Chapter 2
Locally decodable codes via the point removal method

This chapter contains a detailed exposition of the point removal method for constructing locally decodable codes. The method can be broken into two parts. The first part is a reduction that shows how the existence of subsets of finite fields that simultaneously exhibit "nice" properties of two different kinds yields families of locally decodable codes with good parameters. The second part is a construction of "nice" subsets of finite fields.

Sections 2.1 and 2.2 of this chapter are preliminary. In Section 2.3, we give a detailed treatment of the first part of our method for the narrow case of binary codes. We treat binary codes separately to have a simpler setup where we can (in an intuitive yet formal manner) demonstrate the combinatorial and geometric ideas that lie behind our method. While we believe that Section 2.3 may be the most important part of the book (since it explains the intuition behind our approach), it can be skipped by the reader who is interested only in a succinct formal treatment of the constructions. After a detailed treatment of binary codes in Section 2.3, we give a succinct treatment of general codes in Section 2.4. As our main conclusion, we identify the two "nice" properties of subsets of finite fields that (simultaneously) yield good codes. We call those properties combinatorial and algebraic niceness.

The next two sections cover the second part of our method. In Section 2.5, we construct combinatorially nice subsets of prime fields, and in Section 2.6 we construct algebraically nice subsets of prime fields. Finally, in Section 2.7, we put the results of the previous sections together and summarize our improvements in upper bounds for locally decodable codes.

2.1 Notation

We use the following standard mathematical notation:

- $[s] = \{1, \ldots, s\}$.
- $\mathbb{Z}_n$ denotes integers modulo n.

S. Yekhanin, *Locally Decodable Codes and Private Information Retrieval Schemes*,
Information Security and Cryptography, DOI 10.1007/978-3-642-14358-8_2,

- $\mathbb{F}_q$ is a finite field of q elements.
- $\mathbb{F}_q^*$ is the multiplicative group of $\mathbb{F}_q$.
- $d_H(\mathbf{x},\mathbf{y})$ denotes the Hamming distance between vectors $\mathbf{x}$ and $\mathbf{y}$.
- $(\mathbf{u},\mathbf{v})$ stands for the dot product of vectors $\mathbf{u}$ and $\mathbf{v}$.
- For a linear space $L \subseteq \mathbb{F}_r^m$, $L^\perp$ denotes the *dual* space. That is,

$$L^\perp = \{\mathbf{u} \in \mathbb{F}_r^m \mid \forall \mathbf{v} \in L, (\mathbf{u},\mathbf{v}) = 0\}.$$

2.2 Locally decodable codes

In this section we formally define locally decodable codes.

Definition 1. An r-ary code $C : [r]^n \to [r]^N$ is said to be (k,δ,ε)-locally decodable if there exists a randomized decoding algorithm $\mathscr{A}$ such that:

1. For all $\mathbf{x} \in [r]^n$, $i \in [n]$ and $\mathbf{y} \in [r]^N$ such that $d_H(C(\mathbf{x}),\mathbf{y}) \le \delta N$,

$$\Pr\left[\mathscr{A}^{\mathbf{y}}(i) = \mathbf{x}_i\right] \ge 1-\varepsilon,$$

where the probability is taken over the random coin tosses of the algorithm $\mathscr{A}$.

2. $\mathscr{A}$ makes at most k queries to $\mathbf{y}$.

In the special case when r is a prime power and the elements of the alphabet $[r]$ are in one-to-one correspondence with the elements of the finite field $\mathbb{F}_r$, it makes sense to talk about *linear* codes. A locally decodable code C is called linear if C is a linear transformation over $\mathbb{F}_r$. In this book we consider only codes over prime alphabets, and all our codes are linear.

2.3 Binary LDCs via point removal

In this section, we give a detailed treatment of the first part of our method for the narrow case of binary codes. Our goal here is to explain the intuition behind the point removal approach; therefore, we gradually build up our main construction, trying to provide motivation for every choice that we make. Our final result is a claim that subsets of prime fields that exhibit certain properties (combinatorial and algebraic niceness) yield families of LDCs with very good parameters.

In Section 2.3.1, we introduce certain combinatorial objects that we call regular intersecting families of sets. These objects later serve as our tool to construct binary LDCs. In Section 2.3.2, we present a linear algebraic construction of a regular intersecting family that yields locally decodable codes with good (although not the best known) parameters. The notions of combinatorial and algebraic niceness of sets are used implicitly in this section. Our main construction in Section 2.3.3 builds upon the construction of Section 2.3.2 via the *point removal* procedure. We formally introduce combinatorial and algebraic niceness and show how the interplay between these two notions yields locally decodable codes.

2.3.1 Regular intersecting families of sets

The locally decodable codes that we construct are linear. Our decoding algorithms proceed by tossing random coins, reading a certain k-tuple of coordinates of the (corrupted) codeword, and outputting the XOR of the values at these coordinates.

Observe that every linear LDC encoding n-bit messages to N-bit codewords admits a combinatorial description. Let $N, R,$ and n be arbitrary positive integers. For $i \in [n]$, let $\mathbf{e}^i$ denote a binary n-dimensional (unit) vector, whose unique nonzero coordinate is i. In order to define a k-query linear locally decodable code, it is sufficient to specify the following for every $i \in [n]$:

- A set $T_i \subseteq [N]$ of coordinates of $C(\mathbf{e}^i)$ that are set to 1. Such sets completely specify the encoding, since for any message $\mathbf{x}$, $C(\mathbf{x}) = \sum_{i:\mathbf{x}_i=1} C(\mathbf{e}^i)$.
- A family $\{Q_{ir}\}, r \in [R]$, of subsets of $[N]$ of size k that specify collections of codeword coordinates that can be read by a decoding algorithm in order to reconstruct the i-th message bit.

Clearly, not every collection of sets $\{T_i\}$ and $\{Q_{ir}\}$ yields a locally decodable code. Certain combinatorial constraints must be satisfied. We formally define these constraints below.

Definition 2. We say that the subsets $\{T_i\}$ and $\{Q_{ir}\}$ form a (k,n,N,R,s)-regular intersecting family if the following conditions are satisfied:

1. k is odd.
2. For all $i \in [n]$, $|T_i| = s$.
3. For all $i \in [n]$ and $r \in [R]$, $|Q_{ir}| = k$.
4. For all $i \in [n]$ and $r \in [R]$, $Q_{ir} \subseteq T_i$.
5. For all $i \in [n]$ and $w \in T_i$, $|\{r \in [R] \mid w \in Q_{ir}\}| = (Rk)/s$, (i.e., T_i is uniformly covered by the sets Q_{ir}).
6. For all $i, j \in [n]$ and $r \in [R]$ such that $i \neq j$, $\left|Q_{ir} \cap T_j\right| \equiv 0 \bmod (2)$.

We now formally show how regular intersecting families yield binary locally decodable codes.

Proposition 1. *A (k,n,N,R,s)-regular intersecting family yields a binary linear code encoding n bits to N bits that is $(k, \delta, \delta Nk/s)$-locally decodable for all δ.*

Proof. For a set $S \subseteq [N]$, let $I(S) \in \{0,1\}^N$ denote its *incidence vector*. Formally, for $w \in [N]$, we set $I(S)_w = 1$ if $w \in S$, and $I(S)_w = 0$ otherwise. We define a linear code C via its generator matrix $G \in \{0,1\}^{n \times N}$. For $i \in [n]$, we set the i-th row of G to be the incidence vector of the set T_i. Below is the description of the decoding algorithm $\mathscr{A}$. Given oracle access to $\mathbf{y}$ and input $i \in [n]$, the algorithm $\mathscr{A}$ does the following.

1. It picks $r \in [R]$ uniformly at random.
2. It outputs the dot product $(\mathbf{y}, I(Q_{ir}))$ over $\mathbb{F}_2$.

Note that since $|Q_{ir}| = k$, $\mathscr{A}$ needs only k queries to $\mathbf{y}$ to compute the dot product. It is easy to verify that the decoding is correct if $\mathscr{A}$ picks $r \in [R]$ such that all bits of $\mathbf{x}G$ in locations $h \in Q_{ir}$ are not corrupted:

$$(\mathbf{x}G, I(Q_{ir})) = \sum_{j=1}^{n} \mathbf{x}_j \left(I(T_j), I(Q_{ir})\right) = \mathbf{x}_i \left(I(T_i), I(Q_{ir})\right) = \mathbf{x}_i. \tag{2.1}$$

The second equality in (2.1) follows from part 6 of Definition 2 and the last equality follows from parts 1, 3 and 4 of Definition 2.

Now assume that up to δN bits of the encoding $\mathbf{x}G$ have been corrupted. Part 5 of Definition 2 implies that there are at most $(\delta NRk)/s$ sets Q_{ir} that contain at least one corrupted location. Thus, with probability at least $1 - (\delta Nk)/s$, the algorithm $\mathscr{A}$ outputs the correct value. □

To the best of our knowledge, regular intersecting families of sets have not been studied previously. The closest combinatorial objects that have some literature are Ruzsa–Szemeredi (hyper)graphs [40, 79, 80].

2.3.2 Basic construction

In this section we present our basic construction of regular intersecting families, which yields binary k-query locally decodable codes of length $\exp\left(n^{1/(k-1)}\right)$ for prime values of $k \geq 3$. Note that for $k > 3$, the parameters that we get are inferior to the parameters of LDCs of the second generation (see Section 1.3.2).

There is a strong geometric intuition underlying our construction. We choose our universe $[N]$ to be a high-dimensional linear space over a prime field $\mathbb{F}_p$. We choose the sets $\{T_i\}$ to be unions of cosets of certain hyperplanes, and the sets $\{Q_{ir}\}$ to be affine lines. We argue the intersection properties based on elementary linear algebra. Let p be an odd prime, and let $m \geq p - 1$ be an integer.

