Improved Maximally Recoverable LRCs Using Skew Polynomials

Sivakanth Gopi and Venkatesan Guruswami, *Fellow, IEEE*

***Abstract*— An (n, r, h, a, q)-Local Reconstruction Code (LRC) is a linear code over $\mathbb{F}_q$ of length n, whose codeword symbols are partitioned into n/r local groups each of size r. Each local group satisfies 'a' local parity checks to recover from 'a' erasures in that local group and there are further h global parity checks to provide fault tolerance from more global erasure patterns. Such an LRC is Maximally Recoverable (MR), if it offers the best blend of locality and global erasure resilience—namely it can correct all erasure patterns whose recovery is information-theoretically feasible given the locality structure (these are precisely patterns with up to 'a' erasures in each local group and an additional h erasures anywhere in the codeword). Random constructions can easily show the existence of MR LRCs over very large fields, but a major algebraic challenge is to construct MR LRCs, or even show their existence, over smaller fields, as well as understand inherent lower bounds on their field size. We give an explicit construction of (n, r, h, a, q)-MR LRCs with field size q bounded by $(O(\max\{r, n/r\}))^{\min\{h, r-a\}}$. This significantly improves upon known constructions in many practically relevant parameter ranges. Moreover, it matches the lower bound from Gopi *et al.* (2020) in an interesting range of parameters where $r = \Theta(\sqrt{n})$, $r - a = \Theta(\sqrt{n})$ and h is a fixed constant with $h \leqslant a + 2$, achieving the optimal field size of $\Theta_h(n^{h/2})$. Our construction is based on the theory of skew polynomials. We believe skew polynomials should have further applications in coding and complexity theory; as a small illustration we show how to capture algebraic results underlying list decoding folded Reed-Solomon and multiplicity codes in a unified way within this theory.**

***Index Terms*— Erasure coding, distributed storage, local reconstruction codes, maximally recoverable codes, skew polynomials.**

I. Introduction

WE PRESENT an approach to construct Maximally Recoverable Local Reconstruction Codes (MR LRCs) based on the theory of skew polynomials. Our construction matches or improves the field size of MR LRCs for most parameter regimes. We now describe the motivation of MR LRCs in the context of coding for distributed storage, and then formally define them and describe our results.

Manuscript received 24 December 2021; accepted 3 May 2022. Date of publication 20 May 2022; date of current version 21 October 2022. This work was supported in part by NSF Grant CCF-1563742, Grant CCF-1814603, and Grant CCF-2210823; and in part by a Simons Investigator Award. *(Corresponding author: Sivakanth Gopi.)*

Sivakanth Gopi is with Microsoft Research, Redmond, WA 98052 USA (e-mail: sigopi@microsoft.com).

Venkatesan Guruswami was with the Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213 USA. He is now with the Departments of Electrical Engineering and Computer Science (EECS) and Mathematics, University of California at Berkeley, Berkeley, CA 94720 USA, and also with the Simons Institute for the Theory of Computing, Berkeley, CA 94720 USA (e-mail: venkatg@berkeley.edu).

Communicated by I. Tamo, Associate Editor for Coding and Decoding.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2022.3176807.

Digital Object Identifier 10.1109/TIT.2022.3176807

In modern large-scale distributed storage systems (DSS), data is partitioned and stored in individual servers, each with a small storage capacity of a few terabytes. A server can crash any time losing all the data it contains. Less catastrophically, a server often tends to become temporarily unavailable either due to system updates, network bottlenecks, or being busy serving requests of other users. There are thus two design objectives for a DSS. The first one is to never lose user data in the event of crashes (or at least make it highly improbable). The second is to service user requests with low latency despite some servers becoming temporarily unavailable. As the simple approach of replicating data is prohibitive in terms of storage costs, erasure codes are employed in DSS. Using a Reed-Solomon code, if we add $n-k$ parity check servers to k data servers, we can recover user data from any k available servers. But as k gets larger, this does not meet our second objective of servicing user requests with low latency. Local Reconstruction Codes (LRCs) were invented precisely for achieving both the objectives while still maintaining storage efficiency. These codes have *locality* which means that for a small number of erasures, any codeword symbol can be recovered quickly based on a small number of other codeword symbols. At the same time, they can also recover the missing codeword symbols in the unlikely event of a larger number of erasures (but can do so less efficiently). Locality in distributed storage was first introduced in [2], [3], but LRCs were first formally defined and studied in [4] and [5]. Suitably optimized LRCs have been implemented in several large scale systems such as Microsoft Azure [6] and Facebook [7], leading to enormous savings in storage costs and improved system reliability.

An (n, r, h, a, q)-LRC is a linear code over $\mathbb{F}_q$ of length n, whose codeword symbols are partitioned into n/r local groups each of size r. The coordinates in each local group satisfy 'a' local parity checks and there are further h global parity checks that all the n coordinates satisfy. 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 parity check matrix H of an (n, r, h, a, q)-LRC has the structure shown in Equation 1.

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

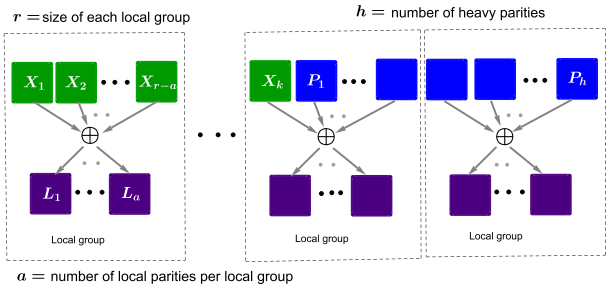

Fig. 1. An LRC with k data symbols, h heavy parities and 'a' local parities per local group. The length of the code $n = k + h + a \cdot \frac{k+h}{r-a}$.

Here $g = n/r$ is the number of local groups. $A_1, A_2, \ldots, A_g$ 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.

Equivalently, from an encoding point of view, an (n, r, h, a, q)-LRC is obtained by adding h global parity checks to k data symbols, partitioning these $k + h$ symbols into local groups of size $r - a$, and then adding 'a' local parity checks for each local group. As a result we have $n = k + h + a \cdot \frac{k+h}{r-a}$ codeword symbols. This is shown in Figure 1.

Information-theoretically, one can show that we can at best hope to correct an additional h erasures distributed across global groups on top of the 'a' erasures in each local group. LRCs which can correct all such erasure patterns which are information-theoretically possible to correct are called *Maximally Recoverable (MR) LRCs*. The notion of maximal recoverability was first introduced by [2], [3] and extended to more general settings in [8]. But MR LRCs were specifically studied first by [9], [10] where they are called *Partial-MDS (Maximum Distance Separable) codes*.

Definition 1: Let C be an arbitrary (n, r, h, a, q)-local reconstruction code. 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.

For a code C with parity check matrix H, an erasure pattern E is correctable iff the submatrix of H formed by columns corresponding the coordinates in E has full column rank. Therefore, we have the following characterization of an MR LRC in terms of its parity check matrix.

Proposition 1: An (n, r, h, a, q)-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.

It is known that MR-LRCs exist over exponentially large fields [4]. This can be easily seen by instantiating the parity check matrix H from Equation 1 randomly from an exponentially large field and verifying that the condition in Proposition 1 is satisfied with high probability by Schwartz-Zippel lemma. But codes deployed in practice require small fields for computational efficiency, typically fields such as $\mathbb{F}_{2^8}$ or $\mathbb{F}_{2^{16}}$ are preferred. Therefore a lot of prior work focused on explicit constructions of MR LRCs over small fields.

TABLE I

TABLE SHOWING THE BEST KNOWN UPPER BOUNDS ON THE FIELD SIZE OF (n, r, h, a, q)-MR LRCS

Field size q	
$O\left(r \cdot n^{(a+1)h-1}\right)$	[11]
$\max\left(O(n/r), O(r)^{\min\{r,h+a\}}\right)^{\min\{h,g\}}$	[12]
$\left(O(\max\{n/r, r\})\right)^{r-a}$	[13]

A. Prior Work

1) Upper Bounds: There are several known constructions of MR LRCs which are incomparable to each other in terms of the field size [1], [8], [10]–[18]. Some constructions are better than others based on the range of parameters. Since there are too many parameters and there is no dominant regime of interest, it is helpful to think about what are the typical ranges of parameters that are useful in deployments of MR LRCs in practice.

2) Parameter Ranges Useful in Practice: One should think of the number of local groups (g) as a constant and n as growing. So $r = n/g$ is growing linearly with n. Typical values of g used in practice are $g = 2, 3, 4$. The number of global parities (h) should also be thought of as a small constant and the number of local parities a is usually 1 or 2. The length n of the code can range from 14 to 60. For example, an early version of Microsoft's Azure storage used $(n = 14, r = 7, h = 2, a = 1)$-MR LRCs with $g = 2$ local groups [6]. These choices are mostly guided by the need to maximize storage efficiency (rate of the code) while balancing durability and fast reconstruction. This is different from the parameters of interest from a theoretical point of view, where to get locality we set r to be sublinear in n.

A few of the important prior constructions that work for all parameter ranges are shown in Table I. The first bound by [11] is good when r is close to n. The second bound by [12] is better when $h \ll r \ll n$. The bound by [13] is better when $r - a \leqslant h$. The construction in [13] is also significantly different from the previous constructions and our construction is inspired by the construction in [13].

In some special cases, there are better constructions. [8] construct MR LRCs over fields of size $O_r\left(n^{\lceil (h-1)(1-1/2^r) \rceil}\right)$ when $a = 1$ and $r = O(1)$. In the special case when $h = 2$,

a construction over linear sized fields for all ranges of other parameters is given in [1].

3) Lower Bounds: The best known lower bounds on the field size required for (n, r, h, a, q)-MR LRCs (with $g = n/r$ local groups) is from [1] who show that for $h \geqslant 2$,

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

The lower bound (2) simplifies to

$$q \geqslant \Omega_{h,a}\left(n r^{\min\{a,h-2\}}\right) \quad (3)$$

when $g = n/r \geqslant h$. When $2 \leqslant h \leqslant \min\{a + 2, g\}$, we have:

$$q \geqslant \Omega_{h}\left(\frac{n(r-a)^{h-1}}{r}\right). \quad (4)$$

Note that the hidden constant in (4) only depends on h.

B. Our Results

We are now ready to present our main result.

Theorem 1 (Main): Let $q_0 \geqslant \max\{g + 1, r - 1\}$ be any prime power where $g = n/r$ is the number of local groups. Then there exists an explicit (n, r, h, a, q)-MR LRC with $q = q_0^{\min\{h,r-a\}}$. Asymptotically, the field size satisfies

$$q \leqslant \left(O\big(\max\{r, n/r\}\big)\right)^{\min\{h,r-a\}}. \quad (5)$$

Our construction is better than (or matches) the first three bounds in Table I for *all* parameter ranges. Moreover when h is a fixed constant with $h \leqslant a + 2$ and $r = \Theta(\sqrt{n})$ and $r - a = \Theta(\sqrt{n})$, our construction matches the lower bound (4), achieving the optimal field size of $\Theta_h(n^{h/2})$. This is the first non-trivial case (other than when $h = 2$ [1]) where we know the optimal field size for MR LRCs.

Corollary 1: Suppose $r = \Theta(\sqrt{n})$, $r - a = \Theta(\sqrt{n})$ and h is a fixed constant independent of n such that $h \leqslant a+2$. Then the optimal field size of an (n, r, h, a, q)-LRC is $q = \Theta_h(n^{h/2})$.

We also remark that the h that appears in the field size upper bound in Theorem 1 can be replaced with h_{local}, if we only want to correct erasure patterns formed by erasing 'a' erasures in each local group and h additional erasures, which are distributed in such a way that no local group has more than $a + h_{\text{local}}$ erasures in total.

MR LRCs used in practice typically have only a small constant number of local groups i.e. $g = n/r$ is typically a small constant such as $g = 2, 3, 4$ [6] and the number of local parities $a = 1$. We can further improve the construction from Theorem 1 in this important regime.

Theorem 2: Suppose the number of local parities $a = 1$ and $g = n/r$ is the number of local groups. Let $q_0 \geqslant g + 1$ be any prime power and let C_0 be any $[r, r - s, d]_{\mathbb{F}_{q_0}}$-code such that its parity check matrix contains a full weight row and it has distance $d \geqslant \min\{h, r - 1\} + 2$.[1] Then there exists an explicit $(n, r, h, a = 1, q)$-MR LRC with field size $q = q_0^{s-1}$. Asymptotically, by instantiating C_0 with BCH codes, we obtain a field size of

$$q \leqslant \left(O(n)\right)^{\lceil \min\{h,r-1\}(1-1/q_0)\rceil}.$$

[1]Equivalently, the dual code $C_{in}^{\perp}$ has a full weight codeword.

We also remark that our constructions can be easily modified to the variant of MR LRCs where the global parities are not protected by the local parity checks. Since we did not define this variant of MR LRCs in this paper, we omit these constructions.

1) Related Work: Shortly before we published our results, we learned that [19] have independently obtained a result analogous to Theorem 1 with a very similar construction. They construct (n, r, h, a, q)-MR LRCs with a field size of

$$q = \left(O\big(\max\{r, n/r\}\big)\right)^{h}. \quad (6)$$

Compared to this, we have a $\min\{h, r-a\}$ in the exponent in our field size bound (5). The construction in the independent work [19] is very similar to ours, we get $\min\{h, r-a\}$ in the exponent by being more careful in our analysis.

Soon after [19], two more constructions of MR LRCs were published by [20] with the following field sizes:

$$q \leqslant \left(\max\left\{(2r)^{r-a}, \frac{g}{r}\right\}\right)^{\min\{h,\lfloor g/r \rfloor\}}, \quad (7)$$

$$q \leqslant (2r)^{r-a}\left(\left\lfloor \frac{g}{r} \right\rfloor + 1\right)^{h-1}. \quad (8)$$

The constructions in (7) and (8) are incomparable to our construction in (5). For example when $r = O(1)$, the construction (8) achieves $O(n)^{h-1}$ field size, whereas our construction achieves $O(n)^{\min\{h,r-a\}}$ field size. In the regime when $r = \Theta(\sqrt{n})$ and $r - a = \Theta(\sqrt{n})$ and $h \leqslant a + 2$ is a fixed constant, our construction achieves the optimal field size of $\Theta_h(n^{h/2})$, whereas the constructions from [20] require fields of size $n^{\Theta(\sqrt{n})}$.

C. Our Techniques

Our constructions are based on the theory of skew polynomials and is inspired by the construction from [13]. Skew polynomials are a non-commutative generalization of polynomials, but they retain many of the familiar and important properties of polynomials. Just as Reed-Solomon codes are constructed using the fact that a degree d polynomial can have at most d roots, our codes will use an analogous theorem that a degree d skew polynomial can have at most d roots *when counted appropriately* (see Theorem 3). Unlike the roots of the usual degree d polynomials which do not have any structure, the roots of degree d skew polynomials have an interesting linear-algebraic structure which we exploit in our constructions. The roots in $\mathbb{F}_{q^m}$ of a degree d skew polynomial over $\mathbb{F}_{q^m}$ can be partitioned into *conjugacy classes* such that the roots in each conjugacy class form a subspace over the base field $\mathbb{F}_q$. Moreover the sum of dimensions of these subspaces across conjugacy classes is at most d.

