = 1, q)$-LRCs is technically disjoint from our results in the previous sections. We observe that in this narrow case, maximally recoverable LRCs are equivalent to families of *matching collinear triples* in the projective plane $\mathbb{PF}_q^2$, i.e., sets of points partitioned into collinear triples, where no three points other than those forming a triple are collinear. In Section 6.1 we state the quantitative parameters of such a family A that we can obtain and translate those to parameters of an MR LRC. The goal of Section 6.2 is to construct the family A using elliptic curves and 3-AP free sets. In Section 6.2.1 we develop the necessary machinery of elliptic curves, and in Section 6.2.2 we carry out the construction.

6.1 LRCs from matching collinear triples

We will reduce the problem of constructing maximally recoverable codes for $h = 3, r = 3, a = 1$ to the problem of constructing matching collinear triples in $\mathbb{PF}_q^2$ which we define below.

Definition 6.1. *We say that $A \subset \mathbb{PF}_q^2$ has matching collinear triples if A can be partitioned into triples, $A = \sqcup_{i=1}^m \{a_i, b_i, c_i\}$, such that the only collinear triples in A are $\{a_i, b_i, c_i\}$ for $i \in [m]$.*

What is the largest subset $A \subset \mathbb{PF}_q^2$ with matching collinear triples? If we consider all the $q + 1$ lines through some fixed point of A, at most one line can contain two other points of A. All other lines can contain at most one other point of A. So $|A| \leqslant q + 3$. The following lemma shows that we can construct a set A with size $|A| \geqslant q^{1-o(1)}$. It is an interesting open question if we can get $|A| \geqslant \Omega(q)$.

Lemma 6.2. *For any prime power q, there is an explicit set $A \subset \mathbb{PF}_q^2$ with matching collinear triples of size $|A| \geqslant q \cdot \exp(-C\sqrt{\log q})$ where $C > 0$ is some absolute constant.*

We will prove Lemma 6.2 in Section 6.2.2.

Lemma 6.3. *Assume $g \geqslant 2$. There exists a subset $S \subset \mathbb{PF}_q^2$ that has g matching collinear triples* if and only if *there exists a maximally recoverable $(3g, r = 3, h = 3, a = 1, q)$-local reconstruction code.*

Proof. We first show how to obtain codes from families of collinear triples. Let $S = \cup_{i=1}^g \{a_i, b_i, c_i\}$ be such that the only collinear triples in S are $\{a_i, b_i, c_i\}$ for $i \in [g]$. From now, we will think of elements of S as vectors in $\mathbb{F}_q^3$ such that every triple of points except for the triples $\{a_i, b_i, c_i\}$ are linearly independent. We can scale each vector with nonzero elements in $\mathbb{F}_q$ such that $a_i + b_i + c_i = 0$ in $\mathbb{F}_q^3$ for every $i \in [g]$. For $i \in [g]$, define blocks A_i and B_i of the parity check matrix (3) as:

$$A_i = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}; \; B_i = \begin{bmatrix} 0 & -b_i & c_i \end{bmatrix}.$$

We need to correct 1 erasure per group and any 3 extra erasures. We can correct groups with a single erasure because A_i is a simple parity check constraint on all the coordinates of the group. We now have to correct groups with more than one erasure, there are two cases:

Case 1: The three extra erasures are in two groups.
Suppose the two groups are i, j and in group i all the coordinates are erased and in group j the second and third coordinates are erased (the other two cases are similar). To correct these erasures, we have to argue that the following matrix is full rank:

$$\left[\begin{array}{ccc|cc} 1 & 1 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 1 \\ \hline 0 & -b_i & c_i & -b_j & c_j \end{array}\right]$$

Subtract the first column in each group from the rest, it is equivalent to the following matrix being full rank:

$$\left[\begin{array}{ccc|cc} 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ \hline 0 & -b_i & c_i & -b_j & c_j + b_j \end{array}\right] = \left[\begin{array}{ccc|cc} 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ \hline 0 & -b_i & c_i & -b_j & a_j \end{array}\right]$$

which is true because b_i, c_i, a_j are linearly independent.