Lemma 1. *Let $n = \binom{m}{p-1}$. There exist two families of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ in $\mathbb{F}_p^m$ such that*

- *For all $i \in [n]$, $(\mathbf{u}_i, \mathbf{v}_i) = 0$.*
- *For all $i, j \in [n]$ such that $i \neq j$, $(\mathbf{u}_j, \mathbf{v}_i) \neq 0$.*

Proof. Let $\mathbf{e} \in \mathbb{F}_p^m$ be the vector that contains 1's in all the coordinates. We set the vectors $\{\mathbf{u}_i\}$ to be the incidence vectors of all possible $\binom{m}{p-1}$ subsets of $[m]$ of cardinality $(p-1)$. For every $i \in [n]$, we set $\mathbf{v}_i = \mathbf{e} - \mathbf{u}_i$. It is straightforward to verify that this family satisfies the condition of the lemma. □

Now we are ready to present our regular intersecting family. We set $N = p^m$ and $n = \binom{m}{p-1}$. We assume some bijection between the set $[N]$ and the space $\mathbb{F}_p^m$. For $i \in [n]$, we set

$$T_i = \left\{ \mathbf{w} \in \mathbb{F}_p^m \mid (\mathbf{u}_i, \mathbf{w}) \in \mathbb{F}_p^* \right\}.$$

We set

$$R = s = (p-1) \cdot p^{m-1}.$$

For each $i \in [n]$, we assume some bijection between points of T_i and elements of $[R]$. For $i \in [n]$ and $r \in [R]$, let $\mathbf{w}_{ir}$ be the r-th point of T_i. We set

$$Q_{ir} = \left\{ \mathbf{w}_{ir} + \lambda \mathbf{v}_i \mid \lambda \in \mathbb{F}_p \right\}.$$[1]

Lemma 2. *For $i \in [n]$ and $r \in [R]$, the sets $\{T_i\}$ and $\{Q_{ir}\}$ defined above form a (p,n,N,R,s)-regular intersecting family.*

Proof. We simply need to verify that all six conditions listed in Definition 2 are satisfied.

1. Condition 1 is trivial.
2. Condition 2 is trivial.
3. Condition 3 is trivial.
4. Fix $i \in [n]$ and $r \in [R]$. Given that $(\mathbf{u}_i, \mathbf{w}_{ir}) \in \mathbb{F}_p^*$ let us show that $Q_{ir} \subseteq T_i$. By Lemma 1, $(\mathbf{u}_i, \mathbf{v}_i) = 0$. Thus, for every $\lambda \in \mathbb{F}_p$,

   $$(\mathbf{u}_i, \mathbf{w}_{ir} + \lambda \mathbf{v}_i) = (\mathbf{u}_i, \mathbf{w}_{ir}).$$

   Condition 4 follows.
5. Fix $i \in [n]$ and $\mathbf{w} \in T_i$. Note that

   $$\begin{aligned} |\{r \in [R] \mid \mathbf{w} \in Q_{ir}\}| &= \left| \left\{ \mathbf{w}_{ir} \in T_i \mid \exists \lambda \in \mathbb{F}_p, \mathbf{w} = \mathbf{w}_{ir} + \lambda \mathbf{v}_i \right\} \right| \\ &= \left| \left\{ \mathbf{w}_{ir} \in T_i \mid \exists \lambda \in \mathbb{F}_p, \mathbf{w}_{ir} = \mathbf{w} - \lambda \mathbf{v}_i \right\} \right| = p. \end{aligned}$$

   It remains to note that $Rp/s = p$. Condition 5 follows.
6. Fix $i, j \in [n]$, and $r \in [R]$ such that $i \neq j$. Note that

   $$\begin{aligned} \left| Q_{ir} \cap T_j \right| &= \left| \left\{ \lambda \in \mathbb{F}_p \mid (\mathbf{u}_j, \mathbf{w}_{ir} + \lambda \mathbf{v}_i) \in \mathbb{F}_p^* \right\} \right| \\ &= \left| \left\{ \lambda \in \mathbb{F}_p \mid ((\mathbf{u}_j, \mathbf{w}_{ir}) + \lambda (\mathbf{u}_j, \mathbf{v}_i)) \in \mathbb{F}_p^* \right\} \right| = p - 1. \end{aligned}$$

   The last equality follows from the fact that $(\mathbf{u}_j, \mathbf{v}_i) \neq 0$, and therefore the univariate linear function $(\mathbf{u}_j, \mathbf{w}_{ir}) + \lambda (\mathbf{u}_j, \mathbf{v}_i)$ takes every value in $\mathbb{F}_p$ exactly once. It remains to note that $p - 1$ is even. Condition 6 follows.

This completes the proof. □

Combining Lemma 2 and Proposition 1 we get the following corollary.

[1] Note that the sets Q_{ir} are not all distinct.

Corollary 1. *Let p be an odd prime and let $m \geq p-1$ be an integer. There exists a binary linear code encoding $\binom{m}{p-1}$ bits to p^m bits that is $\left(p, \delta, \delta p^2/(p-1)\right)$-locally decodable for all δ.*

It is now easy to convert the above result into a *dense family* (i.e., one that has a code for every message length n, as opposed to infinitely many n's) of p-query LDCs of length $\exp\left(n^{1/(p-1)}\right)$.

Theorem 1. *Let p be a fixed odd prime. For every positive integer n there exists a code of length* $\exp\left(n^{1/(p-1)}\right)$ *that is $\left(p, \delta, \delta p^2/(p-1)\right)$-locally decodable for all δ.*

Proof. Given n, choose m to be the smallest integer such that $n \leq \binom{m}{p-1}$. Set $n' = \binom{m}{p-1}$. It is easy to verify that if n is sufficiently large, we have $n' \leq 2n$. Given a message $\mathbf{x}$ of length n, we pad it with zeros to a length n' and use the code in Corollary 1 that encodes $\mathbf{x}$ with a codeword of length $p^m = \exp\left(n^{1/(p-1)}\right)$. □

2.3.3 The main construction: point removal

In the previous section, we presented our basic linear algebraic construction of regular intersecting families of sets. We chose the sets $\{T_i\}$ to be unions of cosets of certain hyperplanes. We chose the sets $\{Q_{ir}\}$ to be affine lines.

The high-level idea behind our main construction is to reduce the number of codeword locations queried by *removing some points from lines*; i.e., by choosing the sets $\{Q_{ir}\}$ to be *proper subsets of lines* rather than whole lines while preserving the right intersection properties.

Before we proceed to our main construction, we introduce two central technical concepts of our method, namely *combinatorial* and *algebraic niceness* of sets. We now give some narrow definitions that are needed to construct binary codes via the point removal method in linear spaces over prime fields. Later, in Section 2.4 we shall give more general definitions. Let p be an odd prime.

Definition 3. A set $S \subseteq \mathbb{F}_p^*$ is called (m,n)-combinatorially nice if there exist two families of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ in $\mathbb{F}_p^m$ such that:

- For all $i \in [n]$, $(\mathbf{u}_i, \mathbf{v}_i) = 0$.
- For all $i, j \in [n]$ such that $i \neq j$, $(\mathbf{u}_j, \mathbf{v}_i) \in S$.

Remark 1. Note that in Lemma 1 we established that the set $S = \mathbb{F}_p^*$ is $\left(m, \binom{m}{p-1}\right)$-combinatorially nice for every integer $m \geq p-1$.

Definition 4. A set $S \subseteq \mathbb{F}_p^*$ is called k-algebraically nice if k is odd and there exist two sets $S_0, S_1 \subseteq \mathbb{F}_p$ such that:

- S_0 is not empty.

- $|S_1| = k$.
- For all $\alpha \in \mathbb{F}_p$ and $\beta \in S$, $|S_0 \cap (\alpha + \beta S_1)| \equiv 0 \bmod (2)$.

Remark 2. It is easy to verify that the set $S = \mathbb{F}_p^*$ is p-algebraically nice. We simply pick $S_1 = \mathbb{F}_p$ and $S_0 = \mathbb{F}_p^*$.

The next lemma shows how an interplay between combinatorial and algebraic niceness yields regular intersecting families. It is the core of our construction.

Lemma 3. *Assume that $S \subseteq \mathbb{F}_p^*$ is simultaneously (m,n)-combinatorially nice and k-algebraically nice. Let S_0 be the set in the definition of the algebraic niceness of S. The set S yields a $\left(k, n, p^m, |S_0|p^{m-1}, |S_0|p^{m-1}\right)$-regular intersecting family.*

Proof. For $i \in [n]$, let $\mathbf{u}_i, \mathbf{v}_i$ be the vectors in the definition of combinatorial niceness. Set $N = p^m$ and

$$R = s = |S_0|p^{m-1}.$$

Assume a bijection between $[N]$ and $\mathbb{F}_p^m$. For all $i \in [n]$, set

$$T_i = \left\{\mathbf{w} \in \mathbb{F}_p^m \mid (\mathbf{u}_i, \mathbf{w}) \in S_0\right\}.$$

For each $i \in [n]$, assume some bijection between $[R]$ and T_i. Let $\mathbf{w}_{ir}$ denote the r-th point of T_i. Set

$$Q_{ir} = \{\mathbf{w}_{ir} + \lambda \mathbf{v}_i \mid \lambda \in S_1\}.$$

It remains to verify that all six conditions listed in Definition 2 are satisfied.

1. Condition 1 is trivial.
2. Condition 2 is trivial.
3. Condition 3 is trivial.
4. Fix $i \in [n]$ and $r \in [R]$. Given that $(\mathbf{u}_i, \mathbf{w}_{ir}) \in S_0$, let us show that $Q_{ir} \subseteq T_i$. Definition 3 implies that $(\mathbf{u}_i, \mathbf{v}_i) = 0$. Thus, for every $\lambda \in S_1$,

$$(\mathbf{u}_i, \mathbf{w}_{ir} + \lambda \mathbf{v}_i) = (\mathbf{u}_i, \mathbf{w}_{ir}).$$

   Condition 4 follows.
5. Fix $i \in [n]$ and $\mathbf{w} \in T_i$. Note that

$$\begin{aligned} |\{r \in [R] \mid \mathbf{w} \in Q_{ir}\}| &= |\{\mathbf{w}_{ir} \in T_i \mid \exists \lambda \in S_1, \mathbf{w} = \mathbf{w}_{ir} + \lambda v_i\}| \\ &= |\{\mathbf{w}_{ir} \in T_i \mid \exists \lambda \in S_1, \mathbf{w}_{ir} = \mathbf{w} - \lambda \mathbf{v}_i\}| = |S_1| = k. \end{aligned}$$

   It remains to note that $Rk/s = k$. Condition 5 follows.
6. Fix $i, j \in [n]$ and $r \in [R]$ such that $i \neq j$. Note that

$$\begin{aligned} \left|Q_{ir} \cap T_j\right| &= \left|\left\{\lambda \in S_1 \mid (\mathbf{u}_j, \mathbf{w}_{ir} + \lambda \mathbf{v}_i) \in S_0\right\}\right| \\ &= \left|\left\{\lambda \in S_1 \mid ((\mathbf{u}_j, \mathbf{w}_{ir}) + \lambda (\mathbf{u}_j, \mathbf{v}_i)) \in S_0\right\}\right| \\ &= \left|S_0 \cap ((\mathbf{u}_j, \mathbf{w}_{ir}) + (\mathbf{u}_j, \mathbf{v}_i) S_1)\right| \equiv 0 \bmod (2). \end{aligned}$$

The last equality follows from the fact that $(\mathbf{u}_j, \mathbf{v}_i) \in S$, and Definition 4. Condition 6 follows.