To exploit this root structure of skew polynomials in an MR LRC construction, we associate each local group with a conjugacy class, and the matrices B_i in (1) are chosen so that $\lambda^T B_i$ is the evaluation of a skew polynomial of degree d (with coefficients given by λ) over different points in the same conjugacy class. Across different local groups, we automatically get linear independence of columns of matrices $B_1, B_2, \ldots, B_g$ as these are associated with different conjugacy classes. Inside

each local group, to argue linear independence, the local parities A_i will be chosen as a Vandermonde matrix over the base field $\mathbb{F}_q$ (we can choose all the A_i's to be equal), and the B_i will be chosen carefully to combine well with the Vandermonde matrix A (see Equations (12), (13), (14)). In particular, we choose B_i so that the $(a+m) \times r$ matrix formed by adding the first row of B_i with entries in $\mathbb{F}_{q^m}$ (but interpreted as an $m \times r$ matrix over the base field $\mathbb{F}_q$) to A_i is an MDS matrix. This allows us to argue that any $a+m$ erasures in that local group can be corrected and we choose $m = \min\{h, r-a\}$. This is also the main difference between our work and [13], which is also implicitly based on skew polynomials.

In this paper, we make this connection explicit in the hope that the theory of skew polynomials will lead to further developments in the constructions of MR LRCs and coding theory more broadly. As an illustration, in Appendix E we show how skew polynomials can give an explanation of algebraic results concerning (generalizations of) Wronskian and Moore matrices that have recently been used in the context of list decoding algorithms for folded Reed-Solomon and univariate multiplicity codes [21], rank condensers [22]–[24], and subspace designs [25], [26]. We also reproduce a construction of maximum sum-rank distance (MSRD) codes due to [27] using the framework of skew polynomials in Appendix F. Skew polynomials have also been explicitly used before to define skew Reed-Solomon codes in [28]. Readers familiar with the theory of skew polynomials or who directly want to get to the construction can skip most of the preliminaries in Section II except for Section II-D.

II. Preliminaries

A. Skew Polynomial Ring

Skew polynomials generalize polynomials while inheriting many of the nice properties of polynomials. Skew polynomials can be defined over division rings[2] and most of the results about skew polynomials are true in this more general setting. It is known that every finite division ring is a field. Since we will only work with skew polynomial rings defined over fields, we will only define them over fields for simplicity. Most of the theory of skew polynomials presented here is from [29], [30], but we reprove the main results in a more accessible way. Skew polynomials were first defined by Ore [31] in 1933 where it was shown that they are the unique non-commutative generalization of polynomials which satisfy (1) associativity (2) distributivity on both sides and (3) the fact that the degree of product of two polynomials is the sum of their degrees.

Let $\mathbb{K}$ be a field. We will first define the key concepts of 'endomorphism' and 'derivation'.

Definition 2 (Endomorphism): A map $\sigma : \mathbb{K} \to \mathbb{K}$ is called an endomorphism if:

1) σ is a linear map i.e. $\sigma(a+b) = \sigma(a) + \sigma(b)$ for all $a, b \in K$ and
2) $\sigma(ab) = \sigma(a)\sigma(b)$ for all $a, b \in K$.

[2]Rings where every non-zero element has a multiplicative inverse, but multiplication may not be commutative.

For example, if $\mathbb{K} = \mathbb{F}_{q^m}$, then $\sigma(x) = x^q$ is an endomorphism called the Frobenius endomorphism. If $\mathbb{K} = \mathbb{F}(x)$ is the field of rational functions and $\gamma \in \mathbb{F}^*$, then $\sigma(f(x)) = f(\gamma x)$ is an endomorphism.

Definition 3 (Derivation): A map $\delta : \mathbb{K} \to \mathbb{K}$ is called a σ-derivation if:

1) δ is a linear map i.e. $\delta(a+b) = \delta(a) + \delta(b)$ for all $a, b \in K$ and
2) $\delta(ab) = \sigma(a)\delta(b) + \delta(a)b$ for all $a, b \in K$.

We will now define the skew polynomial ring.

Definition 4 (Skew polynomial ring): Let σ be an endomorphism of $\mathbb{K}$ and δ be a σ-derivation. The skew polynomial ring in variable t, denoted by $\mathbb{K}[t; \sigma, \delta]$, is a non-commutative ring of skew polynomials in t of the form $\{\sum_{i=0}^{d} a_i t^i : d \geqslant 0, a_i \in \mathbb{K}\}$ (where we always write the coefficients to the left). Degree of a polynomial $f(t) = \sum_i a_i t^i$, denoted by $\deg(f)$, is the largest d such that $a_d \neq 0$.[3] Addition in $\mathbb{K}[t; \sigma, \delta]$ is component wise. But multiplication is distributive and done according to the following rule:

$$\text{For } a \in \mathbb{K},\ t \cdot a = \sigma(a)t + \delta(a). \tag{9}$$

To multiply $f(t)g(t)$, we can first use distributivity to get $f(t)g(t) = \sum_{ij} f_i t^i \cdot g_j t^j$ where $f_i, g_j \in \mathbb{K}$ are coefficients of f, g respectively. Then we use the rule (9) for i times to move the coefficient g_j to the left of t^i. This multiplication turns out to be associative, but may not be commutative. Also $\deg(f \cdot g) = \deg(f) + \deg(g)$. Therefore the skew polynomial ring has no zero divisors. We will now give some examples of skew-polynomials.

The simplest derivation is the zero map i.e. $\delta(a) = 0$ for all $a \in \mathbb{K}$. In this case, the skew polynomial ring is denoted by $\mathbb{K}[t; \sigma]$ and is said to be of endomorphism type. Skew polynomials are interesting even in this case, and in fact the constructions in this paper only use skew polynomials with $\delta \equiv 0$. So the reader can imagine that the derivation is the zero map on a first reading. We include the general case to discuss the applications of skew polynomials to coding and complexity theory later in Appendix E and in the hope that skew polynomial rings with non-zero derivations will find applications in future. For more interesting examples of skew polynomial rings, see Appendix A

We will now collect some simple facts about skew polynomials rings. Let $\mathbb{K}[t; \sigma, \delta]$ be a skew polynomial ring.

Lemma 1 ([29]): $t^n a = \sum_{i=0}^{n} f_i^n(a) t^i$ where $f_0^n = \delta^n$, $f_1^n = \delta^{n-1}\sigma + \delta^{n-2}\sigma\delta + \cdots + \sigma\delta^{n-1}, \ldots, f_n^n = \sigma^n$ are linear maps.

It turns out that the skew polynomial ring has Euclidean algorithm for right division.

Lemma 2 (Euclidean algorithm for right division [29]): For every two polynomial $f, g \in \mathbb{K}[t; \sigma, \delta]$, there exist unique polynomials $q(t), r(t)$ such that $f = q \cdot g + r$ where $\deg(r) < \deg(g)$ or $r = 0$.

This brings us to the most important definition about skew polynomial rings. In the usual polynomial world, we can define the evaluation of a polynomial $f(t) = \sum_i f_i t^i$ at $t = a$ as $\sum_i f_i a^i$. With this definition, it is true that

[3]We will define the degree of the zero polynomial to be ∞.

$f(t) = q(t)(t-a) + f(a)$. But for skew polynomials, these two notions of evaluation differ with each other and the right definition is the second one.

Definition 5 (Evaluation): The evaluation of a polynomial $f \in \mathbb{K}[t;\sigma,\delta]$ at a point $a \in \mathbb{K}$, denoted by $f(a)$, is defined as the remainder obtained when we divide f by $t-a$ on the right i.e. $f(t) = q(t)(t-a) + f(a)$.

Note that evaluation is a linear map i.e. $(f+g)(a) = f(a)+g(a)$. But it is not always true that $(fg)(a) = f(a)g(a)$. We will see shortly how to compute $(fg)(a)$. The evaluation map can be expressed using "power functions," which are the evaluations of monomials of the form t^i.

Definition 6 (Power functions): The power functions are defined inductively as follows. For every $a \in \mathbb{K}$

1) $N_0(a) = 1$ and
2) $N_{i+1}(a) = \sigma(N_i(a))a + \delta(N_i(a))$.

When $\delta \equiv 0$, we have $\mathbb{N}_i(a) = \sigma^{i-1}(a)\sigma^{i-2}(a)\cdots\sigma(a)a$. Additionally if $\sigma \equiv \mathrm{Id}$, then $N_i(a) = a^i$ which explains the terms "power functions."

Lemma 3: Let $f = \sum_i f_i t^i$. Then $f(a) = \sum_i f_i N_i(a)$.

Proof: It is easy to prove by induction that evaluation of t^i at a is $N_i(a)$. The general claim follows by linearity of evaluation. □

We now come to the problem of evaluating $(fg)(a)$. For this, it is useful to define the notion of *conjugates*, which play a big role in this theory.

B. Conjugation and Product Rule

Definition 7 (Conjugation): Let $a \in \mathbb{K}$ and $c \in \mathbb{K}^*$. We define the c-conjugate of a, denoted by ${}^c a$, as

$$ {}^c a = \sigma(c)ac^{-1} + \delta(c)c^{-1}. $$

We say that b is a conjugate of a if there exists some $c \in \mathbb{K}^*$ such that $b = {}^c a$.

We have the following lemma which shows that conjugacy is an equivalence relation, we prove it in Appendix B.

Lemma 4:
1) ${}^d({}^c a) = {}^{dc}a$
2) Conjugacy is an equivalence relation, i.e., we can partition $\mathbb{K}$ into conjugacy classes where elements in each part are conjugates of each other, but elements in different parts are not conjugates.

So $\mathbb{K}$ will get partitioned into conjugacy classes. To understand the structure of each conjugacy class, we need the notion of *centralizer*.

Definition 8 (Centralizer): The centralizer of $a \in \mathbb{K}$ is defined as:

$$ \mathbb{K}_a = \{c \in \mathbb{K}^* : {}^c a = a\} \cup \{0\}. $$

The following lemma shows that centralizers are subfields, we prove it in Appendix B.

Lemma 5:
1) $\mathbb{K}_a$ is a subfield of $\mathbb{K}$.[4]
2) If $a, b \in \mathbb{K}$ are conjugates, then $\mathbb{K}_a = \mathbb{K}_b$. [5]

Because of the above lemma, we can associate a centralizer subfield to each conjugacy class.

[4]When $\mathbb{K}$ is a division ring, $\mathbb{K}_a$ will be a sub-division ring of $\mathbb{K}$.

[5]When $\mathbb{K}$ is a division ring and not a field, we have $\mathbb{K}_{({}^x a)} = x\mathbb{K}_a x^{-1}$.

Example 1: Let $\mathbb{K} = \mathbb{F}_{q^m}$, $\sigma(a) = a^q$ and $\delta \equiv 0$. Then ${}^c a = c^{q-1}a$. Suppose γ is a generator for $\mathbb{F}_{q^m}^*$. There are q equivalence classes, $E_{-1}, E_0, E_1, \ldots, E_{q-2}$, where $E_\ell = \{\gamma^i : i \equiv \ell \mod (q-1).\}$ and $E_{-1} = \{0\}$. The centralizer of an element $a \in \mathbb{K}^*$ is

$$ \mathbb{K}_a = \{c : c^{q-1}a = a\} \cup \{0\} = \{c : c^{q-1} = 1\} \cup \{0\} = \mathbb{F}_q. $$

Therefore the centralizer of every non-zero element is $\mathbb{F}_q$ and the centralizer of 0 is $\mathbb{K}_0 = \mathbb{K}$.

We will now show how to evaluate $(fg)(a)$ using conjugation which plays a key role. The proof of this really important lemma is given in Appendix B.

Lemma 6 (Product evaluation rule [29], [30]): If $g(a) = 0$, then $(fg)(a) = 0$. If $g(a) \neq 0$ then

$$ (fg)(a) = f\left({}^{g(a)}a\right) g(a). $$

Using the product rule, one can prove an interpolation theorem for skew polynomials just like ordinary polynomials. For any $A \subset \mathbb{K}$ be of size n, there exists a non-zero degree $\leqslant n$ skew polynomial $f \in \mathbb{K}[t;\sigma,\delta]$ which vanishes on A [29]. We will later need the following lemma.

Lemma 7: Let f be any skew polynomial. Fix some $a \in \mathbb{K}$. Then $D_{f,a}(y) = f({}^y a)y$ is an $\mathbb{K}_a$-linear map from $\mathbb{K} \to \mathbb{K}$.

Proof: Linearity follows since $f({}^y a)y$ is equal to the evaluation of the polynomial $f(t)y$ at a by Lemma 6. And clearly the evaluation is linear in y. $\mathbb{K}_a$-linearity follows since $\forall c \in \mathbb{K}_a$,

$$ D_{f,a}(yc) = f({}^{yc}a)yc = f({}^y({}^c a))yc = f({}^y a)yc = D_{f,a}(y)c. $$

□

C. Roots of Skew Polynomials

The most important and useful fact about usual polynomials is that a degree d non-zero polynomial can have at most d roots. It turns out that this statement is false for skew polynomials! A skew polynomial can have many more roots than its degree. But when counted in the right way, we can recover an analogous statement for skew polynomials. In this section, we will prove the "fundamental theorem" about roots of skew polynomials which shows that a degree d skew polynomial cannot have more than d roots when counted the right way. Before we state the fundamental theorem, let us try to understand the roots of a skew polynomial in the same conjugacy class. The following lemma shows that they form a vector space over a subfield of $\mathbb{K}$.

Lemma 8: Let $f \in \mathbb{K}[t;\sigma,\delta]$ be a non-zero polynomial and fix some $a \in \mathbb{K}$ and let $\mathbb{F} = \mathbb{K}_a$ be the centralizer of a (which is a subfield of $\mathbb{K}$). Define $V_f(a) = \{y \in \mathbb{K}^* : f({}^y a) = 0\} \cup \{0\}$. Then $V_f(a)$ is a vector space over $\mathbb{F}$.

Proof: For any $\lambda \in \mathbb{F}$ and $y \in V_f(a)$, $f({}^{\lambda y}a) = f({}^y({}^\lambda a)) = f({}^y a) = 0$. Therefore $\lambda y \in V_f(a)$. If $y_1, y_2 \in V_f(a)$ where $y_1 + y_2 \neq 0$, then by Lemma 7, $f({}^{y_1+y_2}a) = 0$. Therefore $y_1 + y_2 \in V_f(a)$. □

We are now ready to state the "fundamental theorem" about roots of skew polynomials, the proof appears in Appendix C.

Theorem 3 ([29], [30]): Let $f \in \mathbb{K}[t;\sigma,\delta]$ be a degree d non-zero polynomial. Let A be the set of roots of f in $\mathbb{K}$ and

let $A = \cup_i A_i$ be a partition of A into conjugacy classes. Fix some representatives $a_i \in A_i$. Let $V_i = \{y : {}^y a_i \in A_i\} \cup \{0\}$ which is a linear subspace over $\mathbb{F}_i = \mathbb{K}_{a_i}$ by Lemma 8. Then

$$\sum_i \dim_{\mathbb{F}_i}(V_i) \leqslant d.$$

In particular, this implies that a non-zero degree d polynomial can have roots in at most d distinct conjugacy classes. And the dimension (over the centralizer subfield) of the subspace of roots in a single conjugacy class is at most d.