Case 2: The three extra erasures are in distinct groups.
Suppose the three groups are i, j, k and in each group the second and third columns are erased (the other cases are similar). To correct these erasures, we have to argue that the following matrix is full rank:

$$\left[\begin{array}{cc|cc|cc} 1 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 1 \\ \hline -b_i & c_i & -b_j & c_j & -b_k & c_k \end{array}\right]$$

Subtract the first column in each group from the rest, it is equivalent to the following matrix being full rank:

$$\left[\begin{array}{cc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ \hline -b_i & c_i + b_i & -b_j & c_j + b_j & -b_k & c_k + b_k \end{array}\right] = \left[\begin{array}{cc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ \hline -b_i & -a_i & -b_j & -a_j & -b_k & -a_k \end{array}\right]$$

which is true because a_i, a_j, a_k are linearly independent.

Reverse connection. We now proceed to show how to obtain a set with matching collinear triples from codes. Given a maximally recoverable $(3g, r = 3, h = 3, a = 1, q)$-local reconstruction code with a parity check matrix (3), without loss of generality assume that for all $i \in [g]$,

$$A_i = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}; \; B_i = \begin{bmatrix} v_i^1 & v_i^2 & v_i^3 \end{bmatrix},$$

where $\{v_i^s\}_{s\in[3], i\in[g]} \subseteq \mathbb{F}_q^3$. For each $i \in [g]$, define

$$a_i = v_i^2 - v_i^1 \qquad b_i = v_i^3 - v_i^2 \qquad c_i = v_i^1 - v_i^3.$$

Clearly, for all $i \in [g]$, $a_i + b_i + c_i = 0$. Consider $\{a_i, b_i, c_i\}_{i\in[g]}$ as elements of $\mathbb{PF}_q^2$ and define our family to be $S = \cup_{i=1}^{g} \{a_i, b_i, c_i\}$. It remains to show that all triples of elements of S other than $\{a_i, b_i, c_i\}$ are non-collinear. When all three elements $v_i^\alpha - v_i^\beta, v_j^\gamma - v_j^\delta, v_k^\varepsilon - v_k^\zeta$ belong to different groups this follows from

the fact that, as implied by the MR property, the matrix

$$\left[\begin{array}{cc|cc|cc} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ \hline v_i^\beta & v_i^\alpha & v_j^\delta & v_j^\gamma & v_k^\zeta & v_k^\varepsilon \end{array}\right] = \left[\begin{array}{cc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline v_i^\beta & v_i^\alpha - v_i^\beta & v_j^\delta & v_j^\gamma - v_j^\delta & v_k^\zeta & v_k^\varepsilon - v_k^\zeta \end{array}\right]$$

is full rank. When triples come from two groups, (say, $v_i^\beta - v_i^\alpha, v_i^\gamma - v_i^\alpha, v_j^\delta - v_j^\varepsilon$) this again follows from the MR property, as the matrix

$$\left[\begin{array}{ccc|cc} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ \hline v_i^\alpha & v_i^\beta & v_i^\gamma & v_j^\varepsilon & v_j^\delta \end{array}\right] = \left[\begin{array}{ccc|cc} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \hline v_i^\alpha & v_i^\beta - v_i^\alpha & v_i^\gamma - v_i^\alpha & v_j^\varepsilon & v_j^\delta - v_j^\varepsilon \end{array}\right]$$

is also full rank. □

Combining Lemma 6.2 and Lemma 6.3 along with the fact that all the constructions are explicit gives the following theorem.

Theorem 6.4. *For any $n > 3$ which is a multiple of 3 and for any finite field $\mathbb{F}_q$, there exists an explicit maximally recoverable $(n, r = 3, h = 3, a = 1, q)$-local reconstruction code provided that $q \geqslant \Omega\left(n \cdot \exp\left(C\sqrt{\log n}\right)\right)$ where $C > 0$ is some absolute constant.*

6.2 Matching Collinear Triples from AP free sets

In this section, we will prove Lemma 6.2 by constructing a large $A \subset \mathbb{PF}_q^2$ with matching collinear triples. The main idea is to reduce the problem to constructing a large subset $A \subset \mathbb{Z}/N\mathbb{Z}$ with *matching tri-sums* where $N = \Omega(q)$. A subset $A \subset \mathbb{Z}/N\mathbb{Z}$ has matching tri-sums if A can partitioned into disjoint triples, $A = \sqcup_i \{a_i, b_i, c_i\}$ such that the only 3 element subsets of A which sum to zero are the triples $\{a_i, b_i, c_i\}$ in the partition. Such sets can be constructed from subsets of $[N]$ without any non-trivial arithmetic progressions. The best known construction of a subset of $[N]$ with no non-trivial three term arithmetic progressions is due to Behrend [Beh46] which was slightly improved in [Elk11]. An explicit construction with similar bounds as [Beh46] was given in [Mos53].