This completes the proof. □

Observe that one can derive a regular intersecting family with the parameters of Lemma 2 using Lemma 3 in combination with Remarks 1 and 2.

The next proposition, which follows immediately by combining Proposition 1 with Lemma 3 is the heart of the first part of our construction of LDCs (for the case of binary codes).

Proposition 2. *Let p be an odd prime. Assume that $S \subseteq \mathbb{F}_p^*$ is simultaneously (m,n)-combinatorially nice and k-algebraically nice. Let S_0 be the set in the definition of the algebraic niceness of S. The set S yields a binary linear code encoding n bits to p^m bits that is $(k, \delta, \delta pk/|S_0|)$-locally decodable for all δ.*

Later, we will see that for every Mersenne prime $p = 2^t - 1$, the multiplicative subgroup generated by the element 2 in $\mathbb{F}_p^*$ is three-algebraically nice (Lemma 14) and sufficiently combinatorially nice (Lemma 6) to yield three-query LDCs of length $\exp\left(n^{1/t}\right)$ via the proposition above.

2.4 General LDCs via point removal

In this section, we present a general treatment of the first part of our construction of locally decodable codes. We extend the results of the previous section in two ways: (1) we consider codes over alphabets $\mathbb{F}_r$, for arbitrary primes r, rather than only binary codes; (2) we consider nice subsets of arbitrary finite fields $\mathbb{F}_q$, rather than only prime fields. We start by defining the combinatorial and algebraic niceness of subsets in the general setup, and then proceed to a succinct formal proof of the main propositions.

Definition 5. Let q be a prime power. A set $S \subseteq \mathbb{F}_q^*$ is called (m,n)-combinatorially nice if there exist two families of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ in $\mathbb{F}_q^m$ such that:

- For all $i \in [n]$, $(\mathbf{u}_i, \mathbf{v}_i) = 0$.
- For all $i, j \in [n]$ such that $i \neq j$, $(\mathbf{u}_j, \mathbf{v}_i) \in S$.

In many cases, it will be more convenient for us to use the following definition of combinatorial niceness that involves a single parameter t.

Definition 6. Let q be a prime power. A set $S \subseteq \mathbb{F}_q^*$ is called t-combinatorially nice if for some $c > 0$ and every positive integer m, S is $(m, \lfloor cm^t \rfloor)$-combinatorially nice.

Given a map f from a finite set to a field let $\mathrm{supp}(f)$ denote its *support* i.e., the number of elements of the set that are not mapped to zero. Now we proceed to the general definition of algebraic niceness.

Definition 7. Let q be a prime power and r be a prime. A set $S \subseteq \mathbb{F}_q^*$ is called k-algebraically nice over $\mathbb{F}_r$ if there exist two maps, $S_0 : \mathbb{F}_q \to \mathbb{F}_r$ and $S_1 : \mathbb{F}_q \to \mathbb{F}_r$ such that:

- $\mathrm{supp}(S_0) \neq 0$.
- $\mathrm{supp}(S_1) \leq k$.
- $\sum\limits_{\lambda \in \mathbb{F}_q} S_1(\lambda) \neq 0$.
- For all $\alpha \in \mathbb{F}_q$ and $\beta \in S$, $\sum\limits_{\lambda \in \mathbb{F}_q} S_0(\alpha + \beta\lambda) S_1(\lambda) = 0$.

We now proceed to our core lemma, which shows how sets exhibiting both combinatorial and algebraic niceness yield locally decodable codes.

Lemma 4. *Let q be a prime power and let r be a prime. Assume that $S \subseteq \mathbb{F}_q^*$ is simultaneously (m,n)-combinatorially nice, and k-algebraically nice over $\mathbb{F}_r$. Let S_0 be the map in the definition of the algebraic niceness of S. The set S yields an $\mathbb{F}_r$-linear code encoding messages of length n to codewords of length q^m that is $(k, \delta, \delta q k/\mathrm{supp}(S_0))$-locally decodable for all δ.*

Proof. Our proof has three steps. We specify encoding and local decoding procedures for our codes and then argue a lower bound for the probability of correct decoding. We use notation from Definitions 5 and 7.

Encoding. Our code will be linear. Therefore it suffices to specify the encoding of *unit vectors* $\mathbf{e}_1, \ldots, \mathbf{e}_n$, where $\mathbf{e}_j$ has length n and a unique nonzero coordinate j. We define the encoding of $\mathbf{e}_j$ to be a vector of length q^m, whose coordinates are labeled by elements of $\mathbb{F}_q^m$. For all $\mathbf{w} \in \mathbb{F}_q^m$, we set

$$\mathrm{Enc}(\mathbf{e}_j)_{\mathbf{w}} = S_0\left((\mathbf{u}_j, \mathbf{w})\right). \tag{2.2}$$

Local decoding. Suppose that the decoding algorithm $\mathscr{A}$ needs to recover the i-th coordinate of the message, $i \in [n]$. To simplify the notation, we put

$$c = \frac{1}{\left(S_0\left((\mathbf{u}_i, \mathbf{w})\right) \sum\limits_{\lambda \in \mathbb{F}_q} S_1(\lambda)\right)}.$$

Given a (possibly corrupted) codeword $\mathbf{y}$, $\mathscr{A}$ picks $\mathbf{w} \in \mathbb{F}_q^m$ such that $S_0((\mathbf{u}_i, \mathbf{w})) \neq 0$ uniformly at random, reads $\mathrm{supp}(S_1) \leq k$ coordinates of y, and outputs the sum

$$c \sum_{\lambda \in \mathbb{F}_q : S_1(\lambda) \neq 0} S_1(\lambda) \mathbf{y}_{\mathbf{w}+\lambda \mathbf{v}_i}. \tag{2.3}$$

Probability of correct decoding. First we argue that the decoding is always correct if $\mathscr{A}$ picks $\mathbf{w} \in \mathbb{F}_q^m$ such that all coordinates of $\mathbf{y}$ with labels in the set $\{\mathbf{w} + \lambda \mathbf{v}_i\}_{\lambda : S_1(\lambda) \neq 0}$ are not corrupted. We need to show that for all $i \in [n]$, $\mathbf{x} \in \mathbb{F}_r^n$, and $\mathbf{w} \in \mathbb{F}_q^m$, such that $S_0((\mathbf{u}_i, \mathbf{w})) \neq 0$,

$$c \sum_{\lambda \in \mathbb{F}_q : S_1(\lambda) \neq 0} S_1(\lambda) \left(\sum_{j=1}^{n} \mathbf{x}_j \operatorname{Enc}(\mathbf{e}_j) \right)_{\mathbf{w}+\lambda \mathbf{v}_i} = \mathbf{x}_i. \tag{2.4}$$

Note that

$$\begin{aligned} & c \sum_{\lambda \in \mathbb{F}_q} S_1(\lambda) \left(\sum_{j=1}^{n} \mathbf{x}_j \operatorname{Enc}(\mathbf{e}_j) \right)_{\mathbf{w}+\lambda \mathbf{v}_i} \\ &= c \sum_{j=1}^{n} \mathbf{x}_j \left(\sum_{\lambda \in \mathbb{F}_q} S_1(\lambda) \operatorname{Enc}(\mathbf{e}_j)_{\mathbf{w}+\lambda \mathbf{v}_i} \right) \\ &= c \sum_{j=1}^{n} \mathbf{x}_j \left(\sum_{\lambda \in \mathbb{F}_q} S_1(\lambda) S_0((\mathbf{u}_j, \mathbf{w}+\lambda \mathbf{v}_i)) \right). \end{aligned} \tag{2.5}$$

Now note that

$$\begin{aligned} \sum_{\lambda \in \mathbb{F}_q} S_1(\lambda) S_0((\mathbf{u}_j, \mathbf{w}+\lambda \mathbf{v}_i)) &= \sum_{\lambda \in \mathbb{F}_q} S_1(\lambda) S_0((\mathbf{u}_j, \mathbf{w}) + \lambda (\mathbf{u}_j, \mathbf{v}_i)) \\ &= \begin{cases} 1/c, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

For $i = j$, the last identity above follows from $(\mathbf{u}_i, \mathbf{v}_i) = 0$ and the definition of the constant c. For $i \neq j$, the identity follows from $(\mathbf{u}_j, \mathbf{v}_i) \in S$ and the algebraic niceness of S. Combining (2.5) with the identity above, we get (2.4).

Now assume that up to a fraction δ of the coordinates of $\mathbf{y}$ are corrupted. Let T_i denote the set of coordinates whose labels belong to

$$\left\{ \mathbf{w} \in \mathbb{F}_q^m \mid S_0((\mathbf{u}_i, \mathbf{w})) \neq 0 \right\}.$$

It is not hard to see that $|T_i| = q^{m-1} \operatorname{supp}(S_0)$. Thus at most a fraction $\delta q / \operatorname{supp}(S_0)$ of the coordinates in T_i are corrupted. Let

$$Q_i = \left\{ \{\mathbf{w} + \lambda \mathbf{v}_i\}_{\lambda \in \mathbb{F}_q : S_1(\lambda) \neq 0} \mid \mathbf{w} : S_0((\mathbf{u}_i, \mathbf{w})) \neq 0 \right\}$$

be the family of $\operatorname{supp}(S_1)$-tuples of coordinates that may be queried by $\mathscr{A}$. $(\mathbf{u}_i, \mathbf{v}_i) = 0$ implies that the elements of Q_i uniformly cover the set T_i. Combining the last two observations, we conclude that with probability at least $1 - \delta q k / \operatorname{supp}(S_0)$, $\mathscr{A}$ picks an uncorrupted $\operatorname{supp}(S_1) \leq k$-tuple and outputs the correct value of $\mathbf{x}_i$. □

The parameters of the locally decodable code that one gets by applying Lemma 4 to a (nice) set S depend on the support of S_0, where S_0 is the map in the definition of the algebraic niceness of S. The next lemma shows that one can always ensure that the support of S_0 is large, and thus obtain a good dependence of the decoding error on the fraction of corrupted locations.