D. Vandermonde Matrix

Definition 9 (Vandermonde matrix): Let $A = \{a_1, \ldots, a_n\} \subset \mathbb{K}$. The $d \times n$ Vandermonde matrix formed by A, denoted by $V_d(a_1, \ldots, a_n)$, is defined as:

$$V_d(a_1, \ldots, a_n) = \begin{bmatrix} N_0(a_1) & N_0(a_2) & \cdots & N_0(a_n) \\ N_1(a_1) & N_1(a_2) & \cdots & N_1(a_n) \\ \vdots & \vdots & & \vdots \\ N_{d-1}(a_1) & N_{d-1}(a_2) & \cdots & N_{d-1}(a_n) \end{bmatrix}.$$

When the order of $a_1, a_2, \ldots, a_n$ is not important, we sometimes denote $V_d(a_1, a_2, \ldots, a_n)$ be $V_d(A)$. If $f(t) = \sum_{i=0}^{d-1} f_i t^i$ is a skew polynomial of degree at most $d-1$, then by Lemma 3,

$$[f_0 f_1 \cdots f_{d-1}] \cdot V_d(a_1, a_2, \ldots, a_n) = [f(a_1) f(a_2) \cdots f(a_n)]. \tag{10}$$

Lemma 9: Let $A \subset \mathbb{K}$ of size d. Let $A = A_1 \cup A_2 \cup \cdots \cup A_r$ be the partition of A into different conjugacy classes. Let $n_i = |A_i|$ and let $A_i = \{{}^{c_{ij}} a_i : j \in [n_i]\}$. Then $V_d(A)$ is full rank if for each $i \in [r]$, $\{c_{ij} : j \in [n_i]\}$ are linearly independent over the centralizer subfield $\mathbb{K}_{a_i}$.

Proof: If $V_d(A)$ is not full rank then there exists some non-zero row vector $[f_0 \ f_1 \ \ldots \ f_{d-1}]$ such that $[f_0 \ f_1 \ \ldots \ f_{d-1}] \cdot V_d(A) = 0$. Therefore the non-zero skew polynomial $f(t) = \sum_{i=0}^{d-1} f_i t^i$, with degree at most $d-1$, has roots at all points of A. This violates Theorem 3. □

We will now see two corollaries of Lemma 9 which are useful for our MR LRC construction.

Corollary 2: Let $\gamma \in \mathbb{F}_{q^m}^*$ be a generator of the multiplicative group. Let $d \leqslant q-1$ and $\ell_1, \ldots, \ell_d \in \{0, 1, 2, \ldots, q-2\}$ be distinct. Then the following matrix M is full rank.

$$M = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \gamma^{\ell_1} & \gamma^{\ell_2} & \cdots & \gamma^{\ell_d} \\ \gamma^{\ell_1(1+q)} & \gamma^{\ell_2(1+q)} & & \\ \vdots & \vdots & & \vdots \\ \gamma^{\ell_1(1+q+\cdots+q^{d-2})} & \gamma^{\ell_2(1+q+\cdots+q^{d-2})} & \cdots & \gamma^{\ell_d(1+q+\cdots+q^{d-2})} \end{bmatrix}.$$

Proof: Let $\mathbb{K} = \mathbb{F}_{q^m}$, $\sigma(a) = a^q$ and $\delta \equiv 0$. Then $N_i(a) = a^{1+q+q^2+\cdots+q^{i-1}}$. By Lemma 9, it is enough to show that $\ell_1, \ldots, \ell_d$ fall in distinct conjugacy classes. This is shown in Example 1. □

Note that when $m = 1$, the matrix in the above corollary reduces to the usual Vandermonde matrix one is familiar with.

Corollary 3: Let $\gamma \in \mathbb{F}_{q^m}^*$ be a generator of the multiplicative group and let $\ell \in \{0, 1, \ldots, q-2\}$. Let $\beta_1, \ldots, \beta_m \in \mathbb{F}_{q^m}$ be linearly independent over $\mathbb{F}_q$. Then the following matrix M is full rank.

$$M = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \gamma^{\ell}\beta_1^{q-1} & \gamma^{\ell}\beta_2^{q-1} & \cdots & \gamma^{\ell}\beta_m^{q-1} \\ \gamma^{\ell(1+q)}\beta_1^{q^2-1} & \gamma^{\ell(1+q)}\beta_2^{q^2-1} & \cdots & \gamma^{\ell(1+q)}\beta_m^{q^2-1} \\ \vdots & \vdots & & \vdots \\ \gamma^{\ell(1+q+\cdots+q^{m-2})}\beta_1^{q^{m-1}-1} & \gamma^{\ell(1+q+\cdots+q^{m-2})}\beta_2^{q^{m-1}-1} & \cdots & \gamma^{\ell(1+q+\cdots+q^{m-2})}\beta_m^{q^{m-1}-1} \end{bmatrix}.$$

Proof: Let $\mathbb{K} = \mathbb{F}_{q^m}$, $\sigma(a) = a^q$ and $\delta \equiv 0$. Then $N_i(a) = a^{1+q+q^2+\cdots+q^{i-1}}$. Let $a = \gamma^{\ell}$ then $M = V_m({}^{\beta_1} a, \ldots, {}^{\beta_m} a)$. Therefore M is full rank by Lemma 9. □

III. Skew Polynomials Based MR LRC Constructions

Let us recall that an (n, r, h, a, q)-LRC admits a parity check matrix H of the following form

$$H = \left[\begin{array}{c|c|c|c} A_1 & 0 & \cdots & 0 \\ \hline 0 & A_2 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & A_g \\ \hline B_1 & B_2 & \cdots & B_g \end{array}\right]. \tag{11}$$

Here $A_1, A_2, \cdots, A_g$ are $a \times r$ matrices over $\mathbb{F}_q$ which represent the local parity checks, $B_1, B_2, \cdots, B_g$ are $h \times r$ matrices over $\mathbb{F}_q$ which together represent the h global parity checks. The rest of the matrix is filled with zeros. By Proposition 1, C is an MR LRC iff (1) any 'a' columns of each matrix A_i are linearly independent and (2) any submatrix of H formed by selecting a columns in each local group and any h additional columns is full rank.

A. Construction: Proof of Theorem 1

In this section, we will prove Theorem 1 by presenting a construction of MR LRCs over fields of size $q = O\left(\max(g, r)\right)^{\min\{h, r-a\}}$. The construction presented here is inspired from [13], where they achieve a field size of $O\left(\max(g, r)\right)^{r-a}$.

Let $q_0 \geqslant \max\{g+1, r\}$ be a prime power. Choose $\alpha_1, \alpha_2, \ldots, \alpha_r \in \mathbb{F}_{q_0}$ to be distinct. Define

$$A_\ell = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \alpha_1 & \alpha_2 & \ldots & \alpha_r \\ \alpha_1^2 & \alpha_2^2 & \ldots & \alpha_r^2 \\ \vdots & \vdots & & \vdots \\ \alpha_1^{a-1} & \alpha_2^{a-1} & \ldots & \alpha_r^{a-1} \end{bmatrix}. \tag{12}$$

Note that $A_1 = A_2 = \cdots = A_g$. Let $m = \min\{r-a, h\}$ and let γ be a generator for $\mathbb{F}^*_{q_0^m}$. Our codes will be defined over the field $\mathbb{F}_q = \mathbb{F}_{q_0^m}$. Define $\beta_1, \beta_2, \ldots, \beta_r \in \mathbb{F}_{q_0^m}$ as

$$\beta_i = \begin{bmatrix} \alpha_i^a \\ \alpha_i^{a+1} \\ \vdots \\ \alpha_i^{a+m-1} \end{bmatrix}, \tag{13}$$

where we are expressing β_i in some basis for $\mathbb{F}_{q_0^m}$ (which is a $\mathbb{F}_{q_0}$-vector space of dimension m). The improvement in our construction over [13] comes from choosing β_i carefully in our construction. In [13], β_i are chosen independently of the local parity check matrix A_i and they are chosen to satisfy $(r-a)$-wise independence over the base field $\mathbb{F}_{q_0}$. By choosing them carefully in combination with the local parity check matrix A_i, we only require $m = \min\{h, r-a\}$-wise independence of $\beta_1, \beta_2, \ldots, \beta_r$. Moreover [13] constructs a generator matrix for the code, whereas we construct a parity check matrix.

Define

$$B_\ell = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_r \\ \gamma^\ell \beta_1^{q_0} & \gamma^\ell \beta_2^{q_0} & \cdots & \gamma^\ell \beta_r^{q_0} \\ \gamma^{\ell(1+q_0)} \beta_1^{q_0^2} & \gamma^{\ell(1+q_0)} \beta_2^{q_0^2} & \cdots & \gamma^{\ell(1+q_0)} \beta_r^{q_0^2} \\ \vdots & \vdots & & \vdots \\ \gamma^{\ell(1+q_0+\cdots+q_0^{h-2})} \beta_1^{q_0^{h-1}} & \gamma^{\ell(1+q_0+\cdots+q_0^{h-2})} \beta_2^{q_0^{h-1}} & \cdots & \gamma^{\ell(1+q_0+\cdots+q_0^{h-2})} \beta_r^{q_0^{h-1}} \end{bmatrix}. \tag{14}$$

To prove that the above construction is an MR LRC, we will use properties of the skew field $\mathbb{F}_{q_0^m}[x;\sigma]$ where $\sigma(a) = a^{q_0}$. We know that $\mathbb{F}_{q_0^m}$ will get partitioned into $q_0 - 1$ conjugacy classes as shown in Example 1. If $\gamma \in \mathbb{F}^*_{q_0^m}$ is a generator of $\mathbb{F}^*_{q_0^m}$, then $\{1, \gamma, \gamma^2, \ldots, \gamma^{q_0-2}\}$ fall in distinct conjugacy classes. Intuitively, in the construction each local group corresponds to one conjugacy class. This is possible since we chose $q_0 \geqslant g+1$. The stabilizer subfield of each conjugacy class is $\mathbb{F}_{q_0}$ as shown in Example 1. Therefore we choose the matrices B_i for local group i as a (skew) Vandermonde matrix where the evaluation points $\beta_1, \cdots, \beta_r$ are from the conjugacy class of γ^i, but are linearly independent over the stabilizer subfield $\mathbb{F}_{q_0}$.

Claim 1: The above construction is an MR LRC over fields of size $q = q_0^{\min\{h, r-a\}}$.

Proof: For a matrix M and a subset X of its columns, we will use $M(X)$ to denote the submatrix of M formed by columns in X. Given an erasure pattern E of size $|E| = ag+h$, composed of a erasures in each local group and h additional erasures, we want to argue that the submatrix $H(E)$ is full rank. WLOG, assume that the h additional erasures happen in local groups $1, 2, \ldots, t \in [g]$ for $t \leqslant h$. Let E_i be the set of erasures that happen in the i^{th} local group. Let $S_i \subset E_i$ be an arbitrary subset of size $|S_i| = a$ and let $T_i = E_i \setminus S_i$. Note that $|T_i| \leqslant m$ for all i. We need to show that $H(E)$ (which is an $(ag+h) \times (ag+h)$ matrix) is full rank where

$$H(E) = \left[\begin{array}{c|c|c|c} A_1(S_1 \cup T_1) & 0 & \cdots & 0 \\ \hline 0 & A_2(S_2 \cup T_2) & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & A_g(S_g \cup T_g) \\ \hline B_1(S_1 \cup T_1) & B_2(S_2 \cup T_2) & \cdots & B_g(S_g \cup T_g) \end{array}\right].$$

Note that $A_1(S_1), A_2(S_2), \cdots, A_g(S_g)$ are $a \times a$ matrices of full rank. By doing column operations on $H(E)$, in each local group we can use the columns of $A_i(S_i)$ to remove the columns of $A_i(T_i)$. This results in the lower block $B_i(T_i)$ to change into a Schur complement as follows:

$$\left[\begin{array}{c|c} A_i(S_i) & A_i(T_i) \\ \hline B_i(S_i) & B_i(T_i) \end{array}\right] \rightarrow \left[\begin{array}{c|c} A_i(S_i) & 0 \\ \hline B_i(S_i) & B_i(T_i) - B_i(S_i)A_i(S_i)^{-1}A_i(T_i) \end{array}\right].$$

Note that $T_i = \phi$ for $i > t$. So by doing row and column operations on $H(E)$, we can set it in a block diagonal form, where the diagonal blocks are given by $A_1(S_1), A_2(S_2), \ldots, A_g(S_g)$ and one additional $h \times h$ block given by

$$C = \left[\begin{array}{c|c|c} B_1(T_1) - B_1(S_1)A_1(S_1)^{-1}A_1(T_1) & \cdots & \\ \hline & \cdots & B_t(T_t) - B_t(S_t)A_t(S_t)^{-1}A_t(T_t) \end{array}\right].$$

Note that all the entries in $A(S_i)^{-1}A_i(T_i)$ are in the base field $\mathbb{F}_{q_0}$. Also column operations on B_i with $\mathbb{F}_{q_0}$ coefficients retain its structure with β's replaced by their corresponding $\mathbb{F}_{q_0}$-linear combinations. Therefore by Lemma 9, it is enough to show that the following t matrices $D_1, D_2, \ldots, D_t$ are full rank:

$$D_i = \left[\beta(T_i) - \beta(S_i)A_i(S_i)^{-1}A_i(T_i)\right]$$

where $\beta = [\beta_1, \ldots, \beta_r]$ is a $m \times r$ matrix over $\mathbb{F}_{q_0}$. Note that $[D_1|D_2|\ldots|D_t]$ is just the first row of C (with entries in $\mathbb{F}_{q_0^m}$) expressed as a matrix over $\mathbb{F}_{q_0}$. Consider following matrices given by

$$F_i = \left[\begin{array}{c|c} A_i(S_i) & A_i(T_i) \\ \hline \beta(S_i) & \beta(T_i) \end{array}\right]$$

where each F_i is of size $(a+m) \times (a + |T_i|)$. Each F_i is a Vandermonde matrix by construction. Since $|T_i| \leqslant m$, each F_i is full rank. Now if we do column operations to get F_i into block diagonal form we get:

$$\left[\begin{array}{c|c} A_i(S_i) & 0 \\ \hline \beta(S_i) & \beta(T_i) - \beta(S_i)A_i(S_i)^{-1}A(T_i) \end{array}\right] = \left[\begin{array}{c|c} A_i(S_i) & 0 \\ \hline \beta(S_i) & D_i \end{array}\right].$$

This implies that $D_1, D_2, \ldots, D_t$ are full rank over $\mathbb{F}_{q_0}$ which completes the proof. □

A slightly better construction which only requires $q_0 \geqslant \max\{g+1, r-1\}$ can be obtained by choosing

$$A_\ell = \begin{bmatrix} 1 & \alpha_2^{m+a-1} & \alpha_3^{m+a-1} & \dots & \alpha_r^{m+a-1} \\ 0 & \alpha_2^{m+a-2} & \alpha_3^{m+a-2} & \dots & \alpha_r^{m+a-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \alpha_2^{m+1} & \alpha_3^{m+1} & \dots & \alpha_r^{m+1} \\ 0 & \alpha_2^{m} & \alpha_3^{m} & \dots & \alpha_r^{m} \end{bmatrix}$$

and $\beta_1, \beta_2, \dots, \beta_r \in \mathbb{F}_{q_0}^m$ as:

$$\beta_1 = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \text{ and } \beta_i = \begin{bmatrix} \alpha_i^{m-1} \\ \vdots \\ \alpha_i \\ 1 \end{bmatrix} \text{ for } i \in \{2, 3, \dots, r\}.$$

B. Construction: Proof of Theorem 2

When $a = 1$ and g is a fixed constant, we can improve the construction from the previous section using ideas from BCH codes. Let $q_0 \geqslant g+1$ be a prime power. Define

$$A_\ell = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Note that $A_1 = A_2 = \cdots = A_g$. Let $H_{s\times r}$ be the parity check matrix of the $[r, r-s, d]_{\mathbb{F}_{q_0}}$-code C_0. By scaling the columns of H and permuting the rows (which doesn't change the distance of C_0), we can assume that the first row of H is $[11\cdots1]$. Let $\tilde{H}_{(s-1)\times r}$ be the submatrix of H formed by removing the first row. Now define $\beta_1, \beta_2, \dots, \beta_r \in \mathbb{F}_{q_0}^s$ as the columns of $\tilde{H}$, i.e.,

$$\begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_r \end{bmatrix} = \tilde{H}.$$

Here we are expressing β_i in some basis for $\mathbb{F}_{q_0^{s-1}}$ (which is a $\mathbb{F}_{q_0}$-vector space of dimension $s-1$). Let γ be a generator of $\mathbb{F}^*_{q_0^{s-1}}$. Define B_ℓ as in (14).