Theorem 6.5 ([Beh46, Mos53, Elk11])**.** *For some absolute constant $C > 0$, there exists an explicit $A \subset \{1, 2, \cdots, N\}$ with $|A| \geqslant N \cdot \exp(-C\sqrt{\log N})$ which doesn't contain any 3 term arithmetic progressions i.e. there doesn't exist distinct $x, y, z \in A$ such that $x + z = 2y$.*

It is also known that any set $A \subset \{1, 2, \cdots, N\}$ with no non-trivial 3 term arithmetic progressions should have size $|A| \lesssim \frac{(\log\log N)^4}{\log N} \cdot N$ [Blo16].

The reduction from matching collinear triples in $\mathbb{F}_q^2$ to subsets of $\mathbb{Z}/N\mathbb{Z}$ with matching tri-sums is simple when q is a prime. In this case we can set $N = q$. Three points $(x_1, y_1), (x_2, y_2), (x_3, y_3) \in \mathbb{F}_q^2$ on the cubic curve $Y = X^3$ are collinear iff $x_1 + x_2 + x_3 = 0$. So we can get a large subset of $\mathbb{PF}_q^2$ with matching collinear triples, from a large subset of $\mathbb{F}_q \cong \mathbb{Z}/q\mathbb{Z}$ with matching tri-sums. And from Theorem 6.5, we can get such a set of size $\geqslant q \cdot \exp(-O(\sqrt{\log q}))$.

When q is not prime, the additive group of $\mathbb{F}_q$ is not cyclic anymore and subsets of $\mathbb{F}_q$ with matching tri-sums are much smaller. For example, if $\mathbb{F}_q$ has characteristic 2, which is the main setting of interest for us, the size of the largest subset of $\mathbb{F}_q$ with matching tri-sums is $\leqslant q^c$ for some absolute constant $c < 1$ [Kle16]. We will use some results on elliptic curves which are a special kind of cubic curves to make the reduction work over any field.

6.2.1 Elliptic curves

We will give a quick introduction to elliptic curves, please refer to [Sil09, MBG+13] for proofs and formal definitions. Let $\mathbb{K}$ be a finite field and $\overline{\mathbb{K}}$ be its algebraic closure. A *Weierstrass equation* defined over $\mathbb{K}$ is homogeneous cubic equation in three variables of the following form:

$$F(X, Y, Z) = Y^2Z + a_1XYZ + a_3YZ^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3 = 0$$

where $a_1, a_2, \cdots, a_6 \in \mathbb{K}$. A point $p \in \mathbb{P}\overline{\mathbb{K}}^2$ is called a *singular point* if

$$\frac{\partial F}{\partial X}(p) = \frac{\partial F}{\partial Y}(p) = \frac{\partial F}{\partial Z}(p) = 0.$$

If there are no such points, we call the equation *non-singular*, else we call the equation *singular*. Since the equation is cubic, it can have at most one singular point. There is an explicit polynomial function Δ in variables $a_1, a_2, \cdots, a_6$ and coefficients in $\mathbb{K}$ called the *discriminant*, such that the Weierstrass equation is singular iff $\Delta(a_1, \cdots, a_6) = 0$ (see Section III.1 in [Sil09] for the explicit polynomial). A singular Weierstrass equation[§] E with singularity at $(X, Y, Z) = (0, 0, 1)$ can be written as:

$$E: \ Y^2Z + a_1XYZ - a_3X^2Z = X^3.$$

We associate with E the set of all points in $\mathbb{P}\overline{\mathbb{K}}^2$ which satisfy the equation E. There is exactly one point in E with Z-coordinate equal to 0, namely $(0 : 1 : 0)$, we call this special point *the point at infinity* and denote it by O. The set of non-singular $\mathbb{K}$-rational points of E, denoted by $E_{ns}(\mathbb{K})$ is defined as follows:

$$E_{ns}(\mathbb{K}) = \{(x : y : 1) | F(x, y, 1) = 0, \ x, y \in \mathbb{K}, (x, y) \neq (0, 0)\} \cup \{O\}.$$