Lemma 5. *Let q be a prime power and let r be a prime. Let $S \subseteq \mathbb{F}_q^*$ be a k-algebraically nice set over $\mathbb{F}_r$. Let S_0, S_1 be the maps in the definition of the algebraic niceness of S. One can always redefine the map S_0 to satisfy* $\mathrm{supp}(S_0) \geq \lceil q(1-1/r) \rceil$.

Proof. The algebraic niceness of S implies that for all $\alpha \in \mathbb{F}_q$ and $\beta \in S$,

$$\sum_{\lambda \in \mathbb{F}_q} S_0(\alpha + \beta\lambda) S_1(\lambda) = 0.$$

Equivalently, for all $\alpha \in \mathbb{F}_q$ and $\beta \in S$,

$$\sum_{\lambda \in \mathbb{F}_q} S_0(\lambda) S_1((\lambda - \alpha)\beta^{-1}) = 0. \tag{2.6}$$

Our goal is to redefine the map S_0 to satisfy both (2.6) and $\mathrm{supp}(S_0) \geq \lceil q(1-1/r) \rceil$.

Consider a linear space $M = \mathbb{F}_r^q$ where the coordinates of vectors are labeled by elements of $\mathbb{F}_q$. Note that there is a natural one-to-one correspondence between vectors in M and maps from $\mathbb{F}_q$ to $\mathbb{F}_r$. Specifically, a map $f : \mathbb{F}_q \to \mathbb{F}_r$ corresponds to a vector $\mathbf{v} \in M$ such that $\mathbf{v}_\lambda = f(\lambda)$ for all $\lambda \in \mathbb{F}_q$.

Let $L \subseteq M$ be a linear subspace spanned by the vectors corresponding to all maps $f(\lambda) = S_1((\lambda - \alpha)\beta^{-1})$, where $\alpha \in \mathbb{F}_q$ and $\beta \in S$. Observe that L is invariant under the actions of a 1-transitive permutation group (that is permuting the coordinates in accordance with addition in $\mathbb{F}_q$). This implies that the dual space $L^\perp$ is also invariant under the actions of the same group. Note that $L^\perp$ has positive dimension since it contains the vector corresponding to the map S_0. The last two observations imply that $L^\perp$ has *full support*, i.e., for every $i \in [q]$ there exists a vector $\mathbf{v} \in L^\perp$ such that $\mathbf{v}_i \neq 0$. It is easy to verify that any linear subspace of $\mathbb{F}_r^q$ that has full support contains a vector of Hamming weight at least $\lceil q(1-1/r) \rceil$. Let $\mathbf{v} \in L^\perp$ be such a vector. By redefining the map S_0 to be the map from $\mathbb{F}_q$ to $\mathbb{F}_r$ corresponding to the vector $\mathbf{v}$, we conclude the proof. □

The following propositions are the heart of the first part of our construction of LDCs. Combining Lemmas 4 and 5, we get the following proposition.

Proposition 3. *Let q be a prime power and let r be a prime. Assume that $S \subseteq \mathbb{F}_q^*$ is simultaneously (m,n)-combinatorially nice, and k-algebraically nice over $\mathbb{F}_r$. The set S yields an $\mathbb{F}_r$-linear code encoding messages of length n to codewords of length q^m that is $(k, \delta, \delta k r/(r-1))$-locally decodable for all δ.*

Using proposition 3 in combination with the single-parameter definition of combinatorial niceness, we get the following proposition.

Proposition 4. *Let q be a prime power and let r be a prime. Assume that $S \subseteq \mathbb{F}_q^*$ is simultaneously t-combinatorially nice, and k-algebraically nice over $\mathbb{F}_r$; then, for every $n > 0$ there exists an $\mathbb{F}_r$-linear code encoding messages of length n to codewords of length* $\exp\left(n^{1/t}\right)$ *that is $(k, \delta, \delta k r/(r-1))$-locally decodable for all δ.*

Proof. Let $c > 0$ be the constant in the (single-parameter) definition of the combinatorial niceness of S. Given a message of length n, we pad it with zeros to get a message of length n', where $n' \geq n$ is the smallest integer of the form $\lfloor cm^t \rfloor$, and then use the code in proposition 3. It is not hard to verify that the padding results in at most a constant (multiplicative) blowup in the message length, and thus the length of our code is $\exp\left(n^{1/t}\right)$. □

Propositions 3 and 4 identify two properties of subsets of finite fields that together yield good locally decodable codes. These properties are combinatorial and algebraic niceness. Our next goal is to construct nice subsets. In the next sections, we show that if the primes p and r are such that p is a large factor of $r^t - 1$, then the multiplicative subgroup generated by the number r in $\mathbb{F}_p^*$ is sufficiently (algebraically and combinatorially) nice to yield constant-query LDCs of length $\exp\left(n^{1/t}\right)$ over $\mathbb{F}_r$ for all message lengths n.

2.5 Combinatorially nice subsets of $\mathbb{F}_p^*$

In this section we study combinatorial niceness and show that multiplicative subgroups of prime fields are combinatorially nice.

For $\mathbf{w} \in \mathbb{F}_p^m$ and a positive integer l, let $\mathbf{w}^{\otimes l} \in \mathbb{F}_p^{m^l}$ denote the l-th tensor power of $\mathbf{w}$. The coordinates of $\mathbf{w}^{\otimes l}$ are labeled by all possible sequences in $[m]^l$, and

$$\mathbf{w}^{\otimes l}_{i_1,\ldots,i_l} = \prod_{j=1}^{l} \mathbf{w}_{i_j}.$$

Our next goal is to establish the following lemma.

Lemma 6. *Let p be a prime and let $m \geq p-1$ be an integer. Suppose that S is a subgroup of $\mathbb{F}_p^*$; then S is $\left(\binom{m-1+(p-1)/|S|}{(p-1)/|S|}, \binom{m}{p-1}\right)$-combinatorially nice.*

Proof. Let $n = \binom{m}{p-1}$. For $i \in [n]$, let the vectors $\mathbf{u}_i''$ and $\mathbf{v}_i''$ in $\mathbb{F}_p^m$ be the same as the vectors $\mathbf{u}_i, \mathbf{v}_i$ in the proof of Lemma 1, i.e., the vectors $\mathbf{u}_i''$ are incidence vectors of all possible subsets of $[m]$ of cardinality $(p-1)$, and the vectors $\mathbf{v}_i''$ are their complements. Recall that:

- For all $i \in [n]$, $(\mathbf{u}_i'', \mathbf{v}_i'') = 0$.
- For all $i, j \in [n]$ such that $i \neq j$, $(\mathbf{u}_j'', \mathbf{v}_i'') \neq 0$.

Let l be a positive integer and let $\mathbf{u}, \mathbf{v}$ be vectors in $\mathbb{F}_p^m$. Observe that

$$\begin{aligned}\left(\mathbf{u}^{\otimes l}, \mathbf{v}^{\otimes l}\right) &= \sum_{(i_1,\dots,i_l)\in[m]^l} \left(\prod_{j=1}^{l} \mathbf{u}_{i_j} \prod_{j=1}^{l} \mathbf{v}_{i_j}\right) \\ &= \sum_{(i_1,\dots,i_l)\in[m]^l} \left(\prod_{j=1}^{l} \mathbf{u}_{i_j}\mathbf{v}_{i_j}\right) = \left(\sum_{i_1\in[m]} \mathbf{u}_{i_1}\mathbf{v}_{i_1}\right) \dots \left(\sum_{i_l\in[m]} \mathbf{u}_{i_l}\mathbf{v}_{i_l}\right) = (\mathbf{u},\mathbf{v})^l.\end{aligned} \tag{2.7}$$

Let $l = (p-1)/|S|$. For $i \in [n]$ set $\mathbf{u}'_i = \mathbf{u}''^{\otimes l}_i$ and $\mathbf{v}'_i = \mathbf{v}''^{\otimes l}_i$. Equation (2.7) and the fact that $\mathbb{F}_p^*$ is a cyclic group yield the following:

- For all $i \in [n]$, $(\mathbf{u}'_i, \mathbf{v}'_i) = 0$.
- For all $i, j \in [n]$ such that $i \neq j$, $(\mathbf{u}'_j, \mathbf{v}'_i) \in S$.

Note that the vectors $\mathbf{u}'_i$ and $\mathbf{v}'_i$ have $m^{(p-1)/|S|}$ coordinates. Therefore, at this point, we have already shown that the set S is $\left(m^{(p-1)/|S|}, \binom{m}{p-1}\right)$-combinatorially nice.