Claim 2: The above construction is an MR LRC over fields of size $q = q_0^{s-1}$.

Proof: The proof is analogous to the proof of Claim 1. Let $m = \min\{h, r-1\}$. We only need $\mathbb{F}_{q_0}$-linear independence of any $m+1$ columns of

$$H = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \beta_1 & \beta_2 & \cdots & \beta_r \end{bmatrix}.$$

This follows from the fact that the code C_0 has minimum distance at least $m+2$, and therefore any $m+1$ columns of the parity check matrix H must be linearly independent. □

To get the asymptotic field size bound, we instantiate the code C_0 with BCH codes.

Proposition 2: There exist $[r, r-s, d]_{\mathbb{F}_{q_0}}$ BCH code with

$$s = 1 + \big((d-2) - \lfloor (d-2)/q_0 \rfloor\big) \lceil \log_{q_0} r \rceil.$$

Proof: Let $\ell = \lceil \log_{q_0} r \rceil$ so that $q_0^\ell \geqslant r$. Choose distinct $\theta_1, \theta_2, \dots, \theta_r \in \mathbb{F}_{q_0^\ell}$. The parity check matrix of the BCH code is given by:

$$H_{in} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \theta_1 & \theta_2 & \dots & \theta_r \\ \vdots & \vdots & \ddots & \vdots \\ \theta_1^{q_0-1} & \theta_2^{q_0-1} & \dots & \theta_r^{q_0-1} \\ \theta_1^{q_0+1} & \theta_2^{q_0+1} & \dots & \theta_r^{q_0+1} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_1^{d-2} & \theta_2^{d-2} & \dots & \theta_r^{d-2} \end{bmatrix},$$

where we removed powers which are multiples of q_0. Each row of H other than the first row of 1's should be thought of as ℓ rows over the base field $\mathbb{F}_{q_0}$. Therefore the codimension of the code is $s \leqslant 1 + \ell((d-2) - \lfloor (d-2)/q_0 \rfloor)$. Finally, the distance of the code is at least d. This is because to argue about $\mathbb{F}_{q_0}$ linear independence of any $d-1$ columns, we can add back the rows whose powers are multiples of q_0 to H which is a Vandermonde matrix over $\mathbb{F}_{q_0^\ell}$. □

Therefore we can choose $s = 1 + \big(m - \lfloor m/q_0 \rfloor\big) \lceil \log_{q_0} r \rceil$ where $m = \min\{h, r-1\}$. Therefore we get a field size of

$$q = q_0^{s-1} \leqslant (O(n))^{m - \lfloor m/q_0 \rfloor}.$$

APPENDIX A
EXAMPLES OF SKEW POLYNOMIAL RINGS

In Section II, we discussed a few examples of skew polynomial rings such as when the derivation is the zero map, i.e., $\delta(a) = 0$ for all $a \in \mathbb{K}$. In this case, the skew ring is denoted by $\mathbb{K}[t;\sigma]$ and is said to be of endomorphism type. Here we give a few more interesting examples.

Example 2 (Skew Polynomial Rings):

1) Let $\mathbb{K}$ be any field and let $\sigma : \mathbb{K} \to \mathbb{K}$ be an endomorphism. Then for any $\lambda \in \mathbb{K}$, $\delta(a) = \lambda(\sigma(a) - a)$ is a σ-derivation.[6] These are called inner-derivations and the skew polynomial ring defined using such a derivation is isomorphic to the skew polynomial ring over $\mathbb{K}$ with the same σ and $\delta = 0$.[7] The concept of q-derivatives [32] is a special case of this for $\mathbb{K} = \mathbb{F}(x)$. For some fixed $q \in \mathbb{F} \setminus \{1\}$, the q-derivative $f \in \mathbb{F}(x)$ is defined as $(f(qx) - f(x))/(qx - x)$. This is a derivation w.r.t. the endomorphism $\sigma : f(x) \to f(qx)$.
2) Let $\mathbb{K} = \mathbb{F}(x)$ and σ be the identity map. Then $\delta(f(x))$ defined as the formal derivative of $f(x)$ is a σ-derivation. This can be extended to rational functions in a consistent way using power series. When σ is the identity map, the skew ring is denoted by $\mathbb{K}[t;\delta]$ and is said to be of derivation type.
3) Let $\mathbb{K}$ be the set of smooth real-valued functions over $\mathbb{R}$ and σ be the identity map. Then $\delta(f(x))$ defined as the derivative $f'(x)$ is a σ-derivation. This is an important skew polynomial ring for the study of linear differential equations. For a skew polynomial $g(t) = g_d t^d + \cdots + g_1 t + g_0 \in \mathbb{K}[t;\delta]$ and a smooth function $f : \mathbb{R} \to \mathbb{R}$,

[6] If $\mathbb{K}$ is a division ring, then $\delta(a) = \sigma(a)\lambda - \lambda a$ is a σ-derivation.

[7] The isomorphism is $\phi : \mathbb{K}[t;\sigma,\delta] \to \mathbb{K}[\tilde{t};\sigma]$ defined as $\phi(t) = \tilde{t} - \lambda$ and $\phi|_{\mathbb{K}} \equiv \mathrm{Id}$.

${}^f0$ is a root of $g(t)$ iff f satisfies the linear differential equation

$$g_d D^d f + \cdots + g_1 Df + g_0 f = 0$$

where $D = \frac{d}{dx}$ is the derivative operator. Theorem 3 implies that the set of roots to $g(t)$ forms a vector space of dimension at most d over the centralizer subfield $\mathbb{K}_0 = \{f : {}^f0 = 0\} = \{f : f' = 0\} \cong \mathbb{R}$. This is consistent with the well-known fact that the space of solutions of a degree d homogeneous linear differential equation has dimension at most d.

The following two propositions classify skew polynomial rings over fields and finite fields.

Proposition 3: When $\mathbb{K}$ is a field (as opposed to being a division ring), up to isomorphisms, the only possible skew polynomial rings over $\mathbb{K}$ are either of endomorphism type (i.e., $\delta \equiv 0$) or derivation type (i.e., $\sigma \equiv \mathrm{Id}$).

Proof: This is because if $\sigma \neq \mathrm{Id}$, then there exists some element $a_0 \in \mathbb{K}$ such that $\sigma(a_0) \neq a_0$. Now using commutativity of $\mathbb{K}$, we have $\delta(aa_0) = \delta(a_0 a)$ for any $a \in \mathbb{K}$. Expanding both sides, we get that for any $a \in \mathbb{K}$, $\delta(a) = \lambda(\sigma(a) - a)$ where $\lambda = \delta(a_0)/(\sigma(a_0) - a_0)$ is a fixed constant, i.e., δ is an inner-derivation. As we discussed above, this skew polynomial ring is isomorphic to the skew polynomial ring with $\delta \equiv 0$ and the same endomorphism σ. □

Proposition 4: When $\mathbb{K} = \mathbb{F}_q$ is a finite field, up to isomorphisms, the only possible skew polynomial rings are of the endomorphism type (i.e., $\delta \equiv 0$).

Proof: By Proposition 3, we already know that the skew polynomial ring has to be either of endomorphism type or derivation type. So we just have to rule out the derivation type. Suppose there is a skew polynomial ring of derivation type, i.e., $\sigma \equiv \mathrm{Id}$ and $\delta \neq 0$. Suppose $\mathrm{char}(\mathbb{F}_q) = p$. Then by repeatedly applying chain rule for δ, for any $a \in \mathbb{K}$,

$$\delta(a^p) = a\delta(a^{p-1}) + \delta(a)a^{p-1} = \cdots = p\delta(a)a^{p-1} = 0.$$

This is a contradiction. □

Appendix B
Missing Proofs From Section II

Lemma 10 (Lemma 4): 1) ${}^d({}^c a) = {}^{dc}a$

2) Conjugacy is an equivalence relation, i.e., we can partition $\mathbb{K}$ into conjugacy classes where elements in each part are conjugates of each other, but elements in different parts are not conjugates.

Proof: (1) follows easily from the definition of conjugation and the using the fact that $\delta(cd) = \sigma(c)\delta(d) + \delta(c)d$.

$$\begin{aligned}
{}^d({}^c a) &= \sigma(d) \cdot {}^c a \cdot d^{-1} + \delta(d)d^{-1} \\
&= \sigma(d)(\sigma(c)ac^{-1} + \delta(c)c^{-1})d^{-1} + \delta(d)d^{-1} \\
&= \sigma(dc)ac^{-1}d^{-1} + \sigma(d)\delta(c)c^{-1}d^{-1} + \delta(d)d^{-1} \\
&= \sigma(dc)a(dc)^{-1} + (\sigma(d)\delta(c) + \delta(d)c)c^{-1}d^{-1} \\
&= \sigma(dc)a(dc)^{-1} + \delta(dc)(dc)^{-1} \\
&= {}^{dc}a.
\end{aligned}$$

We now prove (2). Suppose a is a conjugate of b, i.e., $a = {}^x b$ for some $x \in \mathbb{K}^*$. Then ${}^{x^{-1}}a = {}^{x^{-1}}({}^x b) = {}^{x^{-1}x}b = b$. Therefore b is a conjugate of a. Suppose a is a conjugate of b, with $a = {}^x b$, and c is a conjugate of b, with $b = {}^y c$. Then $a = {}^x b = {}^x({}^y c) = {}^{xy}c$. So a is a conjugate of c. □

Lemma 11 (Lemma 5): 1) $\mathbb{K}_a$ is a subfield of $\mathbb{K}$.[8]

2) If $a, b \in \mathbb{K}$ are conjugates, then $\mathbb{K}_a = \mathbb{K}_b$. [9]

Proof: (1) Let $x, y \in \mathbb{K}_a \setminus \{0\}$ i.e. ${}^x a = {}^y a = a$. Then

$$\begin{aligned}
{}^{x+y}a(x+y) &= \sigma(x+y)a + \delta(x+y) \\
&= \sigma(x)a + \sigma(y)a + \delta(x) + \delta(y) \\
&= {}^x ax + {}^y ay \\
&= ax + ay = a(x+y).
\end{aligned}$$

Therefore ${}^{x+y}a = a$. Also ${}^{yx}a = {}^y({}^x a) = a$. And finally ${}^{x^{-1}}a = {}^{x^{-1}}({}^x a) = {}^{x^{-1}x}a = a$.

(2) Suppose $b = {}^d a$ and let $c \in \mathbb{K}_a$.Then ${}^c b = {}^c({}^d a) = {}^{cd}a = {}^{dc}a = {}^d({}^c a) = {}^d a = b$. Therefore $\mathbb{K}_a \subset \mathbb{K}_b$. By symmetry, $\mathbb{K}_b \subset \mathbb{K}_a$. □

Lemma 12 (Product evaluation rule (Lemma 6)): If $g(a) = 0$, then $(fg)(a) = 0$. If $g(a) \neq 0$ then

$$(fg)(a) = f\left({}^{g(a)}a\right) g(a).$$

Proof: If $g(a) = 0$, then $g(t) = b(t)(t-a)$ for some $b(t) \in \mathbb{K}[t; \sigma, \delta]$. Therefore $f(t)g(t) = f(t)b(t)(t-a)$, and so $(fg)(a) = 0$. Suppose $g(a) \neq 0$. Let $g(t) = b(t)(t-a) + g(a)$ and $f(t) = a(t)\left(t - {}^{g(a)}a\right) + f\left({}^{g(a)}a\right)$. Then

$$\begin{aligned}
&f(t)g(t) \\
&= f(t) \cdot (b(t)(t-a) + g(a)) \\
&= f(t)b(t)(t-a) + f(t)g(a) \\
&= f(t)b(t)(t-a) + \left(a(t)\left(t - {}^{g(a)}a\right) + f\left({}^{g(a)}a\right)\right) g(a) \\
&= f(t)b(t)(t-a) + a(t)\left(tg(a) - {}^{g(a)}a \cdot g(a)\right) \\
&\qquad + f\left({}^{g(a)}a\right) g(a) \\
&= f(t)b(t)(t-a) \\
&\qquad + a(t)\left(\sigma(g(a))t + \delta(g(a)) - \sigma(g(a))a - \delta(g(a))\right) \\
&\qquad + f\left({}^{g(a)}a\right) g(a) \\
&= f(t)b(t)(t-a) + a(t)\sigma(g(a))(t-a) + f\left({}^{g(a)}a\right) g(a) \\
&= \left(f(t)b(t) + a(t)\sigma(g(a))\right)(t-a) + f\left({}^{g(a)}a\right) g(a).
\end{aligned}$$

Therefore $(fg)(a) = f\left({}^{g(a)}a\right) g(a)$. □

Appendix C
Roots of Skew Polynomials

The most important and useful fact about usual polynomials is that a degree d non-zero polynomial can have at most d roots. It turns out that this statement is false for skew polynomials! A skew polynomial can have many more roots than its degree. But when counted in the right way, we can recover an analogous statement for skew polynomials. In this section, we will prove the "fundamental theorem" about roots of skew polynomials which shows that a degree d skew

[8]When $\mathbb{K}$ is a division ring, $\mathbb{K}_a$ will be a sub-division ring of $\mathbb{K}$.

[9]When $\mathbb{K}$ is a division ring and not a field, we have $\mathbb{K}_{({}^x a)} = x\mathbb{K}_a x^{-1}$.

polynomial cannot have more than d roots when counted the right way. We will begin with showing that any non-zero degree d skew polynomial can have at most d roots in distinct conjugacy classes.

Lemma 13: Let $f \in \mathbb{K}[t;\sigma,\delta]$ be a degree d non-zero polynomial. Then f can have at most d roots in distinct conjugacy classes.

Proof: We will prove it using induction on the degree. For the base case, it is clear that a degree 0 polynomial which is a non-zero constant cannot have any roots. Suppose $a_0, a_1, \ldots, a_d \in \mathbb{K}$ be roots of f in distinct conjugacy classes. Since $f(a_0) = 0$, we can write $f(t) = h(t)(t - a_0)$ where $\deg(h) = d - 1$. By Lemma 6, $f(a_i) = h({}^{a_i - a_0}a_i)(a_i - a_0)$. Therefore $b_i = {}^{a_i - a_0}a_i$ for $i \in \{1, \ldots, d\}$ are d roots of h and they lie in distinct conjugacy classes because a_i lie in distinct conjugacy classes. Thus by induction $h = 0$ and therefore $f = 0$ which is a contradiction. □

Now let us try to understand, the roots of a skew polynomial in the same conjugacy class. Let $f \in \mathbb{K}[t;\sigma,\delta]$ be a non-zero polynomial and fix some $a \in \mathbb{K}$ and let $\mathbb{K}_a$ be the centralizer of a (which is a subfield of $\mathbb{K}$). Define $V_f(a) = \{y \in \mathbb{K}^* : f({}^y a) = 0\} \cup \{0\}$. Lemma 8 shows that $V_f(a)$ is a vector space over $\mathbb{K}_a$. The next lemma shows that the dimension of $V_f(a)$ can be at most $\deg(f)$.

Lemma 14: Let $f \in \mathbb{K}[t;\sigma,\delta]$ be a degree d non-zero polynomial and fix some $a \in \mathbb{K}$ and let $\mathbb{F} = \mathbb{K}_a$ be the centralizer subfield of a. Define $V_f(a) = \{y \in \mathbb{K}^* : f({}^y a) = 0\} \cup \{0\}$. Then $V_f(a)$ is a vector space over $\mathbb{F}$ of dimension at most d.