$E_{ns}(\mathbb{K})$ is an Abelian group under a certain addition operation '+', with the point at infinity O as the group identity. Under this operation, three points $a, b, c \in E_{ns}(\mathbb{K})$ satisfy $a + b + c = O$ iff a, b, c are collinear in $\mathbb{P}\mathbb{K}^2$. The following theorem shows that $E_{ns}(\mathbb{K})$ is isomorphic to $\mathbb{K}^*$ when E is of a special form.

Theorem 6.6 (Theorem 8.1 in [MBG+13])**.** *Let $E : (Y - \alpha X)(Y - \beta X)Z = X^3$ be a singular Weierstrass equation with $\alpha, \beta \in \mathbb{K}$ and $\alpha \neq \beta$. Then the map $\phi : E_{ns}(\mathbb{K}) \to \mathbb{K}^*$ defined as:*

$$\phi : O \mapsto 1 \qquad \phi : (x, y, 1) \mapsto \frac{y - \beta x}{y - \alpha x}$$

is a group isomorphism.

Since $\mathbb{K}^*$ is a cyclic group for any finite field $\mathbb{K}$, $E_{ns}(\mathbb{K})$ is isomorphic to $\mathbb{Z}/N\mathbb{Z}$ for $N = |\mathbb{K}| - 1$ when E is a singular Weierstrass equation as in Theorem 6.6.

6.2.2 Proof of Lemma 6.2

Proof. Let E be a singular Weierstrass equation[¶] defined over $\mathbb{F}_q$ as in Theorem 6.6. By Theorem 6.6, $E_{ns}(\mathbb{F}_q) \cong \mathbb{Z}/N\mathbb{Z}$ where $N = q - 1$. Recall that $a, b, c \in E_{ns}(\mathbb{F}_q)$ satisfy $a + b + c = 0$ in the group iff they are collinear.

[§]Usually elliptic curves are defined as curves given by non-singular Weierstrass equations. But for our purpose, it is easier to work with singular Weierstrass equations.

[¶]It is not essential to work with singular Weierstrass equations. The proof also works with non-singular elliptic curves as long the group of $\mathbb{K}$-rational points is cyclic or has a large cyclic subgroup.

Let $B \subset \{1, 2, \cdots, N/20\}$ be an explicit subset of size $|B| \gtrsim N \cdot \exp(-C\sqrt{\log N})$ with no 3-term arithmetic progressions, as guaranteed by Theorem 6.5. Now define subsets $A_1, A_2, A_3 \subset \mathbb{Z}/N\mathbb{Z}$ as

$$A_1 = \{x : x \in B\}, A_2 = \left\{\left\lfloor \frac{N}{3} \right\rfloor + x : x \in B\right\}, A_3 = \left\{N - \left\lfloor \frac{N}{3} \right\rfloor - 2x : x \in B\right\}.$$

Clearly, A_1, A_2, A_3 are disjoint. Finally we define $\tilde{A} = A_1 \cup A_2 \cup A_3$. Now we claim that the only triples from $\tilde{A}$ which sum to zero in $\mathbb{Z}/N\mathbb{Z}$ are $\{x, \lfloor N/3 \rfloor + x, N - \lfloor N/3 \rfloor - 2x\}$ for $x \in B$ and these triples form a partition of $\tilde{A}$.

It is not hard to see that if three distinct elements $a, b, c \in \tilde{A}$ satisfy $a + b + c = 0$, then a, b, c should come from 3 different sets A_1, A_2, A_3. So after reordering, we can assume

$$a = x, b = \lfloor N/3 \rfloor + y, c = N - \lfloor N/3 \rfloor - 2z$$

for some $x, y, z \in B$. Thus $a + b + c = 0$ implies that $x + y = 2z$, which implies that $x = y = z$ since B is free from 3 arithmetic progressions.