Let $\mathbf{w}$ be an arbitrary vector in $\mathbb{F}_p^m$. Note that the value of $\mathbf{w}^{\otimes l}_{i_1,\dots,i_l}$ depends on the *multiset* $\{i_1,\dots,i_l\}$ rather than the sequence $i_1,\dots,i_l$. Thus many coordinates of $\mathbf{w}^{\otimes l}$ contain identical (and therefore redundant) values. We are going to reduce the number of coordinates in the vectors $\{\mathbf{u}'_i\}$ and $\{\mathbf{v}'_i\}$ using this observation. Let $F(m,l)$ denote the family of all multi-subsets of $[m]$ of cardinality l. Note that

$$|F(m,l)| = \binom{m-1+l}{l}.$$

For a multiset $\sigma \in F(m,l)$, let $c(\sigma)$ denote the number of sequences in $[m]^l$ that represent σ. Now we are ready to define the vectors $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ in $\mathbb{F}_p^{|F(m,l)|}$. The coordinates of the vectors $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ are labeled by multisets $\sigma \in F(m,l)$. For all $i \in [n]$ and $\sigma \in F(m,l)$, we set

$$(\mathbf{u}_i)_\sigma = c(\sigma)(\mathbf{u}'_i)_\sigma \text{ and } (\mathbf{v}_i)_\sigma = (\mathbf{v}'_i)_\sigma.$$

It is easy to verify that for all $i, j \in [n]$, $(\mathbf{u}_j, \mathbf{v}_i) = \left(\mathbf{u}'_j, \mathbf{v}'_i\right)$. Combining this observation with the properties of the vectors $\mathbf{u}'_i$ and $\mathbf{v}'_i$ that were established earlier, we conclude that the set S is $\left(\binom{m-1+(p-1)/|S|}{(p-1)/|S|}, \binom{m}{p-1}\right)$-combinatorially nice. □

We now give a simple corollary to Lemma 6 that uses a single-parameter definition of combinatorial niceness.

Lemma 7. *Let p be a prime. Suppose that S is a multiplicative subgroup of $\mathbb{F}_p^*$; then S is $|S|$-combinatorially nice.*

Proof. Let $t = |S|$. We need to specify a constant $c > 0$ such that for every positive integer m, there exist two collections of size $n = \lfloor cm^t \rfloor$-sized of m-dimensional vectors over $\mathbb{F}_p$ satisfying:

- For all $i \in [n]$, $(\mathbf{u}_i, \mathbf{v}_i) = 0$.

- For all $i, j \in [n]$ such that $i \neq j$, $(\mathbf{u}_j, \mathbf{v}_i) \in S$.

First, assume that m has the form $m = \binom{m'-1+(p-1)/t}{(p-1)/t}$, for some integer $m' \geq p-1$. In this case Lemma 6 gives us a collection of $n = \binom{m'}{p-1}$ vectors with the right properties. Observe that $n \geq cm^t$ for a constant c that depends only on p and t. Now assume that m does not have the right form, and let m_1 be the largest integer smaller than m that does have the right form. In order to get vectors of dimension m, we use vectors of dimension m_1, obtained from Lemma 6 padded with zeros. It is not hard to verify that such a construction still gives us families of vectors of size $n \geq cm^t$ for a suitably chosen constant c. □

2.6 Algebraically nice subsets of $\mathbb{F}_p^*$

In the previous section, we studied the concept of combinatorial niceness and established that multiplicative subgroups of prime fields are combinatorially nice. In this section we study the concept of algebraic niceness, and show that (under certain constraints on p and r) the multiplicative subgroup generated by r in $\mathbb{F}_p^*$ is algebraically nice over $\mathbb{F}_r$.

We start by introducing some notation. Let p and r be distinct primes.

- The order of r modulo p, which is commonly denoted by $\mathrm{ord}_p(r)$, is the smallest integer t such that $p \mid r^t - 1$.
- $\langle r \rangle \subseteq \mathbb{F}_p^*$ denotes the multiplicative subgroup of $\mathbb{F}_p^*$ generated by the element r. Clearly, $|\langle r \rangle| = \mathrm{ord}_p(r)$.
- $\overline{\mathbb{F}}$ denotes the algebraic closure of the field $\mathbb{F}$.
- $C_r^p \subseteq \overline{\mathbb{F}}_r^*$ denotes the multiplicative subgroup of p-th roots of unity in $\overline{\mathbb{F}}_r$.

Definition 8. Let p and r be distinct primes. We say that there is a *nontrivial k-dependence* between the elements of C_r^p if there exist $\zeta_1, \ldots, \zeta_k \in C_r^p$ and $\sigma_1, \ldots, \sigma_k \in \mathbb{F}_r$ such that

$$\sigma_1 \zeta_1 + \ldots + \sigma_k \zeta_k = 0 \ \text{ and } \ \sigma_1 + \ldots + \sigma_k \neq 0. \tag{2.8}$$

Lemma 8. *Let p and r be distinct primes. Suppose there exists a nontrivial k-dependence between the elements of C_r^p; then $\langle r \rangle \subseteq \mathbb{F}_p^*$ is k-algebraically nice over the field $\mathbb{F}_r$.*

Proof. In what follows, we define a map $S_1 : \mathbb{F}_p \to \mathbb{F}_r$ and prove the existence of a map $S_0 : \mathbb{F}_p \to \mathbb{F}_r$ such that, together, S_0 and S_1 yield k-algebraic niceness of $\langle r \rangle$ over $\mathbb{F}_r$. The identity (2.8) implies that for some $k' \leq k$ there exist k' *distinct* p-th roots of unity $\zeta_1, \ldots, \zeta_{k'} \in C_r^p$ such that for some $\sigma_1, \ldots, \sigma_{k'} \in \mathbb{F}_r$,

$$\sigma_1 \zeta_1 + \ldots + \sigma_{k'} \zeta_{k'} = 0 \ \text{ and } \ \sigma_1 + \ldots + \sigma_{k'} \neq 0. \tag{2.9}$$

Let $t = \mathrm{ord}_p(r)$. Observe that $C_r^p \subseteq \mathbb{F}_{r^t}$. Let g be a multiplicative generator of C_r^p. The identity (2.9) yields

$$\sigma_1 g^{\gamma_1} + \ldots + \sigma_{k'} g^{\gamma_{k'}} = 0,$$

for some distinct values $\{\gamma_i\}_{i\in[k']}$ in $\mathbb{Z}_p$. We define

$$S_1(\lambda) = \begin{cases} \sigma_i, & \text{if } \lambda = \gamma_i, \text{ for some } i \in [k'], \\ 0, & \text{otherwise.} \end{cases}$$

The identity (2.9) yields $\mathrm{supp}(S_1) \leq k$ and

$$\sum_{\lambda\in\mathbb{F}_p} S_1(\lambda) \neq 0.$$

Now our goal is to prove the existence of a (nonzero) map $S_0 : \mathrm{F}_p \to \mathbb{F}_r$ such that for all $\alpha \in \mathbb{F}_p$ and $\beta \in S$,

$$\sum_{\lambda\in\mathbb{F}_p} S_0(\alpha + \beta\lambda) S_1(\lambda) = 0.$$

Equivalently, we need (a nonzero) map S_0 such that for all $\alpha \in \mathbb{F}_p$ and $\beta \in S$,

$$\sum_{\lambda\in\mathbb{F}_p} S_0(\lambda) S_1((\lambda-\alpha)\beta^{-1}) = 0. \tag{2.10}$$

Consider a natural one-to-one correspondence between maps $S' : \mathbb{F}_p \to \mathbb{F}_r$ and polynomials $\phi_{S'}(x)$ in the ring $\mathbb{F}_r[x]/(x^p - 1)$,

$$\phi_{S'}(x) = \sum_{\lambda\in\mathbb{Z}_p} S'(\lambda) x^\lambda.$$

Clearly, for every map $S' : \mathbb{F}_p \to \mathbb{F}_r$ and every fixed $\alpha, \beta \in \mathbb{F}_p$ such that $\beta \neq 0$,

$$\begin{aligned} \phi_{S'((\lambda-\alpha)\beta^{-1})}(x) &= \sum_{\lambda\in\mathbb{F}_p} S'((\lambda-\alpha)\beta^{-1}) x^\lambda \\ &= \sum_{\lambda\in\mathbb{F}_p} S'(\lambda) x^{\alpha+\beta\lambda} = x^\alpha \phi_{S'}(x^\beta). \end{aligned}$$

Let α be a variable ranging over $\mathbb{F}_p$, and let β be a variable ranging over $\langle r \rangle$. We are going to argue the existence of a map $S_0 : \mathbb{F}_p \to \mathbb{F}_r$ that satisfies (2.10) by showing that all polynomials $\phi_{S_1((\lambda-\alpha)\beta^{-1})}$ belong to a certain linear space $L \in \mathbb{F}_r[x]/(x^p - 1)$ of dimension less than p. In this case any (nonzero) map $T : \mathbb{F}_p \to \mathbb{F}_r$ such that $\phi_T \in L^\perp$ can be used as the map S_0.

Let

$$\tau(x) = \gcd(x^p - 1, \phi_{S_1}(x)).$$

Note that $\tau(x) \neq 1$, since g is a common root of $x^p - 1$ and $\phi_{S_1}(x)$. Let L be the space of polynomials in $\mathbb{F}_r[x]/(x^p - 1)$ that are multiples of $\tau(x)$. Clearly, $\dim L = p - \deg\tau$. Fix some $\alpha \in \mathbb{F}_p$ and $\beta \in \langle r \rangle$. Let us prove that $\phi_{S_1((\lambda-\alpha)\beta^{-1})}(x)$ is in L :

$$\phi_{S_1((\lambda-\alpha)\beta^{-1})}(x) = x^{\alpha}\phi_{S_1}(x^{\beta}) = x^{\alpha}(\phi_{S_1}(x))^{\beta}.$$

The last identity above follows from the fact that for any $f \in \mathbb{F}_r[x]$ and any positive integer i,

$$f\left(x^{r^i}\right) = (f(x))^{r^i}.$$

This completes the proof. □

Lemma 8 reduces the task of proving the k-algebraic niceness of $\langle r\rangle \subseteq \mathbb{F}_p^*$ to certifying the existence of a nontrivial k-dependence in C_r^p. In the following subsections, we present several sufficient conditions for the existence of such a dependence.

Our first sufficient condition (Lemma 9) is the following: p is a Mersenne prime and $r = 2$. The proof that this condition suffices is simple and self-contained. This result alone yields most of our improvements for binary locally decodable codes (see Lemma 14 and Section 2.7.1). Two weaker sufficient conditions are given in Lemmas 10 and 13. Those lemmas have fairly technical proofs and are used later to obtain the most general form of our results for locally decodable codes (see Section 2.7.2).