Proof: We will use induction on the degree. For the base case, it is clear that for a degree 0 polynomial, which is a non-zero constant, $\dim_{\mathbb{F}}(V_f(a)) = 0$. Suppose for contradiction that there exists $y_0, y_1, \ldots, y_d \in V_f(a)$ which are linearly independent over $\mathbb{F}$. WLOG, we can assume that $y_0 = 1$ (by redefining a to be equal to ${}^{y_0}a$). Since $f(a) = 0$, we can write $f(t) = h(t)(t - a)$ where $\deg(h) = d - 1$. By Lemma 6, $f({}^{y_i}a) = h({}^{y_i({}^{y_i}a - a)}a)({}^{y_i}a - a)$. Since $y_0 = 1$ and y_i is linearly independent from y_0 over $\mathbb{F}$, $y_i \notin \mathbb{F}$. Therefore ${}^{y_i}a - a \neq 0$, and so $b_i = {}^{y_i({}^{y_i}a - a)}a$ for $i \in \{1, \ldots, d\}$ are d roots of h. If we show that $y_i({}^{y_i}a - a)$ for $i \in \{1, \ldots, d\}$ are linearly independent over $\mathbb{F}$, then we are done by induction.

Suppose they are not independent. Then there exists $c_1, \ldots, c_d \in \mathbb{F}$ s.t. $\sum_{i=1}^{d} c_i y_i({}^{y_i}a - a) = 0$. Therefore,

$$\begin{aligned} a\sum_{i=1}^{d} c_i y_i &= \sum_{i=1}^{d} c_i y_i \cdot {}^{y_i}a \\ &= \sum_{i=1}^{d} c_i y_i \cdot {}^{c_i y_i}a \qquad (c_i \in \mathbb{F} = \mathbb{K}_a) \\ &= \left(\sum_{i=1}^{d} c_i y_i\right)\left({}^{\sum_{i=1}^{d} c_i y_i}\right)a \\ &\qquad ({}^{x+y}a(x+y) = {}^{x}ax + {}^{y}ay \text{ for all } x, y \in \mathbb{K}^*) \end{aligned}$$

Since $y_1, \ldots, y_d$ are independent over $\mathbb{F}$, $\sum_{i=1}^{d} c_i y_i \neq 0$. Therefore ${}^{\left(\sum_{i=1}^{d} c_i y_i\right)}a = a$ i.e. $\sum_{i=1}^{d} c_i y_i \in \mathbb{K}_a = \mathbb{F}$. But this contradicts the fact that $\{y_0 = 1, y_1, \ldots, y_d\}$ are linearly independent over $\mathbb{F}$. □

We will now prove the "fundamental theorem" about roots of skew polynomials. It immediately implies Lemma 13 and Lemma 14 as corollaries. But we have proved them before, just to convey some intuition.

Theorem 4 (Theorem 3): Let $f \in \mathbb{K}[t;\sigma,\delta]$ be a degree d non-zero polynomial. Let A be the set of roots of f in $\mathbb{K}$ and let $A = \cup_i A_i$ be a partition of A into conjugacy classes. Fix some representatives $a_i \in A_i$. Let $V_i = \{y : {}^{y}a_i \in A_i\} \cup \{0\}$ which is a linear subspace over $\mathbb{F}_i = \mathbb{K}_{a_i}$ by Lemma 8. Then

$$\sum_i \dim_{\mathbb{F}_i}(V_i) \leqslant d.$$

Proof: We will use induction on the degree. For the base case, it is clear that for a degree 0 polynomial, which is a non-zero constant, $\dim_{\mathbb{F}_i}(V_i) = 0$ for every i. We will now show the induction step.

For each i, let $d_i = \dim_{\mathbb{F}_i}(V_i)$. Fix some basis $y(i,1), y(i,2), \ldots, y(i,d_i) \in \mathbb{K}^*$ which span V_i with coefficients in $\mathbb{F}_i = \mathbb{K}_{a_i}$. WLOG, we can assume that $y(i,1) = 1$ for every i, by reassigning $a_i = {}^{y(i,1)}a_i$.

Fix some conjugacy class i^* s.t. $d_{i*} \geqslant 1$. Since $f(a_{i^*}) = 0$, we can write $f(t) = h(t)(t - a_{i^*})$ where $\deg(h) = d - 1$. Now let A'_i be the roots of h in conjugacy class i and $V'_i = \{y : {}^{y}a_i \in A'_i\} \cup \{0\}$. We claim that $\dim_{\mathbb{F}_i}(V'_i) \geqslant \dim_{\mathbb{F}_i}(V_i)$ for every $i \neq i^*$ and $\dim_{\mathbb{F}_{i^*}}(V'_{i^*}) \geqslant \dim_{\mathbb{F}_{i^*}}(V_{i^*}) - 1$. By induction $\sum_i \dim_{\mathbb{F}_i}(V'_i) \leqslant d - 1$. Therefore we have $\sum_i \dim_{\mathbb{F}_i}(V_i) \leqslant d$. We will now prove the claim in two parts.

Claim 3: $\dim_{\mathbb{F}_i}(V'_i) \geqslant \dim_{\mathbb{F}_i}(V_i)$ for every $i \neq i^*$.

Proof: Fix some conjugacy class $i \neq i^*$. By Lemma 6,

$$f\left({}^{y(i,j)}a_i\right) = h\left({}^{y(i,j)({}^{y(i,j)}a_i - a_{i^*})}a_i\right)\left({}^{y(i,j)}a_i - a_{i^*}\right).$$

Since a_i, a_{i^*} are in different conjugacy classes, ${}^{y(i,j)}a_i - a_{i^*} \neq 0$. So $b_j = {}^{y(i,j)({}^{y(i,j)}a_i - a_{i^*})}a_i$ for $j \in \{1, \ldots, d_i\}$ are d_i roots of h in the i^{th} conjugacy class A'_i. If we show that $y(i,j)({}^{y(i,j)}a_i - a_{i^*})$ for $j \in \{1, \ldots, d_i\}$ are linearly independent over $\mathbb{F}_i$, then this proves the claim.

Suppose they are not independent. Then there exists $c_1, \ldots, c_{d_i} \in \mathbb{F}_i$ s.t. $\sum_{j=1}^{d_i} c_j y(i,j)({}^{y(i,j)}a_i - a_{i^*}) = 0$. Therefore,

$$\begin{aligned} a_{i^*}\sum_{j=1}^{d_i} c_j y(i,j) &= \sum_{j=1}^{d_i} c_j y(i,j) \cdot {}^{y(i,j)}a_i \\ &= \sum_{j=1}^{d_i} c_j y(i,j) \cdot {}^{c_j y(i,j)}a_i \quad (c_j \in \mathbb{F}_i = \mathbb{K}_{a_i}) \\ &= \left(\sum_{i=1}^{d_i} c_j y(i,j)\right)\left({}^{\sum_{j=1}^{d_i} c_j y(i,j)}\right)a_i \\ &\qquad ({}^{x+y}a(x+y) = {}^{x}ax + {}^{y}ay \text{ for all } x, y \in \mathbb{K}^*) \end{aligned}$$

Since $y(i,1), \ldots, y(i,d_i)$ are independent over $\mathbb{F}_i$, $\sum_{j=1}^{d_i} c_j y(i,j) \neq 0$. Therefore ${}^{\left(\sum_{j=1}^{d_i} c_j y(i,j)\right)}a_i = a_{i^*}$. This is a contradiction because a_i, a_{i^*} are in different conjugate classes. □

Claim 4: $\dim_{\mathbb{F}_{i^*}}(V'_{i^*}) \geqslant \dim_{\mathbb{F}_{i^*}}(V_{i^*}) - 1$.

Proof: The proof is exactly similar to that of the previous claim, up until the last. Let $j \in \{2, 3, \ldots, d_{i^*}\}$. By Lemma 6,

$$f\left(^{y(i^*,j)}a_{i^*}\right) = h\left(^{y(i^*,j)\left(^{y(i^*,j)}a_{i^*}-a_{i^*}\right)}a_{i^*}\right)\left(^{y(i^*,j)}a_{i^*} - a_{i^*}\right).$$

Since $y(i^*, 1) = 1$ and $y(i^*, j)$ are linearly independent over $\mathbb{F}_{i^*}$, $y(i^*, j) \notin \mathbb{F}_{i^*}$. Therefore $^{y(i^*,j)}a_{i^*} - a_{i^*} \neq 0$. So $b_j = {}^{y(i^*,j)(^{y(i^*,j)}a_{i^*}-a_{i^*})}a_{i^*}$ for $j \in \{2, \ldots, d_{i^*}\}$ are $d_{i^*} - 1$ roots of h in the i^{*th} conjugacy class A'_{i^*}. If we show that $y(i^*, j)(^{y(i^*,j)}a_{i^*} - a_{i^*})$ for $j \in \{2, \ldots, d_{i^*}\}$ are linearly independent over $\mathbb{F}_{i^*}$, then this proves the claim.

Suppose they are not independent. Then there exists $c_2, \ldots, c_{d_{i^*}} \in \mathbb{F}_{i^*}$ s.t.

$$\sum_{j=2}^{d_{i^*}} c_j y(i^*, j)(^{y(i^*,j)}a_{i^*} - a_{i^*}) = 0\ .$$

Therefore,

$$\begin{aligned} a_{i^*}\sum_{j=2}^{d_{i^*}} c_j y(i^*, j) &= \sum_{j=2}^{d_{i^*}} c_j y(i^*, j) \cdot {}^{y(i^*,j)}a_{i^*} \\ &= \sum_{j=2}^{d_{i^*}} c_j y(i^*, j) \cdot {}^{c_j y(i^*,j)}a_{i^*} \\ &\qquad (c_j \in \mathbb{F}_{i^*} = K_{a_{i^*}}) \\ &= \left(\sum_{j=2}^{d_{i^*}} c_j y(i^*, j)\right)^{\left(\sum_{j=2}^{d_{i^*}} c_j y(i^*,j)\right)} a_{i^*} \\ &\qquad ({}^{x+y}a(x+y) = {}^{x}ax + {}^{y}ay \text{ for all } x, y \in \mathbb{K}^*) \end{aligned}$$

Since $y(i^*, 1), \ldots, y(i^*, d_{i^*})$ are independent over $\mathbb{F}_{i^*}$, $\sum_{j=2}^{d_{i^*}} c_j y(i^*, j) \neq 0$. Therefore

$$\left(^{\sum_{j=2}^{d_{i^*}} c_j y(i^*,j)}\right) a_{i^*} = a_{i^*}\ ,$$

and thus $\sum_{j=2}^{d_{i^*}} c_j y(i^*, j) \in K_{a_{i^*}} = \mathbb{F}_{i^*}$. But this contradicts the fact that

$$\{y(i^*, 1) = 1, y(i^*, 2), \ldots, y(i^*, d_{i^*})\}$$

are linearly independent over $\mathbb{F}_{i^*}$. □

The above two claims finish the proof of Theorem 3. □

Appendix D
Constructions of MR LRCs Where Global Parities Are Outside Local Groups

Sometimes, it is better to keep the global parities outside the local groups, i.e., the global/heavy parities do not participate in any local groups. For a given length of the code, this reduces the size of local groups and therefore improves the reconstruction performance (at the cost of slight decrease in durability). Figure 2 shows such an MR LRC. The encoding is done by partitioning the k data symbols into g local groups of size $r - a$ each and adding 'a' local parities per local group. There are a total of g local groups. Further an additional h global parity checks are added which are placed outside the local groups. The length of the code is therefore $n = k+h+a\cdot\frac{k}{r-a}$. The parity check matrix of an (n, r, h, a, q)-

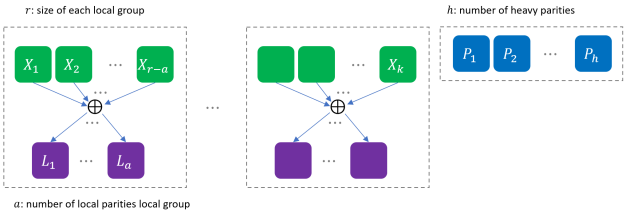

Fig. 2. An LRC with k data symbols, h global/heavy parities and 'a' local parities per local group. The global parities are outside the local groups. The length of the code $n = k + h + a \cdot \frac{k}{r-a}$.

MR LRC where the global parities are outside local groups is of the following form:

$$H = \left[\begin{array}{c|c|c|c|c} A_1 & 0 & \cdots & 0 & 0 \\ \hline 0 & A_2 & \cdots & 0 & 0 \\ \hline \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline 0 & 0 & \cdots & A_g & 0 \\ \hline B_1 & B_2 & \cdots & B_g & B_{global} \end{array}\right]. \tag{15}$$

Here $g = n/r$ is the number of local groups. $A_1, A_2, \ldots, A_g$ 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 B_{global} is a $h \times h$ matrix; together they represent the h global parity checks that the codewords should satisfy.

The set of correctable erasure patterns correctable by such an MR LRC are exactly those obtained by erasing 'a' symbols per local group and h additional symbols arbitrarily. Our constructions can be easily modified to obtain the constructions in this setting as well. For simplicity, we will only state the theorems for the case when $h \leqslant r - a$, since this is the regime that is commonly used in practice. The constructions can be easily modified to also work when $h > r - a$.

Theorem 5: Suppose $h \leqslant r - a$. Let q_0 be any prime power such that one of the following is true:

1) $q_0 \geqslant \max\{g + 2, r - 1\}$ or
2) $q_0 \geqslant \max\{g + 1, r + \lceil (h - 1)/g \rceil - 1\}$

where $g = n/r$ is the number of local groups. Then there exists an explicit (n, r, h, a, q)-MR LRC with $q = q_0^h$.

MR LRCs used in practice typically have only one local parity per local group, i.e., $a = 1$ [6]. We can further improve the construction from Theorem 5 in this regime.

Theorem 6: Suppose $h \leqslant r - a$ and the number of local parities $a = 1$. Choose a prime power q_0 and a positive integer n_0 such that one of the following is true:

1) $q_0 \geqslant g + 2$ and $n_0 = r$ or
2) $q_0 \geqslant g + 1$ and $n_0 = r + \lceil \frac{h-1}{g} \rceil$.

Suppose there exists an $[n_0, n_0 - c, d]_{\mathbb{F}_{q_0}}$ linear code C_0 where c is its codimension and minimum distance $d \geqslant h + 2$. Further we need the dual code $C_0^{\perp}$ to have a codeword of weight exactly r.[10] Then there exists an explicit

[10]This is equivalent to C_0 having a parity check matrix containing a row with exactly r non-zero entries.

$(n, r, h, a = 1, q)$-MR LRC with field size

$$q = q_0^{c-1}$$

where the global parities are outside the local groups.

A. Construction: Proof of Theorem 5

Let us recall the parity check matrix of an (n, r, h, a, q)-LRC where the global parities are outside local groups is of the form given in (15). By Proposition 1, C is an MR LRC iff (1) any 'a' columns of each matrix A_i are linearly independent and (2) any submatrix of H formed by selecting a columns in each local group and any h additional columns is full rank.

Case 1: Let $q_0 \geqslant \max\{g+2, r-1\}$ be a prime power.