Finally let $A \subset \mathbb{PF}_q^2$ be the set of points in $E_{ns}(\mathbb{F}_q)$ which map to the set $\tilde{A} \subset \mathbb{Z}/N\mathbb{Z}$ under the isomorphism $E_{ns}(\mathbb{F}_q) \cong \mathbb{Z}/N\mathbb{Z}$. Now it is easy to see that A has matching collinear triples and we have $|A| \gtrsim q \cdot \exp(-C\sqrt{\log q})$. □

7 Open problems

In this work we made progress towards quantifying the minimal size of finite fields required for existence of maximally recoverable local reconstruction codes and obtained both lower and upper bounds. There is a wide array of questions that remain open. Here we highlight some of them:

- Our lower bound (2) implies that even in the regime of constant a and h, when $h \geqslant 3$ and r grows with n there exist no MR codes over fields of size $O(n)$. It would be of great interest to understand if such codes always exist when all parameters a, h, and r are held constant and only n grows.
- Our Lemma 6.3 provides an equivalence between the parameters of families of matching collinear triples in the projective plane and maximally recoverable local reconstruction codes with $r = 3, h = 3$, and $a = 1$. We hope that this reduction will be useful to obtain an $\omega(n)$ lower bound for the alphabet size of MR $(n, r = 3, h = 3, a = 1, q)$-LRCs, or lead to a construction over fields of linear size. It is also very interesting to see if techniques similar to those in Section 6.2 can be used to get codes over fields of nearly linear size when $r > 3$ or $a > 1$ or $h > 3$.
- It is interesting to understand whether various technical conditions that appear in our theorems are in fact necessary. For instance, can one relax the condition $h \leqslant n/r$ in our main lower bound (Theorems 3.5 and 3.8)? Also, in the case of fields of characteristic two, can one reduce the field sizes in Theorems 4.3 and 5.6 to $O(n)$ and $O(n^3)$ to match the case of prime fields?
- Finally, it is interesting to see if our lower bounds (Theorems 3.5 and 3.8) can be generalized to the setting of non-linear codes. Basic results about LRCs such as distance vs. redundancy tradeoff [GHSY12] have been generalized to non-linear setting in [SAP+13, FY14].

Acknowledgements

We would like to thank Parikshit Gopalan for allowing us to include his Theorem 3.8 in this paper. We would also like to thank Cheng Huang for asking the question that led us to start this project, and Ilya

Shkredov for helpful discussions about this work. And finally we would like to thank Suryateja Gavva for telling us about Theorem A.2.

References

[Alo86] Noga Alon. Eigenvalues, geometric expanders, sorting in rounds, and ramsey theory. *Combinatorica*, 6(3):207–219, 1986. 3, 8

[Bal12] Simeon Ball. On sets of vectors of a finite vector space in which every subset of basis size is a basis. *Journal of European Mathematical Society*, 14:733–748, 2012. 2

[Beh46] Felix A Behrend. On sets of integers which contain no three terms in arithmetical progression. *Proceedings of the National Academy of Sciences*, 32(12):331–332, 1946. 3, 19

[BFI86] Enrico Bombieri, John B Friedlander, and Henryk Iwaniec. Primes in arithmetic progressions to large moduli. *Acta Mathematica*, 156(1):203–251, 1986. 25

[BHH13] 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(7):4510–4519, 2013. 1, 2, 3

[Bla13] Mario Blaum. Construction of PMDS and SD codes extending RAID 5. Arxiv 1305.0032, 2013. 1, 2, 3, 4

[Blo16] Thomas F Bloom. A quantitative improvement for Roth's theorem on arithmetic progressions. *Journal of the London Mathematical Society*, page jdw010, 2016. 19

[BPSY16] Mario Blaum, James 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. 2, 3, 4, 6

[CHL07] Minghua Chen, Cheng Huang, and Jin Li. On maximally recoverable property for multi-protection group codes. In *IEEE International Symposium on Information Theory (ISIT)*, pages 486–490, 2007. 1, 3

[CK17] Gokhan Calis and Ozan Koyluoglu. A general construction fo PMDS codes. *IEEE Communications Letters*, 21(3):452–455, 2017. 1, 2, 5

[DGW+10] Alexandros G. Dimakis, Brighten Godfrey, Yunnan Wu, Martin J. Wainwright, and Kannan Ramchandran. Network coding for distributed storage systems. *IEEE Transactions on Information Theory*, 56(9):4539–4551, 2010. 3