2.6.1 3-dependences between p-th roots: sufficient conditions

Lemma 9. *Suppose that $p = 2^t - 1$ is a Mersenne prime; then there exists a nontrivial three-dependence in C_2^p.*

Proof. Observe that the polynomial

$$x^p - 1 = x^{2^t-1} - 1 \in \mathbb{F}_2[x]$$

splits into distinct linear factors in the finite field $\mathbb{F}_{2^t}$. Therefore $C_2^p = \mathbb{F}_{2^t}^*$. Pick $\zeta_1 \neq \zeta_2$ in C_2^p arbitrarily. Set $\zeta_3 = \zeta_1 + \zeta_2$. Note that $\zeta_3 \in C_2^p$ and

$$\zeta_1 + \zeta_2 + \zeta_3 = 0.$$

This completes the proof. □

Now we generalize Lemma 9 and show that a substantially weaker condition on p and r is still sufficient. Our argument relies on the classical Weil bound [62, p. 330] for the number of rational points on curves over finite fields.

Lemma 10. *Let p and r be distinct primes. Suppose that $\mathrm{ord}_p(r) < (4/3)\log_r p$; then there exists a nontrivial three-dependence in C_r^p.*

Proof. We start with a brief review of some basic concepts of projective algebraic geometry [29]. Let $\mathbb{F}$ be a field, and let $f \in \mathbb{F}[x,y,z]$ be a homogeneous polynomial. A triple $(x_0,y_0,z_0) \in \mathbb{F}^3$ is called a zero of f if $f(x_0,y_0,z_0) = 0$. A zero is called "nontrivial" if it is different from the origin. An equation $f = 0$ defines a projective plane curve χ_f. Nontrivial zeros of f, considered up to multiplication by a scalar, are called $\mathbb{F}$-rational points of χ_f. If $\mathbb{F}$ is a finite field, it makes sense to talk about the number of $\mathbb{F}$-rational points on a curve.

Let $t = \mathrm{ord}_p(r)$. Note that $C_r^p \subseteq \mathbb{F}_{r^t}$. Pick $\{\sigma_i\}_{i\in[3]}$ in $\mathbb{F}_r^*$ such that

$$\sigma_1 + \sigma_2 + \sigma_3 \neq 0.$$

Consider a projective plane curve χ defined by

$$\sigma_1 x^{(r^t-1)/p} + \sigma_2 y^{(r^t-1)/p} + \sigma_3 z^{(r^t-1)/p} = 0. \tag{2.11}$$

Let us call a point $\mathbf{a}$ on χ "trivial" if one of the coordinates of $\mathbf{a}$ is zero. Clearly, there are at most $3(r^t-1)/p$ trivial points on χ. Note that every nontrivial $\mathbb{F}_{r^t}$-rational point of χ yields a nontrivial 3-dependence in C_r^p (since $\mathbb{F}_{r^t}^*$ is cyclic). The classical Weil bound [62, p. 330] provides an estimate

$$\left|N_q - (q+1)\right| \le (d-1)(d-2)\sqrt{q} \tag{2.12}$$

for the number N_q of $\mathbb{F}_q$-rational points on an arbitrary smooth projective plane curve of degree d. Equation (2.12) implies that if

$$r^t + 1 > \left(\frac{r^t-1}{p} - 1\right)\left(\frac{r^t-1}{p} - 2\right) r^{t/2} + 3\frac{r^t-1}{p} \tag{2.13}$$

there exists a nontrivial point on the curve (2.11). Note that (2.13) follows from

$$r^t + 1 > \left(\frac{r^t}{p}\right)\left(\frac{r^t}{p}\right) r^{t/2} - \frac{2r^{3t/2}}{p} + \frac{3r^t}{p}, \tag{2.14}$$

and (2.14) follows from

$$r^t > \frac{r^{2t+t/2}}{p^2} \quad \text{and} \quad 2r^{t/2} > 3.$$

Now note that the first inequality above follows from $t < (4/3)\log_r p$. To prove the second inequality, observe that $r \ge 3$ implies $2r^{1/2} > 3$, and $r = 2$ implies $t \ge 2$. □

2.6.2 *k*-dependences between *p*-th roots: a sufficient condition

In this section, we show that one can relax the conditions of Lemma 10 further and still ensure the existence of nontrivial k-dependences in C_r^p (for $k \ge 3$). Our proof is quite technical and comes in three steps. First, we briefly review the notion of (additive) Fourier coefficients of subsets of $\mathbb{F}_{r^t}$. Next, we invoke a folklore argument to show that subsets of $\mathbb{F}_{r^t}$ with appropriately small nontrivial Fourier coefficients contain nontrivial k-dependences. Finally, we use a recent result of Bourgain and

Chang [23] (which generalizes the classical estimate for Gauss sums) to argue that (under certain constraints on p and r) all nontrivial Fourier coefficients of C_r^p are small.

For a prime r, let $\mathbb{C}_r$ denote the multiplicative group of complex r-th roots of unity. Let $e \in \mathbb{C}_r$ be an r-th root other than the identity. For $x \in \mathbb{F}_{r^t}$, let

$$\mathrm{Tr}(x) = x + x^r + \ldots + x^{r^{t-1}}$$

denote the trace of x. It is not hard to verify that for all x, $\mathrm{Tr}(x) \in \mathbb{F}_r$. The characters of $\mathbb{F}_{r^t}$ are homomorphisms from the additive group of $\mathbb{F}_{r^t}$ into $\mathbb{C}_r$. There exist r^t characters. We denote the characters by χ_a, where a ranges over $\mathbb{F}_{r^t}$, and set

$$\chi_a(x) = e^{\mathrm{Tr}(ax)}.$$

Let $C(x)$ denote the incidence function of a set $C \subseteq \mathbb{F}_{r^t}$. For arbitrary $a \in \mathbb{F}_{r^t}$, the Fourier coefficient $\hat{C}(\chi_a)$ is defined by

$$\hat{C}(\chi_a) = \sum \chi_a(x) C(x),$$

where the sum is over all $x \in \mathbb{F}_{r^t}$. The Fourier coefficient $\hat{C}(\chi_0) = |C|$ is said to be *trivial*, and the other Fourier coefficients are said to be nontrivial. In what follows, $\sum_a$ stands for summation over all r^t elements of $\mathbb{F}_{r^t}$. We need the following two standard properties of characters and Fourier coefficients:

$$\sum_a \chi_a(x) = \begin{cases} r^t, \text{ if } x = 0, \\ 0, \ \text{otherwise}, \end{cases} \tag{2.15}$$

$$\sum_a \left|\hat{C}(\chi_a)\right|^2 = r^t |C|. \tag{2.16}$$

The following lemma is part of mathematical folklore.

Lemma 11. *Let $C \subseteq \mathbb{F}_{r^t}$ and let $k \geq 3$ be an integer such that there exist $\{\sigma_i\}_{i \in [k]}$ in $\mathbb{F}_r^*$, where $\sum_{i \in [k]} \sigma_i \neq 0$. Let F be the largest absolute value of a nontrivial Fourier coefficient of C. Suppose that*

$$\frac{F}{|C|} < \left(\frac{|C|}{r^t} \right)^{1/(k-2)}; \tag{2.17}$$

then there exists a nontrivial k-dependence between the elements of C.

Proof. Let

$$M(C) = \#\left\{ \zeta_1, \ldots, \zeta_k \in C \mid \sigma_1 \zeta_1 + \ldots + \sigma_k \zeta_k = 0 \right\}.$$

The identity (2.15) yields

$$M(C) = \frac{1}{r^t} \sum_{x_1, \ldots, x_k \in \mathbb{F}_{r^t}} C(x_1) \ldots C(x_k) \sum_a \chi_a(\sigma_1 x_1 + \ldots + \sigma_k x_k). \tag{2.18}$$

Note that

$$\chi_a(\sigma_1 x_1 + \ldots + \sigma_k x_k) = \chi_{\sigma_1 a}(x_1) \ldots \chi_{\sigma_k a}(x_k).$$

Changing the order of summation in (2.18), we get

$$M(C) = \frac{1}{r^t} \sum_a \sum_{x_1,\ldots,x_k \in \mathbb{F}_{r^t}} C(x_1) \ldots C(x_k) \chi_{\sigma_1 a}(x_1) \ldots \chi_{\sigma_k a}(x_k). \tag{2.19}$$

Separating the term corresponding to $a = 0$ in the right-hand side of (2.19), we get

$$M(C) = \frac{|C|^k}{r^t} + \frac{1}{r^t} \sum_{a \neq 0} \prod_{i=1}^k \hat{C}(\chi_{\sigma_i a}) \geq \frac{|C|^k}{r^t} - \frac{1}{r^t} \sum_{a \neq 0} \prod_{i=1}^k \left|\hat{C}(\chi_{\sigma_i a})\right|. \tag{2.20}$$

Using the generalized Holder's inequality [13, p. 20], we obtain

$$\sum_{a \neq 0} \prod_{i=1}^k \left|\hat{C}(\chi_{\sigma_i a})\right| \leq \prod_{i=1}^k \left(\sum_{a \neq 0} \left|\hat{C}(\chi_{\sigma_i a})\right|^k \right)^{1/k}. \tag{2.21}$$

Note that for every $i \in [k]$ we have

$$\sum_{a \neq 0} \left|\hat{C}(\chi_{\sigma_i a})\right|^k \leq F^{k-2} \sum_a \left|\hat{C}(\chi_{\sigma_i a})\right|^2 = F^{k-2} r^t |C|, \tag{2.22}$$

where the last identity follows from (2.16). Combining (2.20), (2.21), and (2.22) we get

$$M(C) \geq \frac{|C|^k}{r^t} - F^{k-2}|C|, \tag{2.23}$$

and conclude that (2.17) implies $M(C) > 0$. □

The following lemma is due to Bourgain and Chang [23, Theorem 1].

Lemma 12. *Assume that $n \mid r^t - 1$ and satisfies the condition*

$$\gcd\left(n, \frac{r^t - 1}{r^{t'} - 1}\right) < r^{t(1-\varepsilon)-t'} \quad \textit{for all} \quad 1 \leq t' < t,\ t' \mid t,$$

where $\varepsilon > 0$ is arbitrary and fixed. Then, for all $a \in \mathbb{F}_{r^t}^*$,

$$\left| \sum_{x \in \mathbb{F}_{r^t}} e^{\mathrm{Tr}(ax^n)} \right| < c_1 r^{t(1-\delta)}, \tag{2.24}$$

where $\delta = \delta(\varepsilon) > 0$ and $c_1 = c_1(\varepsilon)$ are constants.