Since we have one extra conjugacy class (note that $q_0 - 1 \geqslant g+1$), we will use it to define B_{global}. Let H_0 be an $(a+h)\times r$ MDS matrix over $\mathbb{F}_{q_0}$ which can be constructed using a Reed-Solomon code. Partition H_0 as follows:

$$H_0 = \left[\begin{array}{c} A_{a\times r} \\ \hline \beta_{h\times r} \end{array}\right].$$

Define $A_1, A_2, \ldots, A_g = A$ where A is formed by the first a rows of H_0. Let $\beta_1, \beta_2, \ldots, \beta_r \in \mathbb{F}_{q_0}^h$ be the columns of $\beta_{h\times r}$, i.e.,

$$\beta = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_r \end{bmatrix}.$$

Let $\widetilde{\beta}_i = e_i \in F_{q_0}^h$ for $1 \leqslant i \leqslant h$ where $e_1, e_2, \ldots, e_h$ are coordinate basis vectors. Note that $\beta_1, \ldots, \beta_r$ and $\widetilde{\beta}_1, \ldots, \widetilde{\beta}_h$ can also be thought of as elements of $\mathbb{F}_{q_0^h}$ by fixing some basis of $\mathbb{F}_{q_0^h}$ as a vector space over $\mathbb{F}_{q_0}$.

Define for $1 \leqslant \ell \leqslant g$,

$$B_\ell = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_r \\ \gamma^\ell \beta_1^{q_0} & \gamma^\ell \beta_2^{q_0} & \cdots & \gamma^\ell \beta_r^{q_0} \\ \gamma^{\ell(1+q_0)} \beta_1^{q_0^2} & \gamma^{\ell(1+q_0)} \beta_2^{q_0^2} & \cdots & \gamma^{\ell(1+q_0)} \beta_r^{q_0^2} \\ \vdots & \vdots & & \vdots \\ \gamma^{\ell(1+q_0+\cdots+q_0^{h-2})} \beta_1^{q_0^{h-1}} & \gamma^{\ell(1+q_0+\cdots+q_0^{h-2})} \beta_2^{q_0^{h-1}} & \cdots & \gamma^{\ell(1+q_0+\cdots+q_0^{h-2})} \beta_r^{q_0^{h-1}} \end{bmatrix}. \tag{16}$$

$$B_{global} = \begin{bmatrix} \widetilde{\beta}_1 & \widetilde{\beta}_2 & \cdots & \widetilde{\beta}_h \\ \gamma^{(g+1)} \widetilde{\beta}_1^{q_0} & \gamma^{(g+1)} \widetilde{\beta}_2^{q_0} & \cdots & \gamma^{(g+1)} \widetilde{\beta}_h^{q_0} \\ \gamma^{(g+1)(1+q_0)} \widetilde{\beta}_1^{q_0^2} & \gamma^{(g+1)(1+q_0)} \widetilde{\beta}_2^{q_0^2} & \cdots & \gamma^{(g+1)(1+q_0)} \widetilde{\beta}_h^{q_0^2} \\ \vdots & \vdots & & \vdots \\ \gamma^{(g+1)(1+q_0+\cdots+q_0^{h-2})} \widetilde{\beta}_1^{q_0^{h-1}} & \gamma^{(g+1)(1+q_0+\cdots+q_0^{h-2})} \widetilde{\beta}_2^{q_0^{h-1}} & \cdots & \gamma^{(g+1)(1+q_0+\cdots+q_0^{h-2})} \widetilde{\beta}_h^{q_0^{h-1}} \end{bmatrix}. \tag{17}$$

Claim 5: The above construction is an MR LRC over fields of size $q = q_0^h$.

Proof Sketch: The proof is very similar to that of Claim 1. The only difference is that some of the h additional erasures can happen in the global parities. Since we defined B_{global} so that it belongs to a conjugacy class distinct from those of $B_1, B_2, \ldots, B_g$, the proof follows similarly. □

Case 2: Let $q_0 \geqslant \max\{g+1, r+\lceil \frac{h-1}{g} \rceil\}$ be a prime power.

In this case, we don't have an extra (non-zero) conjugacy class to define B_{global}. Therefore, we will partition B_{global} into g parts and fold in the parts into the existing g conjugacy classes. Note that we can always include the last column of B_{global} as $[0, 0, \ldots, 1]^T$. Therefore we only need to fold in $h-1$ columns of B_{global} into existing g conjugacy classes. Let $t = \lceil \frac{h-1}{g} \rceil$ and let $n_0 = r + t$. Let H_0 be an $(a+h) \times n_0$ MDS matrix over $\mathbb{F}_{q_0}$ of the following form

$$H_0 = \left[\begin{array}{c|c} A_{a\times r} & 0 \\ \hline \beta_{h\times r} & \widetilde{\beta}_{h\times t} \end{array}\right]. \tag{18}$$

Note that we can construct an MDS matrix of this form, by first starting with a Reed-Solomon MDS matrix over $\mathbb{F}_{q_0}$ and doing row operations to get this form. Moreover, note that the matrix $A_{a\times r}$ is itself an MDS matrix.

Define $A_1, A_2, \ldots, A_g = A$. Let $\beta_1, \beta_2, \ldots, \beta_r \in \mathbb{F}_{q_0}^h$ be the columns of $\beta_{h\times r}$, i.e.,

$$\beta = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_r \end{bmatrix}.$$

and define $\widetilde{\beta}_1, \ldots, \widetilde{\beta}_t \in \mathbb{F}_{q_0}^h$ be the columns of $\widetilde{\beta}_{h\times t}$. Note that $\beta_1, \ldots, \beta_r$ and $\widetilde{\beta}_1, \ldots, \widetilde{\beta}_t$ can also be thought of as elements of $\mathbb{F}_{q_0^h}$ by fixing some basis of $\mathbb{F}_{q_0^h}$ as a vector space over $\mathbb{F}_{q_0}$.

Define for $1 \leqslant \ell \leqslant g$, B_ℓ as in (16) and

$$B_{global}^{\ell} = \begin{bmatrix} \widetilde{\beta}_1 & \widetilde{\beta}_2 & \cdots & \widetilde{\beta}_t \\ \gamma^\ell \widetilde{\beta}_1^{q_0} & \gamma^\ell \widetilde{\beta}_2^{q_0} & \cdots & \gamma^\ell \widetilde{\beta}_t^{q_0} \\ \gamma^{\ell(1+q_0)} \widetilde{\beta}_1^{q_0^2} & \gamma^{\ell(1+q_0)} \widetilde{\beta}_2^{q_0^2} & \cdots & \gamma^{\ell(1+q_0)} \widetilde{\beta}_t^{q_0^2} \\ \vdots & \vdots & & \vdots \\ \gamma^{\ell(1+q_0+\cdots+q_0^{h-2})} \widetilde{\beta}_1^{q_0^{h-1}} & \gamma^{\ell(1+q_0+\cdots+q_0^{h-2})} \widetilde{\beta}_2^{q_0^{h-1}} & \cdots & \gamma^{\ell(1+q_0+\cdots+q_0^{h-2})} \widetilde{\beta}_t^{q_0^{h-1}} \end{bmatrix}. \tag{19}$$

Define

$$\widetilde{B}_{global} = \left[\begin{array}{c|c|c|c|c} B_{global}^1 & B_{global}^2 & \cdots & B_{global}^g & e_h \end{array}\right] \tag{20}$$

where $e_h \in F_{q_0}^h$ is the coordinate vector $e_h = [0, 0, \ldots, 0, 1]^T$. Note that $\widetilde{B}_{global}$ is an $h \times (gt+1)$ matrix and $gt + 1 \geqslant h$. Finally define B_{global} to be an $h \times h$ matrix formed by arbitrary h columns of $\widetilde{B}_{global}$.

Claim 6: The above construction is an MR LRC over fields of size $q = q_0^h$.

Proof Sketch: The proof is very similar to that of Claim 1. The only difference is that some of the h additional erasures can happen in the global parities. We will need to crucially

use the fact the top right corner of the matrix H_0 in (18) used to define A's and B's is zero. Therefore using some 'a' columns of A to remove the rest of the columns in the upper half of H_0, does not affect the $\widetilde{\beta}$ matrix since the top right corner is already forced to be zero. □

B. Construction: Proof of Theorem 6

Let us recall the parity check matrix of an (n, r, h, a, q)-LRC where the global parities are outside local groups is of the form given in (15). By Proposition 1, C is an MR LRC iff (1) any 'a' columns of each matrix A_i are linearly independent and (2) any submatrix of H formed by selecting a columns in each local group and any h additional columns is full rank.

Case 1: $q_0 \geqslant g+2$ and $n_0 = r$.

Since we have one extra conjugacy class (note that $q_0 - 1 \geqslant g+1$), we will use it to define B_{global}. Let C_0 be an $[n_0, n_0-c, d]_{\mathbb{F}_{q_0}}$ linear code with minimum distance $d \geqslant h+2$. Let H_0 be the parity check matrix of C_0 which is a $c \times r$ matrix over $\mathbb{F}_{q_0}$. By the hypothesis that the dual code of C_0 has a full weight vector, we can assume that the first row of H_0 is all 1's vector (scaling the columns if necessary). Partition H_0 as follows:

$$H_0 = \left[\begin{array}{cccc} 1 & 1 & \cdots & 1 \\ \hline \beta_1 & \beta_2 & \cdots & \beta_r \end{array}\right].$$

Define

$$A_1, A_2, \ldots, A_g = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Note that $c \geqslant h+1$ since any $h+1$ columns of H_0 are linearly independent. Therefore we have $h \leqslant c-1$, and so we can define for $1 \leqslant i \leqslant h$,

$$\widetilde{\beta}_i = e_i$$

where $e_i \in \mathbb{F}_{q_0}^{c-1}$ is the i^{th} coordinate basis vector.

Note that $\beta_1, \ldots, \beta_r$ and $\widetilde{\beta}_1, \ldots, \widetilde{\beta}_h$ can also be thought of as elements of $\mathbb{F}_{q_0^{c-1}}$ by fixing some basis of $\mathbb{F}_{q_0^{c-1}}$ as a vector space over $\mathbb{F}_{q_0}$.

Define for $1 \leqslant \ell \leqslant g$, B_ℓ as in (16) and B_{global} as in (17).

Claim 7: The above construction is an MR LRC over fields of size $q = q_0^{c-1}$.

Proof Sketch: The proof is very similar to that of Claim 1. The only difference is that some of the h additional erasures can happen in the global parities. Since we defined B_{global} so that it belongs to a conjugacy class distinct from those of $B_1, B_2, \ldots, B_g$, the proof follows similarly. We will also use the fact that C_0 has minimum distance at least $h+2$ and so any $h+1$ columns of H_0 are linearly independent. □

Case 2: $q_0 \geqslant g+1$ and $n_0 = r + \lceil \frac{h-1}{g} \rceil$.

In this case, we don't have an extra (non-zero) conjugacy class to define B_{global}. Therefore, we will partition B_{global} into g parts and fold in the parts into the existing g conjugacy classes. Note that we can always include the last column of B_{global} as $[0, 0, \ldots, 1]^T$. Therefore we only need to fold in $h-1$ columns of B_{global} into existing g conjugacy classes. Let $t = \lceil \frac{h-1}{g} \rceil$ and let $n_0 = r + t$.

Let C_0 be an $[n_0, n_0 - c, d]_{\mathbb{F}_{q_0}}$ linear code with minimum distance $d \geqslant h+2$. Let H_0 be the parity check matrix of C_0 which is a $c \times n_0$ matrix over $\mathbb{F}_{q_0}$. By the hypothesis that the dual code of C_0 has a vector of weight exactly r, we can assume that the first row of H_0 has exactly r ones and t zeros (after scaling the columns if necessary). Partition H_0 as follows:

$$H_0 = \left[\begin{array}{cccc|cccc} 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ \beta_1 & \beta_2 & \cdots & \beta_r & \widetilde{\beta}_1 & \widetilde{\beta}_2 & \cdots & \widetilde{\beta}_t \end{array}\right]. \tag{21}$$

Define

$$A_1, A_2, \ldots, A_g = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Note that $\beta_1, \ldots, \beta_r$ and $\widetilde{\beta}_1, \ldots, \widetilde{\beta}_t$ can also be thought of as elements of $\mathbb{F}_{q_0^{c-1}}$ by fixing some basis of $\mathbb{F}_{q_0^{c-1}}$ as a vector space over $\mathbb{F}_{q_0}$. Define for $1 \leqslant \ell \leqslant g$, B_ℓ as in (16), B_{global}^{ℓ} as in (19) and $\widetilde{B}_{global}$ as in (20). Note that $\widetilde{B}_{global}$ is an $h \times (gt+1)$ matrix and $gt+1 \geqslant h$. Finally define B_{global} to be an $h \times h$ matrix formed by arbitrary h columns of $\widetilde{B}_{global}$.

Claim 8: The above construction is an MR LRC over fields of size $q = q_0^{c-1}$.

Proof Sketch: The proof is very similar to that of Claim 1. The only difference is that some of the h additional erasures can happen in the global parities. We will need to crucially use the fact the top right corner of the matrix H_0 in (21) is zero. Therefore using some 'a' ones to remove the rest of the ones in the upper half of H_0, does not affect the $\widetilde{\beta}$ matrix since the top right corner is already forced to be zero. □

APPENDIX E
SKEW POLYNOMIAL WRONSKIAN AND MOORE MATRICES

In this section, we will discuss generalizations of Wronskian and Moore matrices using skew polynomials. The non-singularity of special cases of these matrices has been instrumental in works on list decoding [21], [33] and algebraic pseudorandomness such as constructions of rank condensers and subspace designs [24]–[26]. We will need the following simple lemmas.

Lemma 15: Let $\mathbb{F}(x)$ be the field of rational functions in x and let $\mathbb{L} = \mathbb{F}(x^r)$ which is a subfield of $\mathbb{F}(x)$.[11] Let $g_1, g_2, \ldots, g_m \in \mathbb{F}[x]^{<r}$ be polynomials of degree strictly less than r. Then $g_1, g_2, \ldots, g_m$ are $\mathbb{L}$-linearly independent iff they are $\mathbb{F}$-linearly independent.

Proof: One direction is obvious since $\mathbb{F}$ is a subfield of $\mathbb{L}$. To prove the other direction, suppose $g_1, g_2, \ldots, g_m$ are $\mathbb{L}$-linearly dependent, i.e., $\sum_i c_i(x^r) g_i(x) = 0$ for some $c_i \in \mathbb{F}(x)$. WLOG, by clearing denominators and common factors, we can assume that c_i are also polynomials (i.e., $c_i \in \mathbb{F}[x]$) with no common factor. By comparing the coefficients of powers of x between 0 and $r-1$, we immediately get that $\sum_i c_i(0) g_i(x) = 0$. Note that all $c_i(0)$ cannot be zero simultaneously since then x would be a common factor for all c_i. Therefore we get a non-trivial $\mathbb{F}$-linear dependency for $g_1, g_2, \ldots, g_m$. □

Lemma 16: Let $\mathbb{K}[x; \sigma, \delta]$ be a skew polynomial ring. For $a \in \mathbb{K}$, define $\phi_a : \mathbb{K} \to \mathbb{K}$ as $\phi_a(y) = \sigma(y)a + \delta(y)$. Then

[11] $\mathbb{F}(x^r)$ is the set of rational functions of the form $f(x^r)$ for $f \in \mathbb{F}(x)$ i.e. rational functions which only have terms whose powers are multiples of r.

1) $\phi_a^i(y) = N_i({}^y a)y$ where ϕ_a^i is ϕ_a composed with itself i times and
2) ϕ_a is a linear map over the subfield $\mathbb{K}_a$.