[Elk11] Michael Elkin. An improved construction of progression-free sets. *Israel journal of mathematics*, 184(1):93–128, 2011. 19

[FY14] Michael Forbes and Sergey Yekhanin. On the locality of codeword symbols in non-linear codes. *Discrete mathematics*, 324:78–84, 2014. 21

[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. 2, 3, 5

[GHK+17] Parikshit Gopalan, Guangda Hu, Swastik Kopparty, Shubhangi Saraf, Carol Wang, and Sergey Yekhanin. Maximally recoverable codes for grid-like topologies. In *28th Annual Symposium on Discrete Algorithms (SODA)*, pages 2092–2108, 2017. 2, 3, 5

[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. 3, 21

[Gop17] Parikshit Gopalan. Personal communication, 2017. 10

[GW16] Venkatesan Guruswami and Mary Wootters. Repairing Reed-Solomon codes. In *48th ACM Symposium on Theory of Computing (STOC)*, pages 216–226, 2016. 3

[GYBS17] Ryan Gabrys, Eitan Yaakobi, Mario Blaum, and Paul Siegel. Construction of partial MDS codes over small finite fields. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 1–5, 2017. 2, 5

[HCL07] Cheng Huang, Minghua Chen, and Jin Li. Pyramid codes: flexible schemes to trade space for access efficiency in reliable data storage systems. In *6th IEEE International Symposium on Network Computing and Applications (NCA 2007)*, pages 79–86, 2007. 1, 3

[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 *USENIX Annual Technical Conference (ATC)*, pages 15–26, 2012. 1

[HY16] Guangda Hu and Sergey Yekhanin. New constructions of SD and MR codes over small finite fields. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 1591–1595, 2016. 2

[Kle16] Robert Kleinberg. A nearly tight upper bound on tri-colored sum-free sets in characteristic 2. *arXiv preprint arXiv:1605.08416*, 2016. 19

[KLR17] Daniel Kane, Shachar Lovett, and Sankeerth Rao. Labeling the complete bipartite graph with no zero cycles. In *58th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2017. 3

[LN83] Rudolf Lidl and Harald Niederreiter. *Finite Fields*. Cambridge University Press, Cambridge, 1983. 3, 14

[MBG+13] A.J. Menezes, I.F. Blake, X.H. Gao, R.C. Mullin, S.A. Vanstone, and T. Yaghoobian. *Applications of Finite Fields*. The Springer International Series in Engineering and Computer Science. Springer US, 2013. 20

[Mos53] Leo Moser. *On non-averaging sets of integers*. Canadian Mathematical Society, 1953. 19

[MS77] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error Correcting Codes*. North Holland, Amsterdam, New York, 1977. 1, 4

[PD14] Dimitris Papailiopoulos and Alexandros Dimakis. Locally repairable codes. *IEEE Transactions on Information Theory*, 60(10):5843–5855, 2014. 1, 3

[PGM13] J. S. Plank, K. M. Greenan, and E. L. Miller. Screaming fast Galois field arithmetic using Intel SIMD instructions. In *11th Usenix Conference on File and Storage Technologies (FAST)*, pages 299–306, San Jose, February 2013. 1

[SAP+13] Maheswaran Sathiamoorthy, Megasthenis Asteris, Dimitris S. Papailiopoulos, Alexandros G. Dimakis, Ramkumar Vadali, Scott Chen, and Dhruba Borthakur. XORing elephants: novel erasure codes for big data. In *Proceedings of VLDB Endowment (PVLDB)*, pages 325–336, 2013. 1, 21

[Sil09] J.H. Silverman. *The Arithmetic of Elliptic Curves*. Graduate Texts in Mathematics. Springer New York, 2009. 20

[TB14] Itzhak Tamo and Alexander Barg. A family of optimal locally recoverable codes. *IEEE Transactions on Information Theory*, 60:4661–4676, 2014. 1, 3

[TPD16] Itzhak Tamo, Dimitris Papailiopoulos, and Alexandros G. Dimakis. Optimal locally repairable codes and connections to matroid theory. *IEEE Transactions on Information Theory*, 62:6661–6671, 2016. 2