The main result of this subsection is presented below. Recall that C_r^p denotes the set of p-th roots of unity in $\overline{\mathbb{F}}_r$.

Lemma 13. *For every $c > 0$ and prime r, there exists an integer $k = k(c,r)$ such that the following implication holds. If $p \neq r$ is a prime and $\mathrm{ord}_p(r) < c \log_r p$, then there is a nontrivial k-dependence between the elements of C_r^p.*

Proof. Note that the sum of all p-th roots of unity in $\overline{\mathbb{F}}_r$ is zero. Therefore, given r and c, it suffices to prove the existence of a $k = k(c,r)$ that works for all *sufficiently large p.*

Let $t = \mathrm{ord}_p(r)$. Observe that $p > r^{t/c}$. Assume that p is sufficiently large that $t > 2c$. We now show that the precondition of Lemma 12 holds for $n = (r^t - 1)/p$ and $\varepsilon = 1/(2c)$. Let $t' \mid t$ and $1 \leq t' < t$. Clearly, $\gcd(r^{t'} - 1, p) = 1$. Therefore

$$\gcd\left(\frac{r^t - 1}{p}, \frac{r^t - 1}{r^{t'} - 1}\right) = \frac{r^t - 1}{p(r^{t'} - 1)} < \frac{r^{t(1-1/c)}}{r^{t'} - 1}, \tag{2.25}$$

where the inequality follows from $p > r^{t/c}$. Clearly, $t > 2c$ yields $r^{t/(2c)}/2 > 1$. Multiplying the right-hand side of (2.25) by $r^{t/(2c)}/2$ and using $2(r^{t'} - 1) \geq r^{t'}$, we get

$$\gcd\left(\frac{r^t - 1}{p}, \frac{r^t - 1}{r^{t'} - 1}\right) < r^{t(1-1/(2c))-t'}. \tag{2.26}$$

Combining (2.26) with Lemma 12, we conclude that there exist $\delta > 0$ and c_1 such that for all $a \in \mathbb{F}_{r^t}^*$,

$$\left| \sum_{x \in \mathbb{F}_{r^t}} e^{\mathrm{Tr}\left(ax^{(r^t-1)/p}\right)} \right| < c_1 r^{t(1-\delta)}. \tag{2.27}$$

Observe that $x^{(r^t-1)/p}$ takes every value in C_r^p exactly $(r^t - 1)/p$ times when x ranges over $\mathbb{F}_{r^t}^*$. Thus (2.27) implies

$$(r^t - 1)\left(\frac{F}{p}\right) < c_1 r^{t(1-\delta)} + 1, \tag{2.28}$$

where F denotes the largest absolute value of a nontrivial Fourier coefficient of C_r^p. Assuming that t is sufficiently large, we get

$$(r^t - 1)\left(\frac{F}{p}\right) < c_2 r^{t(1-\delta)}, \tag{2.29}$$

for a suitably chosen constant c_2. Equation (2.29) yields $F/p < (2c_2)r^{-\delta t}$. Pick $k \geq 3$ to be an *odd* integer large enough so that $(1 - 1/c)/(k-2) < \delta$. We now have

$$\frac{F}{p} < r^{-\frac{(1-1/c)t}{(k-2)}} \tag{2.30}$$

for all sufficiently large values of p. Combining $p > r^{t/c}$ with (2.30), we get

$$\frac{F}{\left|C_r^p\right|} < \left(\frac{\left|C_r^p\right|}{r^t}\right)^{1/(k-2)},$$

and an application of Lemma 11 together with the observation that for odd k there always exist $\{\sigma_i\}_{i\in[k]}$ in $\mathbb{F}_r^*$, where $\sum_{i\in[k]} \sigma_i \neq 0$, conclude the proof. □

2.6.3 Summary

We now summarize our sufficient conditions on p and r that yield algebraic niceness of $\langle r\rangle \subseteq \mathbb{F}_p^*$ over $\mathbb{F}_r$. Combining Lemmas 8 and 9 we get the following.

Lemma 14. *Suppose that $p = 2^t - 1$ is a Mersenne prime; then $\langle 2\rangle \subseteq \mathbb{F}_p^*$ is three-algebraically nice over $\mathbb{F}_2$.*

Using Lemma 10 instead of Lemma 9 (in combination with Lemma 8) we get a weaker sufficient condition.

Lemma 15. *Suppose that p and r are distinct primes such that* $\mathrm{ord}_p(r) \leq (4/3)\log_r p$; *then $\langle r\rangle \subseteq \mathbb{F}_p^*$ is three-algebraically nice over $\mathbb{F}_r$.*

Finally, combining Lemmas 8 and 13 we get the following.

Lemma 16. *For every $c > 0$ and prime r there exists an integer $k = k(c,r)$ such that the following implication holds. If $p \neq r$ is a prime and* $\mathrm{ord}_p(r) < c\log_r p$, *then $\langle r\rangle \subseteq \mathbb{F}_p^*$ is k-algebraically nice over $\mathbb{F}_r$.*

2.7 Results

In what follows, we put the results of the previous sections together and summarize our improvements in upper bounds for the codeword length of locally decodable codes.

In Section 2.7.1, we present our results for the narrow case of three-query binary codes. First we show that given a single Mersenne prime $p = 2^t - 1$, one can design three-query binary LDCs of length $\exp\left(n^{1/t}\right)$ for every message length n. Next we review the achievements of the centuries-old study of Mersenne primes, and obtain new families of locally decodable codes that yield large improvements upon earlier work.

In Section 2.7.2, we present the general form of our results. We show that if r is a prime and $r^t - 1$ has a polynomially large prime factor $p \geq r^{\gamma t}$, then for every message length n there exists a $k(\gamma)$-query r-ary LDC of length $\exp\left(n^{1/t}\right)$. The query complexity of the codes that we obtain depends on the size of the largest prime factor of $r^t - 1$, and the codeword length depends on the size of $r^t - 1$ itself. The larger the largest prime factor is, the smaller is the query complexity. The larger $r^t - 1$ is the shorter are the codes.

2.7.1 Results for three-query binary codes

By combining proposition 4 with Lemmas 7 and 14, we conclude that every Mersenne prime $p = 2^t - 1$ yields a family of 3-query locally decodable codes of length $\exp\left(n^{1/t}\right)$.

Theorem 2. *Suppose that $p = 2^t - 1$ is a Mersenne prime; then for every message length n, there exists a binary linear code of length $\exp\left(n^{1/t}\right)$ that is $(3, \delta, 6\delta)$-locally decodable for all δ.*

Mersenne primes have been a popular object of study in number theory for the last few centuries. The largest known Mersenne prime (as of June 2007) is $p = 2^{32\,582\,657} - 1$. It was discovered by Cooper and Boone [1] on September 4, 2006. Plugging p into Theorem 2, we get the following theorem.

Theorem 3. *For every message length n there exists a binary linear code of length $\exp\left(n^{1/32\,582\,657}\right)$ that is $(3, \delta, 6\delta)$-locally decodable for all δ.*

It has often been conjectured that the number of Mersenne primes is infinite. If this conjecture holds, we get three-query locally decodable codes of subexponential length *for infinitely many* message lengths n. To prove this, we first combine Proposition 3 with Lemmas 6 and 14 to obtain the following lemma.

Lemma 17. *Let $p = 2^t - 1$ be a Mersenne prime and let $m \geq p - 1$ be an integer. Let $m' = \binom{m-1+(p-1)/t}{(p-1)/t}$. There exists a binary linear code encoding $n = \binom{m}{p-1}$-bit messages to $p^{m'}$-bit codewords that is $(3, \delta, 6\delta)$-locally decodable code for all δ.*

Now we proceed to constructing a family of three-query binary LDCs of subexponential length.

Theorem 4. *Suppose that the number of Mersenne primes is infinite; then for infinitely many values of the message length n, there exists a binary linear code of length $\exp\left(n^{O(1/\log\log n)}\right)$ that is $(3, \delta, 6\delta)$-locally decodable for all δ.*

Proof. Given a Mersenne prime p, set $m = 2^p$. By substituting m and p into Lemma 17 and doing some basic manipulations, we conclude that there exists a $(3, \delta, 6\delta)$-locally decodable code encoding $n = m^{\Theta(\log m)}$ bits to

$$N = \exp\left(m^{O(\log m / \log\log m)}\right)$$

bits. An observation that $\log\log n = \Theta(\log\log m)$ completes the proof. □

Lenstra, Pomerance, and Wagstaff [2, 74, 90] have made the following conjecture regarding the density of Mersenne primes.

Conjecture 1. Let $M(t)$ be the number of Mersenne primes that are less than or equal to $2^t - 1$; then

$$\lim_{t \to \infty} \frac{M(t)}{\log_2 t} = e^{\gamma},$$

where $\gamma \approx 0.577$ is the Euler–Mascheroni constant.

If this conjecture holds, we get three-query locally decodable codes of subexponential length *for all* message lengths n.

Theorem 5. *Let ε be a positive constant. Suppose that conjecture 1 holds; then for every message length n, there exists a binary linear code of length* $\exp\left(n^{O\left(1/\log^{1-\varepsilon}\log n\right)}\right)$ *that is $(3,\delta,6\delta)$ locally decodable for all δ.*

Proof. Conjecture 1 implies that for all sufficiently large integers z, there is a Mersenne prime between $2^{\log^{1-\varepsilon} z}$ and z. Assume that n is sufficiently large. Pick a Mersenne prime p from the interval

$$\left[2^{\log^{1-\varepsilon}\sqrt{\log n}},\sqrt{\log n}\right].$$

Let m be the smallest integer such that $n \leq \binom{m}{p-1}$. Note that $m = pn^{\Theta(1/p)}$. Given an n-bit message x, we pad it with zeros to a length $\binom{m}{p-1}$ and use the code in Lemma 17 to encode x into a codeword of length $p^{m'}$ for

$$m' = \left(n^{1/p}\log p\right)^{O(p/\log p)}.$$

It remains to note that

$$\log m' = O\left(\frac{\log n}{\log p} + \frac{p\log\log p}{\log p}\right) = O\left(\frac{\log n}{\log^{1-\varepsilon}\log n}\right).$$

This completes the proof. □

2.7.2 Results for general codes

For an integer m, let $P(m)$ denote the largest prime factor of m. Our first theorem gets three-query r-ary LDCs from numbers $m = r^t - 1$ such that $P(m) > m^{3/4}$.