Proof: (1) This can be proved by induction, it is true for $i = 1$.

$$\begin{aligned} N_{i+1}({}^y a)y &= \sigma(N_i({}^y a))^y ay + \delta(N_i({}^y a))y \\ &= \sigma(N_i({}^y a))(\sigma(y)a + \delta(y)) + \delta(N_i({}^y a))y \\ &= \sigma(N_i({}^y a)y)a + \sigma(N_i({}^y a))\delta(y) + \delta(N_i({}^y a))y \\ &= \sigma(N_i({}^y a)y)a + \delta(N_i({}^y a)y) \\ &= \phi_a(N_i({}^y a)y) = \phi_a(\phi_a^i(y)) = \phi_a^{i+1}(y). \end{aligned}$$

(2) $\mathbb{K}_a$-linearity follows since $\forall c \in \mathbb{K}_a$,

$$\phi_a(yc) = N_1({}^{yc}a)yc = N_1({}^y({}^c a))yc = N_1({}^y a)yc = \phi_a(y)c\ .$$

□

Using Lemma 16, one can linearize the evaluation of skew-polynomials on any conjugacy class. This gives a bijection between evaluation of skew-polynomials on a particular conjugacy class and *linearized polynomials* which found several applications in coding theory and linear-algebraic pseudorandomness [34]–[36]. In fact this is a ring isomorphism and the product operation denoted by $\otimes$ in [34] is equivalent to the product operation for skew polynomials in the appropriate skew polynomial ring.

A. Wronskian Matrix

The theory of skew polynomials allows us to calculate rank of Wronskian matrices. Let $\mathbb{K}[x;\delta]$ be a skew-polynomial of derivation type i.e. $\sigma \equiv \mathrm{Id}$ is the identity map.

Definition 10 (Wronskian): Let $c_1, \ldots, c_n \in \mathbb{K}^*$. Define the Wronskian

$$W_n(c_1, \ldots, c_n) = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ \delta(c_1) & \delta(c_2) & \cdots & \delta(c_n) \\ \delta^2(c_1) & \delta^2(c_2) & \cdots & \delta^2(c_n) \\ \vdots & \vdots & & \vdots \\ \delta^{n-1}(c_1) & \delta^{n-1}(c_2) & \cdots & \delta^{n-1}(c_n) \end{bmatrix}.$$

Corollary 4: $W_n(c_1, \ldots, c_n)$ is full-rank iff $c_1, \ldots, c_n$ are linearly independent over $\mathbb{F} = \mathbb{K}_0$, the centralizer of 0.

Proof: By Lemma 16, $\delta^i(c) = N_i({}^c 0)c$. Thus the claim follows from Lemma 9. □

Note that when δ is the formal derivative of polynomials, the above is the usual Wronskian of polynomials. Applying the above corollary in this special case, we can relate the non-singularity of the Wronskian to the linear independence of the polynomials.

Proposition 5: Let $f_1, f_2, \ldots, f_s \in \mathbb{F}[x]$ be polynomials of degree at most d. Suppose $\delta^j(f_i)$ is the j^{th} derivative of f_i. Define

$$M = \begin{bmatrix} f_1(x) & f_2(x) & \ldots & f_s(x) \\ \vdots & \vdots & & \vdots \\ \delta^j(f_1)(x) & \delta^j(f_2)(x) & \ldots & \delta^j(f_s)(x) \\ \vdots & \vdots & & \vdots \\ \delta^{s-1}(f_1)(x) & \delta^{s-1}(f_2)(x) & \ldots & \delta^{s-1}(f_s)(x) \end{bmatrix}.$$

Then the following are true:

1) If $\mathrm{char}(\mathbb{F}) = p$ then[12], $\det(M) \neq 0$ iff $f_1, f_2, \ldots, f_2$ are linearly independent over $\mathbb{F}(x^p)$.
2) If $\mathrm{char}(\mathbb{F}) > d$ or $\mathrm{char}(\mathbb{F}) = 0$ then, $\det(M) \neq 0$ iff $f_1, f_2, \ldots, f_s$ are linearly independent over $\mathbb{F}$.

Proof: It is clear that if $f_1, f_2, \ldots, f_d$ are linearly dependent over $\mathbb{F}$, then $\det M = 0$. Now we will prove the converse.

Consider the skew polynomial ring defined in Example 2 where $\mathbb{K} = \mathbb{F}(x), \sigma \equiv \mathrm{Id}$ and $\delta(f)$ is the derivative of f. By Corollary 4, $\det(M)$ is zero iff $f_1, f_2, \ldots, f_s$ are linearly independent over $\mathbb{K}_0$, the centralizer of 0. We have

$$\mathbb{K}_0 = \{g : {}^g 0 = 0\} \cup \{0\} = \{g : \delta(g) = 0\}.$$

If $\mathrm{char}(\mathbb{F}) = 0$, then $\mathbb{K}_0 = \mathbb{F}$ and we are done. If $\mathrm{char}(\mathbb{F}) = p$ for some prime p, then we claim below that $\mathbb{K}_0 = \mathbb{F}(x^p)$, which finishes the proof using Lemma 15. □

Claim 9: If $\mathrm{char}(\mathbb{F}) = p$, then $\mathbb{K}_0 = \mathbb{F}(x^p)$.

Proof: $\mathbb{K}_0 = \{g \in \mathbb{F}(x) : \delta(g) = 0\}$. If $g \in \mathbb{F}[x]$, then it is easy to see that $\delta(g) = 0$ iff $g \in \mathbb{F}[x^p]$. Now suppose g is a rational function of the form $g = a/b$ where $a, b \in \mathbb{F}[x]$ do not have any common factors. By product rule, $\delta(g) = 0 \iff \delta(a)b = a\delta(b)$. Since a, b do not have any common factors, this implies that a divides $\delta(a)$ and b divides $\delta(b)$. Since degree of $\delta(a)$ is smaller than a, this is not possible unless $\delta(a) = 0$ and similarly we can conclude that $\delta(b) = 0$. Therefore $a, b \in \mathbb{F}[x^p]$ and so $g \in \mathbb{F}(x^p)$. □

Using the above, we can now deduce the following result which is the basis of list-size bound for list decoding univariate multiplicity codes [33] and the analysis of the associated subspace design constructed in [25].

Proposition 6: Let $\mathrm{char}(\mathbb{F}) = p$. Let δ be the derivative operator on polynomials in $\mathbb{F}[x]$ and $\delta^i(\cdot)$ be the i^{th} derivative of a polynomial. Let $Q(x, y_0, y_1, \ldots, y_{s-1}) = A(x) + \sum_{i=0}^{s-1} A_i(x)y_i$ where $A(x), A_i(x) \in \mathbb{F}[x]$ and not all A_i are zero. The set of all $f \in \mathbb{F}[x]$ of degree less than p, such that

$$Q(x, f(x), \delta(f)(x), \ldots, \delta^{s-1}(f)(x)) = 0, \qquad (22)$$

form an $\mathbb{F}$-affine subspace of $\mathbb{F}[x]$ of dimension at most $s-1$.

Proof: Equation (22) can be rewritten as $A + \sum_{i=0}^{s-1} A_i\delta^i(f) = 0$. Suppose that the set of solutions to this equation in $\mathbb{F}[x]^{<p}$ form an $\mathbb{F}$-affine subspace of $\mathbb{F}[x]$ of dimension at least s. Then there exist solutions $f_0, f_1, \ldots, f_s \in \mathbb{F}[x]^{<p}$ where $f_1 - f_0, \ldots, f_s - f_0$ are $\mathbb{F}$-linearly independent. Let $g_i = f_i - f_0$. Then for $j \in [s]$ we have, $\sum_{i=0}^{s-1} A_i\delta^i(g_j) = 0$. Therefore the determinant of the matrix $[\delta^i(g_j)]_{ij}$ is zero. Therefore by Proposition 5, $g_1, g_2, \ldots, g_s$ should be $\mathbb{F}$-linearly dependent, which is a contradiction. □

We also remark that solving equation (22) when $A = 0$ is equivalent to finding roots of a skew polynomial of degree $s-1$ in a conjugacy class. This also intuitively explains why the set of solutions is an affine subspace of dimension at most $s-1$. Consider the skew polynomial ring $\mathbb{K}[t;\delta]$ of derivation type where $\mathbb{K} = \mathbb{F}(x)$, $\sigma \equiv \mathrm{Id}$ and δ is the derivative operator. Then by Lemma 16, $N_i({}^f 0)f = \delta^i(f)$.

[12] $\mathrm{char}(\mathbb{F})$ is the characteristic of $\mathbb{F}$.

Therefore the Equation (22), when $A = 0$, can be rewritten as:

$$\sum_{i=0}^{s-1} A_i \delta^i(f) = 0 \iff \sum_{i=0}^{s-1} A_i N_i({}^f 0) f = 0.$$

Define $G(t) \in \mathbb{K}[t;\delta]$ as $G(t) = \sum_{i=0}^{s-1} A_i t^i$ which is a skew polynomial of degree at most $s-1$. Then $G({}^f 0)f = \sum_{i=0}^{s-1} A_i N_i({}^f 0) f$. Therefore the solutions of (22) when $A = 0$ are precisely $\{0\} \cup \{f : G({}^f 0) = 0\}$.

B. Moore Matrix

The theory of skew polynomials also allows us to calculate the rank of Moore matrices. Let $\mathbb{K}[t;\sigma]$ be a skew polynomial ring of endomorphism type i.e. $\delta \equiv 0$. This is completely analogous to Wronskian matrices (Section E-A) once we use the skew polynomial framework.

Definition 11 (Moore matrix): Let $c_1, \ldots, c_n \in \mathbb{K}^*$. Define the Moore matrix

$$M_n(c_1, \ldots, c_n) = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ \sigma(c_1) & \sigma(c_2) & \cdots & \sigma(c_n) \\ \sigma^2(c_1) & \sigma^2(c_2) & \cdots & \sigma^2(c_n) \\ \vdots & \vdots & & \vdots \\ \sigma^{n-1}(c_1) & \sigma^{n-1}(c_2) & \cdots & \sigma^{n-1}(c_n) \end{bmatrix}.$$

Corollary 5: $M_n(c_1, \ldots, c_n)$ is full-rank iff $c_1, \ldots, c_n$ are linearly independent over $\mathbb{F} = \mathbb{K}_1$, the centralizer of 1.

Proof: By Lemma 16, $\sigma^i(c) = N_i({}^c 1)c$. Thus the claim follows from Lemma 9. □

We now apply the above to the case when $\mathbb{K} = \mathbb{F}_q(x)$ and σ is the automorphism which maps $f(x) \in \mathbb{F}_q(x)$ to $f(\gamma x)$ for a generator γ of $\mathbb{F}_q^*$. In this case, the Moore matrix was called the folded Wronskian in [25]. Analogous Moore matrices for function fields were studied in [26].

Proposition 7: Let $f_1, f_2, \ldots, f_s \in \mathbb{F}_q[x]$ be polynomials of degree at most d. Let γ be generator for $\mathbb{F}_q^*$. Define

$$M = \begin{bmatrix} f_1(x) & f_2(x) & \ldots & f_s(x) \\ \vdots & \vdots & & \vdots \\ f_1(\gamma^j x) & f_2(\gamma^j x) & \ldots & f_s(\gamma^j x) \\ \vdots & \vdots & & \vdots \\ f_1(\gamma^{s-1} x) & f_2(\gamma^{s-1} x) & \ldots & f_s(\gamma^{s-1} x) \end{bmatrix}.$$

Then the following are true:

1) $\det(M) \neq 0$ iff $f_1, f_2, \ldots, f_2$ are linearly independent over $\mathbb{F}_q(x^{q-1})$.
2) If $q - 1 > d$ then, $\det(M) \neq 0$ iff $f_1, f_2, \ldots, f_s$ are linearly independent over $\mathbb{F}_q$.

Proof: It is clear that if $f_1, f_2, \ldots, f_d$ are linearly dependent over $\mathbb{F}_q$, then $\det M = 0$. Now we will prove the converse.

Consider the skew polynomial ring defined in Example 2 where $\mathbb{K} = \mathbb{F}_q(x), \sigma(g(x)) = g(\gamma x)$ and $\delta \equiv 0$. By Corollary 5, $\det(M)$ is zero iff $f_1, f_2, \ldots, f_s$ are linearly independent over $\mathbb{K}_1$, the centralizer of 1. We have

$$\mathbb{K}_1 = \{g : {}^g 1 = 1\} \cup \{0\} = \{g : g(\gamma x) = g(x)\}.$$

We now claim that $\mathbb{K}_1 = \mathbb{F}_q(x^{q-1})$ and the rest follows from Lemma 15. □

Claim 10: $\mathbb{K}_1 = \mathbb{F}_q(x^{q-1})$.

Proof: $\mathbb{K}_1 = \{g \in \mathbb{F}_q(x) : g(\gamma x) = g(x)\}$. If $g \in \mathbb{F}_q[x]$, then it is easy to see that $g(\gamma x) = g(x)$ iff $g \in \mathbb{F}_q[x^{q-1}]$. Now suppose g is a rational function of the form $g = a/b$ where $a, b \in \mathbb{F}_q[x]$ do not have any common factors and we can assume that the constant term of a or b is 1. $g(\gamma x) = g(x) \iff a(\gamma x)b(x) = a(x)b(\gamma x)$. Since a, b do not have any common factors, this implies that a divides $a(\gamma x)$ and b divides $b(\gamma x)$. Since degree of $a(\gamma x)$ is the same as that of $a(x)$ and the degree of $b(\gamma x)$ is the same as that of $b(x)$, this implies that $a(\gamma x) = \lambda a(x)$ and $b(\gamma x) = \lambda b(x)$ for some $\lambda \in \mathbb{F}_q$. Since we assumed that a or b has constant term 1, we can conclude that $\lambda = 1$. Therefore $a, b \in \mathbb{F}_q[x^{q-1}]$ and so $g \in \mathbb{F}_q(x^{q-1})$. □

Using the above, we can now deduce the following result which is the basis of list-size bound for list decoding folded Reed-Solomon codes [21], [37] and the analysis of the subspace design constructed using folded Reed-Solomon codes [25].

Lemma 17: Let γ be a generator for $\mathbb{F}_q^*$. Let $Q(x, y_0, y_1, \ldots, y_{s-1}) = A(x) + \sum_{i=0}^{s-1} A_i(x) y_i$ where $A(x), A_i(x) \in \mathbb{F}_q[x]$ and not all A_i are zero. The set of all $f \in \mathbb{F}_q[x]$ of degree less than $q-1$, such that

$$Q(x, f(x), f(\gamma x), \ldots, f(\gamma^{s-1} x)) = 0, \tag{23}$$

form an $\mathbb{F}_q$-affine subspace of $\mathbb{F}_q[x]$ of dimension at most $s-1$.

Proof: Equation (23) can be rewritten as $A + \sum_{i=0}^{s-1} A_i f(\gamma^i x) = 0$. Suppose that the set of solutions to this equation in $\mathbb{F}_q[x]^{<q-1}$ form an $\mathbb{F}_q$-affine subspace of $\mathbb{F}_q[x]$ of dimension at least s. Then there exist solutions $f_0, f_1, \ldots, f_s \in \mathbb{F}_q[x]^{<q-1}$ where $f_1 - f_0, \ldots, f_s - f_0$ are $\mathbb{F}_q$-linearly independent. Let $g_i = f_i - f_0$. Then for $j \in [s]$ we have, $\sum_{i=0}^{s-1} A_i g_j(\gamma^i x) = 0$. Therefore the determinant of the matrix $[g_j(\gamma^i x)]_{ij}$ is zero. Therefore by Proposition 7, $g_1, g_2, \ldots, g_s$ should be $\mathbb{F}_q$-linearly dependent, which is a contradiction. □

Just as we did in Section E-A, we remark that solving Equation (23), when $A = 0$, is equivalent to finding roots of the degree $s-1$ skew polynomial $G(t) = \sum_{i=0}^{s-1} A_i t^i$ in the conjugacy class of 1, where the underlying skew polynomial ring is $\mathbb{K}[t;\sigma]$ where $\mathbb{K} = \mathbb{F}(x)$ and $\sigma(f(x)) = f(\gamma x)$.