[WTB17] Zhiying Wang, Itzhak Tamo, and Jehoshua Bruck. Optimal rebuilding of multiple erasures in MDS codes. *IEEE Transactions on Information Theory*, 63:1084–1101, 2017. 3

[YB17] Min Ye and Alexander Barg. Explicit constructions of high-rate MDS array codes with optimal repair bandwidth. *IEEE Transactions on Information Theory*, 63:2001–2014, 2017. 3

[Yek12] Sergey Yekhanin. Locally decodable codes. *Foundations and trends in theoretical computer science*, 6(3):139–255, 2012. 4

A Proof of Lemma 4.2

The goal of the section is to prove Lemma 4.2 which is restated here for convenience.

Lemma A.1 (Restatement of Lemma 4.2)**.** *Let r, n be some positive integers with $r \leqslant n$. Then there exists a finite field $\mathbb{F}_q$ with $q = O(n)$ such that the multiplicative group $\mathbb{F}_q^*$ contains a subgroup of size at least r and with at least n/r cosets. If additionally we require that the field has characteristic two, then such a field exists with $q = n \cdot \exp(O(\sqrt{\log n}))$.*

We will need some estimates from analytic number theory, we will setup some notation first.

$\pi(x; m, a)$: number of primes $p \leqslant x$ such that $p \equiv a \mod m$

$\pi(x, y; m, a) = \pi(y; m, a) - \pi(x; m, a)$

$\mathrm{Li}(x) = \int_2^x \frac{1}{\ln t} dt$

(m, a) : greatest common divisor of m and a

$\phi(m)$: number of positive integers $a \leqslant m$ such that $(a, m) = 1$ (Euler's totient function)

By the prime number theorem, the number of primes $\leqslant x$ is approximately $\mathrm{Li}(x) = \Theta(x/\log x)$. So if the primes are equidistributed among different congruence classes of m with no obvious divisors (i.e. $a \mod m$ where $(a, m) = 1$), then we expect to see approximately $\mathrm{Li}(x)/\phi(m)$ primes in each such congruence class. The following theorem gives an upper bound on the error term in this approximation averaged over $m < \sqrt{x}(\log x)^A$.

Theorem A.2 (Theorem from [BFI86] (Page 250)). *Let $a \neq 0, A \geqslant 0$ be some fixed constants and $x \geqslant 3$. We then have*

$$\sum_{(m,a)=1;\ m<\sqrt{x}(\log x)^A} \left|\pi(x;m,a) - \frac{\mathrm{Li}(x)}{\phi(m)}\right| \lesssim_{a,A} x\frac{(\log\log x)^B}{(\log x)^3}$$

where B is an absolute constant.

Applying the above theorem with $a = 1, A = 0$ for x and $2x$, and using triangle inequality, we get the following corollary.

Corollary A.3. *For x large enough,*

$$\sum_{m<\sqrt{x}} \left|\pi(x,2x;m,1) - \frac{(\mathrm{Li}(2x) - \mathrm{Li}(x))}{\phi(m)}\right| \lesssim x\frac{(\log\log x)^B}{(\log x)^3}$$

where B is an absolute constant.

Lemma A.4. *Let $a \leqslant b$ be some positive integers. Then there exists $A \geqslant a$,$B \geqslant b$ such that $AB + 1$ is a prime and $AB = O(ab)$.*

Proof. If there exists some A such that $a \leqslant A \leqslant 2a$ and there is a prime p between $4ab + 1$ and $8ab$ which is congruent to $1 \mod A$, then we can take $B = (p-1)/A \geqslant b$. Suppose this is not true, we will arrive at a contradiction. For every $a \leqslant m \leqslant 2a$, we have $\pi(4ab, 8ab; m, 1) = 0$. Applying corollary A.3 with $x = 4ab$, we get

$$\begin{aligned}
ab\frac{(\log\log ab)^B}{(\log ab)^3} &\gtrsim \sum_{m<2\sqrt{ab}} \left|\pi(4ab,8ab;m,1) - \frac{(\mathrm{Li}(8ab) - \mathrm{Li}(4ab))}{\phi(m)}\right| \\
&\geqslant \sum_{a\leqslant m<2a} \left|\pi(4ab,8ab;m,1) - \frac{(\mathrm{Li}(8ab) - \mathrm{Li}(4ab))}{\phi(m)}\right| \\
&= \sum_{a\leqslant m<2a} \frac{(\mathrm{Li}(8ab) - \mathrm{Li}(4ab))}{\phi(m)} \\
&\geqslant a\frac{\mathrm{Li}(8ab) - \mathrm{Li}(4ab)}{2a} \gtrsim \frac{ab}{\log(ab)}
\end{aligned}$$

which is a contradiction when ab is large enough. □

In practice, it is desirable to work with fields of characteristic two, the following lemma gives us such fields.

Lemma A.5. *Let a, b be some positive integers and let $n = ab$. Then there exists $A \geqslant a$, $B \geqslant b$ such that $q = AB + 1$ is a power of two and $q = n \cdot \exp(O\sqrt{\log n})$.*

Proof. Let m be a positive integer to be chosen later. Let ℓ be an integer such that

$$2^{\ell(2^m-1)} \geqslant Cn + 1 > 2^{(\ell-1)(2^m-1)}$$

where $C \geqslant 1$ is some sufficiently large constant to be chosen later and let $x = 2^\ell, q = x^{2^m}$. We will now show that for any $a \leqslant n$, we can factor $q - 1$ as $A \cdot B$ where $A \geqslant a$ and $B \geqslant n/a = b$. We can factor $q - 1 = x^{2^m} - 1$ as:

$$x^{2^m} - 1 = (x-1)\prod_{i\in[m]}(1 + x^{2^{i-1}}).$$

We will rearrange these factors to get the desired factorization of $q-1$. Let $0 \leqslant \alpha \leqslant 2^m - 1$ be such that $x^{\alpha-1} < a \leqslant x^{\alpha}$. Expand α into its binary expansion as $\alpha = \sum_{i \in S} 2^i$ where $S \subset \{0, 1, \cdots, m-1\}$. Define $A = \prod_{i \in S}(1 + x^{2^i})$ and define $B = (x^{2^m} - 1)/A$. Clearly $A \geqslant x^{\alpha} \geqslant a$. We can lower bound B as follows:

$$\begin{aligned} B &= \frac{(x^{2^m}-1)}{\prod_{i \in S}(1+x^{2^i})} = \prod_{i \in S}(1+x^{-2^i})^{-1} \cdot \frac{(x^{2^m}-1)}{\prod_{i \in S} x^{2^i}} \\ &\geqslant \exp(-\sum_{j \geqslant 0} x^{-2^j}) \frac{(x^{2^m}-1)}{x^{\alpha}} \geqslant \exp(-\sum_{j \geqslant 0} 2^{-2^j}) \frac{(x^{2^m}-1)}{xa} \\ &\geqslant \exp(-\sum_{j \geqslant 0} 2^{-2^j}) \frac{(x^{2^m-1}-1)}{a} \geqslant \exp(-\sum_{j \geqslant 0} 2^{-2^j}) \frac{Cn}{a} \geqslant \frac{n}{a} \end{aligned}$$

when $C = \exp(\sum_{j \geqslant 0} 2^{-2^j})$. Now we need to bound $q = x^{2^m}$ as a function of n.

$$\begin{aligned} q = 2^{\ell 2^m} &= 2^{(\ell-1)(2^m-1)} \cdot 2^{\ell} \cdot 2^{2^m-1} \\ &\leqslant (Cn+1) \cdot 2^{\ell} \cdot 2^{2^m-1} \\ &\lesssim n^{1+1/(2^m-1)} \cdot 2^{2^m-1} \\ &\lesssim n \exp(O(\sqrt{\log n})) \end{aligned}$$

if we choose m such that $(2^m - 1) = \Theta(\sqrt{\log n})$.

□

We are now ready to prove Lemma 4.2.

Proof of Lemma 4.2. By Lemma A.4, there exists $A \geqslant r$ and $B \geqslant n/r$ such that $q = AB + 1$ is prime and $q = O(n)$. Since $\mathbb{F}_q^*$ is a cyclic group of size $q-1$ and A divides $q-1$, there exists a subgroup of $\mathbb{F}_q^*$ of size $A \geqslant r$ with $B \geqslant n/r$ cosets. To get a finite field of characteristic two, we use Lemma A.5 instead. □

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il