APPENDIX F
MAXIMUM SUM RANK DISTANCE CODES

In this section, we will present a construction of Maximum Sum-Rank Distance (MSRD) codes due to [27] using the skew polynomial framework. We will first define sum-rank distance codes.

Fix some basis $\mathcal{B}$ for $\mathbb{F}_{q^m}$ as vector space over $\mathbb{F}_q$. Given $z = (z_1, z_2, \ldots, z_r) \in \mathbb{F}_{q^m}^r$, we can think of z as an $m \times r$ matrix with entries in $\mathbb{F}_q$ by expressing each coordinate z_i as a $\mathbb{F}_q^m$ vector using basis $\mathcal{B}$; define $\mathrm{rank}_{\mathbb{F}_q}(z)$ to be the $\mathbb{F}_q$-rank of that matrix. Let $\mathcal{P} = A_1 \sqcup A_2 \sqcup \cdots \sqcup A_s$ be a partition of $[n]$ into s parts. Given $x \in \mathbb{F}_{q^m}^n$, let $x = (x_1, x_2, \ldots, x_s)$

be the partition of x according to $\mathcal{P}$ where $x_i \in \mathbb{F}_{q^m}^{A_i}$. Define $\text{sum-rank}_{\mathcal{P}}(x) = \sum_{i=1}^{s} \text{rank}_{\mathbb{F}_q}(x_i)$.

Definition 12 (sum-rank distance): Fix some partition $\mathcal{P} = A_1 \sqcup A_2 \sqcup \cdots \sqcup A_s$ of $[n]$ into s parts. An $\mathbb{F}_{q^m}$-linear subspace C of $\mathbb{F}_{q^m}^n$ is said to have sum-rank distance d (w.r.t. partition $\mathcal{P}$) if every non-zero codeword $c \in C$, $\text{sum-rank}_{\mathcal{P}}(c) \geqslant d$.

Note that the sum-rank distance generalizes both Hamming metric (by choosing $\mathcal{P} = \{1\} \sqcup \{2\} \sqcup \cdots \sqcup \{n\}$) and rank metric (by choosing $\mathcal{P} = [n]$). Moreover for any partition $\mathcal{P}$ and any $x \in \mathbb{F}_{q^m}^n$, $\text{sum-rank}_{\mathcal{P}}(x)$ is most the Hamming weight of x (as rank is upper bounded by the number of non-zero columns). Therefore by the Singleton bound, any k-dimensional code of $\mathbb{F}_{q^m}^n$, can have sum-rank distance at most $n - k + 1$. A code achieving this bound is called an MSRD code. Therefore MSRD codes generalize both MDS codes and Gabidulin codes. Sum-rank distance was introduced by [38] for applications in network coding. We will now present the construction of MSRD codes.

Theorem 7 (Construction of maximum sum rank distance codes [27]): Let γ be a generator for $\mathbb{F}_{q^m}$ and let $\beta_1, \ldots, \beta_m \in \mathbb{F}_{q^m}$ be linearly independent over $\mathbb{F}_q$. Let $n = (q-1)m$. For $k \leqslant n$, define a $k \times n$ matrix $M = [M_0 | M_1 | \ldots | M_{q-2}]$ where

$$M_\ell = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_m \\ \gamma^\ell \beta_1^q & \gamma^\ell \beta_2^q & \cdots & \gamma^\ell \beta_m^q \\ \gamma^{\ell(1+q)} \beta_1^{q^2} & \gamma^{\ell(1+q)} \beta_2^{q^2} & \cdots & \gamma^{\ell(1+q)} \beta_m^{q^2} \\ \vdots & \vdots & & \vdots \\ \gamma^{\ell(1+q+\cdots+q^{h-2})} \beta_1^{q^{h-1}} & \gamma^{\ell(1+q+\cdots+q^{h-2})} \beta_2^{q^{h-1}} & \cdots & \gamma^{\ell(1+q+\cdots+q^{h-2})} \beta_m^{q^{h-1}} \end{bmatrix}.$$

Then M is the generator matrix of a maximum sum rank distance code, i.e., for every non-zero vector $\lambda \in \mathbb{F}_{q^m}^k$, $\sum_{\ell=0}^{q-2} \text{rank}_{\mathbb{F}_q}(\lambda^T M_\ell) \geqslant n - k + 1$.[13]

Proof: Suppose $\lambda \in \mathbb{F}_{q^m}^k$ is a non-zero vector such that $\sum_{\ell=0}^{q-2} \text{rank}_{\mathbb{F}_q}(\lambda^T M_\ell) \leqslant n - k$. This is equivalent to $\sum_{\ell=0}^{q-2} \dim_{\mathbb{F}_q}(\ker_{\mathbb{F}_q}(\lambda^T M_\ell)) \geqslant k$.

Let $\mathbb{K} = \mathbb{F}_{q^m}$, $\sigma(a) = a^q$ and $\delta \equiv 0$. See Example 1 for the conjugation relation and conjugacy classes in this case. Define $f(t) = \sum_{i=0}^{k-1} \lambda_i t^i$ which is a non-zero skew polynomial of degree at most $k-1$ in $\mathbb{F}_{q^m}[t;\sigma]$. We will find many roots for f which would violate Theorem 3 to get a contradiction.

Fix some $\ell \in \{0, 1, \ldots, q-2\}$. Suppose $\dim_{\mathbb{F}_q}(\ker_{\mathbb{F}_q}(\lambda^T M_\ell)) = d_\ell$. Let $\mu_1, \ldots, \mu_{d_\ell} \in \mathbb{F}_q^m$ be a basis for the kernel. Let $\beta = (\beta_1, \beta_2, \ldots, \beta_m) \in \mathbb{F}_{q^m}^m$. Now $\lambda^T M_\ell \mu_i = 0$ implies that $\beta^T \mu_i$ is root of f. Moreover the d_ℓ roots $\beta^T \mu_1, \ldots, \beta^T \mu_{d_\ell} \in \mathbb{F}_{q^m}$ are linearly independent over $\mathbb{F}_q$ since $\text{rank}_{\mathbb{F}_q}(\beta) = m$.

[13]Here we are interpreting a row vector $c \in \mathbb{F}_{q^m}^r$ as an $m \times r$ matrix over $\mathbb{F}_q$. $\text{rank}_{\mathbb{F}_q}(c)$ is the $\mathbb{F}_q$-rank of this matrix. We will also use $\ker_{\mathbb{F}_q}(c)$ in the proof to denote the kernel of the matrix.

Thus we get $\sum_{\ell=0}^{q-2} d_\ell \geqslant k$ roots for f. And the roots in each conjugacy class are linearly independent over $\mathbb{F}_q$ (which is the centralizer). Therefore by Theorem 3, we get a contradiction. □

It is easy to see that the above construction can be easily modified to work for any partition $\mathcal{P}$ of $[n]$ into at most $(q-1)$ parts, where each part has size at most m. In [39], an efficient decoding algorithm for these codes is given.

ACKNOWLEDGMENT

The authors thank Sergey Yekhanin for several illuminating discussions about MR-LRCs and Umberto Martínez-Peñas for helpful comments on an earlier version of this paper.

REFERENCES

[1] S. Gopi, V. Guruswami, and S. Yekhanin, "Maximally recoverable LRCs: A field size lower bound and constructions for few heavy parities," *IEEE Trans. Inf. Theory*, vol. 66, no. 10, pp. 6066–6083, Oct. 2020.

[2] C. Huang, M. Chen, and J. Li, "Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems," in *Proc. 6th IEEE Int. Symp. Netw. Comput. Appl. (NCA)*, Jul. 2007, pp. 79–86.

[3] M. Chen, C. Huang, and J. Li, "On the maximally recoverable property for multi-protection group codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2007, pp. 486–490.

[4] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6925–6934, Nov. 2011.

[5] D. S. Papailiopoulos and A. G. Dimakis, "Locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5843–5855, Oct. 2014.

[6] C. Huang *et al.*, "Erasure coding in Windows azure storage," in *Proc. USENIX Annu. Tech. Conf. (ATC)*, 2012, pp. 15–26.

[7] M. Sathiamoorthy *et al.*, "XORing elephants: Novel erasure codes for big data," in *Proc. VLDB Endowment (PVLDB)*, 2013, pp. 325–336.

[8] P. Gopalan, C. Huang, B. Jenkins, and S. Yekhanin, "Explicit maximally recoverable codes with locality," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5245–5256, Sep. 2014.

[9] M. Blaum, J. L. Hafner, and S. Hetzler, "Partial-MDS codes and their application to RAID type of architectures," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4510–4519, Jul. 2013.

[10] M. Blaum, "Construction of PMDS and SD codes extending RAID 5," 2013, *arXiv:1305.0032*.

[11] R. Gabrys, E. Yaakobi, M. Blaum, and P. Siegel, "Construction of partial MDS codes over small finite fields," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Dec. 2017, pp. 1–5.

[12] V. Guruswami, L. Jin, and C. Xing, "Constructions of maximally recoverable local reconstruction codes via function fields," *IEEE Trans. Inf. Theory*, vol. 66, no. 10, pp. 6133–6143, Oct. 2020.

[13] U. Martínez-Peñas and F. R. Kschischang, "Universal and dynamic locally repairable codes with maximal recoverability via sum-rank codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 7790–7805, Dec. 2019.

[14] I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis, "Optimal locally repairable codes and connections to Matroid theory," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6661–6671, Dec. 2016.

[15] G. Hu and S. Yekhanin, "New constructions of SD and MR codes over small finite fields," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1591–1595.

[16] P. Gopalan, G. Hu, S. Kopparty, S. Saraf, C. Wang, and S. Yekhanin, "Maximally recoverable codes for grid-like topologies," in *Proc. 28th Annu. ACM-SIAM Symp. Discrete Algorithms*, Jan. 2017, pp. 2092–2108.

[17] G. Calis and O. O. Koyluoglu, "A general construction for PMDS codes," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 452–455, Mar. 2017.

[18] M. Blaum, J. S. Plank, M. Schwartz, and E. Yaakobi, "Construction of partial MDS and sector-disk codes with two global parity symbols," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2673–2681, May 2016.

[19] H. Cai, Y. Miao, M. Schwartz, and X. Tang, "A construction of maximally recoverable codes with order-optimal field size," *IEEE Trans. Inf. Theory*, vol. 68, no. 1, pp. 204–212, Jan. 2022.

[20] U. Martínez-Peñas, "A general family of MSRD codes and PMDS codes with smaller field sizes from extended Moore matrices," 2020, *arXiv:2011.14109*.
[21] V. Guruswami and C. Wang, "Linear-algebraic list decoding for variants of Reed–Solomon codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3257–3268, Jun. 2013.
[22] M. A. Forbes and A. Shpilka, "On identity testing of tensors, low-rank recovery and compressed sensing," in *Proc. 44th Symp. Theory Comput. (STOC)*, 2012, pp. 163–172, doi: 10.1145/2213977.2213995.
[23] M. A. Forbes, R. Saptharishi, and A. Shpilka, "Hitting sets for multilinear read-once algebraic branching programs, in any order," in *Proc. 46th Annu. ACM Symp. Theory Comput.*, May 2014, pp. 867–875.
[24] M. A. Forbes and V. Guruswami, "Dimension expanders via rank condensers," in *Proc. 19th Int. Workshop Randomization Comput. (RANDOM)*, 2015, pp. 800–814.
[25] V. Guruswami and S. Kopparty, "Explicit subspace designs," *Combinatorica*, vol. 36, no. 2, pp. 161–185, 2016, doi: 10.1007/s00493-014-3169-1.
[26] V. Guruswami, C. Xing, and C. Yuan, "Constructions of subspace designs via algebraic function fields," *Trans. Amer. Math. Soc.*, vol. 370, pp. 8757–8775, Jan. 2018.
[27] U. Martínez-Peñas, "Skew and linearized Reed–Solomon codes and maximum sum rank distance codes over any division ring," *J. Algebr.*, vol. 504, pp. 587–612, Jun. 2018.
[28] D. Boucher and F. Ulmer, "Linear codes using skew polynomials with automorphisms and derivations," *Des., Codes Cryptogr.*, vol. 70, no. 3, pp. 405–431, 2014.
[29] T. Y. Lam and A. Leroy, "Vandermonde and wronskian matrices over division rings," *J. Algebra*, vol. 119, pp. 308–336, Dec. 1988.
[30] T.-Y. Lam, *A General Theory of Vandermonde Matrices*. Berkeley, CA, USA: Center for Pure and Applied Mathematics, Univ. California, Berkeley, 1985.
[31] O. Ore, "Theory of non-commutative polynomials," *Ann. Math.*, vol. 34, no. 3, pp. 480–508, 1993. [Online]. Available: http://www.jstor.org/stable/1968173
[32] A. Bostan, B. Salvy, M. F. I. Chowdhury, É. Schost, and R. Lebreton, "Power series solutions of singular (q)-differential equations," in *Proc. 37th Int. Symp. Symbolic Algebr. Comput. (ISSAC)*, 2012, pp. 107–114.
[33] V. Guruswami and C. Wang, "Optimal rate list decoding via derivative codes," in *Proc. APPROX/RANDOM*, Aug. 2011, pp. 593–604.
[34] H. Mahdavifar and A. Vardy, "Algebraic list-decoding of subspace codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7814–7828, Dec. 2013.
[35] V. Guruswami, N. Resch, and C. Xing, "Lossless dimension expanders via linearized polynomials and subspace designs," in *Proc. 33rd Comput. Complex. Conf. (CCC)*, 2018, pp. 1–18.
[36] E. R. Berlekamp, *Algebraic Coding Theory (Revised Edition)*. Singapore: World Scientific, 2015.
[37] V. Guruswami, "Linear-algebraic list decoding of folded Reed–Solomon codes," in *Proc. IEEE 26th Annu. Conf. Comput. Complex.*, Jun. 2011, pp. 77–85.
[38] R. W. Nobrega and B. F. Uchoa-Filho, "Multishot codes for network coding using rank-metric codes," in *Proc. 3rd IEEE Int. Workshop Wireless Netw. Coding*, Jun. 2010, pp. 1–6.
[39] U. Martinez-Penas and F. R. Kschischang, "Reliable and secure multishot network coding using linearized Reed–Solomon codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 4785–4803, Aug. 2019.

Sivakanth Gopi received the bachelor's degree from the Indian Institute of Technology Bombay in 2013 and the Ph.D. degree in computer science from Princeton University in 2018. The title of his dissertation was "Locality in coding theory."

He is currently a Senior Researcher at Microsoft Research Redmond. His main research interests are in coding theory and its applications to both theory and practice, specifically to areas like pseudorandomness, complexity theory, cryptography, privacy, and distributed data storage. He is also interested in differential privacy and its applications to private machine learning.

Venkatesan Guruswami (Fellow, IEEE) received the bachelor's degree from the Indian Institute of Technology, Madras, in 1997, and the Ph.D. degree from MIT in 2001.

He is currently a Professor of computer science and mathematics at UC Berkeley and a Senior Scientist at the Simons Institute for the Theory of Computing. His research interests include coding theory, pseudorandomness, approximate optimization, and computational complexity. He is a fellow of the ACM. He was a recipient of the Simons Investigator Award, the Presburger Award, Packard and Sloan Fellowships, the ACM Doctoral Dissertation Award, and the IEEE Information Theory Society Paper Award. He is currently the Editor-in-Chief of the *Journal of the ACM*.