Theorem 6. *Let r be a prime. Suppose that $P(r^t-1) > r^{0.75t}$; then for every message length n, there exists a three-query r-ary code of length* $\exp\left(n^{1/t}\right)$ *that is $(3,\delta,3\delta r/(r-1))$-locally decodable for all δ.*

Proof. Let $P(r^t-1) = p$. Observe that $p \mid r^t - 1$ and $p > r^{0.75t}$ yield

$$\mathrm{ord}_p(r) < (4/3)\log_r p.$$

By combining Lemmas 15 and 7 with Proposition 4, we obtain the statement of the theorem. □

As an example application of theorem 6, one can observe that

$$P(2^{23}-1) = 178\,481 > 2^{(3/4)*23} \approx 155\,872$$

yields a family of three-query locally decodable codes of length $\exp(n^{1/23})$. Theorem 6 immediately yields the following theorem.

Theorem 7. *Let r be a prime. Suppose, for infinitely many t, we have $P(r^t - 1) > r^{0.75t}$; then for every $\varepsilon > 0$ and for every message length n, there exists a three-query r-ary code of length $\exp(n^{\varepsilon})$ that is $(3, \delta, 3\delta r/(r-1))$-locally decodable for all δ.*

The next theorem gets constant-query LDCs from numbers $m = r^t - 1$ with prime factors larger than m^{γ} for every value of γ.

Theorem 8. *Let r be a prime. For every $\gamma > 0$, there exists an integer $k = k(\gamma, r)$ such that the following implication holds. Suppose that $P(r^t - 1) > r^{\gamma t}$; then for every message length n, there exists a k-query r-ary code of length $\exp\left(n^{1/t}\right)$ that is $(k, \delta, \delta kr/(r-1))$-locally decodable for all δ.*

Proof. Let $P(r^t - 1) = p$. Observe that $p \mid r^t - 1$ and $p > r^{\gamma t}$ yield

$$\mathrm{ord}_p(r) < (1/\gamma) \log_r p.$$

By combining Lemmas 16 and 7 with Proposition 4, we obtain the statement of the theorem. □

As an immediate corollary, we get the following.

Theorem 9. *Let r be a prime. Suppose, for some $\gamma > 0$ and infinitely many t, we have $P(r^t - 1) > r^{\gamma t}$; then there is a fixed k such that for every $\varepsilon > 0$ and every message length n, there exists a k-query r-ary code of length $\exp(n^{\varepsilon})$ that is $(k, \delta, \delta kr/(r-1))$-locally decodable for all δ.*

2.8 Addendum

The locally decodable codes of the third generation that were introduced in this book have been developed further in [20, 36, 38, 55, 75]. Specifically,

- Raghavendra [75] suggested an alternative conceptually simpler framework for viewing the construction. The key observation underlying Raghavendra's view is that the maps S_0 and S_1 in the definition of algebraic niceness (Definition 7) can be fixed in a certain canonical form.
- Using Raghavendra's view, Efremenko [38] generalized the construction to work over composites; i.e., Efremenko replaced the field $\mathbb{F}_q$ by a ring $\mathbb{Z}_b$ for a composite b in Definitions 6, 5, and 7. Efremenko used a powerful result of Grolmusz [49] showing that certain subsets of $\mathbb{Z}_b$ are combinatorially far "nicer" than any subsets of $\mathbb{F}_q$, and obtained substantial improvements in upper bounds for the codeword length.
- Finally, Dvir et al. [36] suggested yet another view of the construction. They also studied code parameters in the regime of super-constant query complexity.

In order to demonstrate the key ideas behind the follow-up work, below we review the construction of LDCs of the third generation following the (most recent) view of Dvir et al. [36].

This view fleshes out an intrinsic similarity between locally decodable codes of the third generation and classical Reed–Muller codes. An r-ary locally decodable code consists of a linear subspace of polynomials in $\mathbb{F}_r[z_1,\ldots,z_m]$, evaluated at all points of

$$C_b^m = C_b \times \ldots \times C_b \quad (m \text{ times}),$$

where C_b is a certain multiplicative subgroup of $\mathbb{F}_r^*$.

The decoding algorithm is similar to the traditional local decoders for Reed–Muller codes. The decoder shoots a line in a certain direction and decodes along it (see the locally decodable code described in Section 1.1). The difference is that the monomials which are used are not of low degree; instead, they are chosen according to a *matching* family of vectors (see the following definition). Further, the lines for decoding are *multiplicative*, a notion that we will define shortly.

Definition 9. Let b be an arbitrary positive integer. We say that the families $\mathscr{U} = \{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ and $\mathscr{V} = \{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ of vectors in $\mathbb{Z}_b^m$ form a matching family if the following two conditions are satisfied:

- For all $i \in [n]$, $(\mathbf{u}_i,\mathbf{v}_i) = 0$.
- For all $i,j \in [n]$ such that $i \neq j$, $(\mathbf{u}_j,\mathbf{v}_i) \neq 0$.

Observe that the concept of a matching family is intimately related to the concept of combinatorial niceness of a set (Definition 3). We now show how one can obtain a locally decodable code out of a matching family of vectors. We start with some notation.

- A D-evaluation of a function f defined over a domain D is a vector of values of f at all points of D.
- Let $\mathbf{w} \in \mathbb{Z}_b^m$ be a vector and let $l \in [m]$ be an integer. In this section, we write $\mathbf{w}(l)$ to denote the l-th coordinate of $\mathbf{w}$.
- We assume that r is a prime power and that b divides $r-1$; we denote a multiplicative subgroup of $\mathbb{F}_r^*$ of order b by C_b.
- We fix some generator g of C_b.
- For $\mathbf{w} \in \mathbb{Z}_b^m$, we define $g^{\mathbf{w}} \in C_b^m$ by $\left(g^{\mathbf{w}(1)},\ldots,g^{\mathbf{w}(m)}\right)$.
- For $\mathbf{w},\mathbf{v} \in \mathbb{Z}_b^m$ we define the multiplicative line $M_{\mathbf{w},\mathbf{v}}$ through $\mathbf{w}$ in the direction $\mathbf{v}$ to be the multiset

$$M_{\mathbf{w},\mathbf{v}} = \left\{g^{\mathbf{w}+\lambda\mathbf{v}} \mid \lambda \in \mathbb{Z}_b\right\}. \tag{2.31}$$

- For $\mathbf{u} \in \mathbb{Z}_b^m$, we define the monomial $\mathrm{mon}_{\mathbf{u}} \in \mathrm{\mathbb{F}_r[z_1,\ldots,z_m]}$ by

$$\mathrm{mon}_{\mathbf{u}}(\mathrm{z_1,\ldots,z_m}) = \prod_{\mathrm{l}\in[\mathrm{m}]} \mathrm{z_l}^{\mathbf{u}(\mathrm{l})}. \tag{2.32}$$

Note that for any $\mathbf{w}, \mathbf{u}, \mathbf{v} \in \mathbb{Z}_b^m$ and $\lambda \in \mathbb{Z}_b$ we have

$$\mathrm{mon}_{\mathbf{u}}\left(g^{\mathbf{w}+\lambda\mathbf{v}}\right) = g^{(\mathbf{u},\mathbf{w})}\left(g^{\lambda}\right)^{(\mathbf{u},\mathbf{v})}. \tag{2.33}$$

The formula above implies that the $M_{\mathbf{w},\mathbf{v}}$-evaluation of a monomial $\mathrm{mon}_{\mathbf{u}}$ is a C_b-evaluation of a (univariate) monomial

$$g^{(\mathbf{u},\mathbf{w})} y^{(\mathbf{u},\mathbf{v})} \in \mathbb{F}_r[y]. \tag{2.34}$$

This observation is the foundation of the decoding algorithm. We are now ready to formally specify the locally decodable code.

2.8.1 The code

Proposition 5. *Let $\mathscr{U}, \mathscr{V}$ be a family of matching vectors in $\mathbb{Z}_b^m$, $|\mathscr{U}| = |\mathscr{V}| = n$. Suppose that $b \mid r-1$, where r is a prime power; then there exists an r-ary linear code encoding messages of length n to codewords of length b^m that is $(b, \delta, b\delta)$-locally decodable for all δ.*

Proof. We specify the encoding and decoding procedures for our code as follows.

Encoding. We encode a message $(\mathbf{x}(1), \ldots, \mathbf{x}(n)) \in \mathbb{F}_r^n$ by the C_b^m-evaluation of the polynomial

$$F(z_1, \ldots, z_m) = \sum_{j=1}^{n} \mathbf{x}(j) \times \mathrm{mon}_{\mathbf{u_j}}(\mathrm{z}_1, \ldots, \mathrm{z_m}). \tag{2.35}$$

Decoding. The input to the decoder is a (corrupted) C_b^m-evaluation of F and an index $i \in [n]$. To recover the value $\mathbf{x}(i)$, the decoder picks $\mathbf{w} \in \mathbb{Z}_b^m$ at random, and queries the (possibly corrupted) $M_{\mathbf{w},\mathbf{v}_i}$-evaluation of F at all b points.

We now claim that the noiseless $M_{\mathbf{w},\mathbf{v}_i}$-evaluation of F uniquely determines $\mathbf{x}(i)$. To see this, note that by (2.33), (2.34), and (2.35), the $M_{\mathbf{w},\mathbf{v}_i}$-evaluation of F is a C_b-evaluation of a polynomial

$$f(y) = \sum_{j=1}^{n} \mathbf{x}(j) \times g^{(\mathbf{u}_j,\mathbf{w})} y^{(\mathbf{u}_j,\mathbf{v}_i)} \in \mathbb{F}_r[y]. \tag{2.36}$$

We observe further that the properties of the matching family $\mathscr{U}, \mathscr{V}$ and (2.36) yield

$$f(y) = \mathbf{x}(i) \times g^{(\mathbf{u}_i,\mathbf{w})} + \sum_{s \in \mathbb{Z}_b \setminus \{0\}} \left(\sum_{j\,:\,(\mathbf{u}_j,\mathbf{v}_i)=s} \mathbf{x}(j) \times g^{(\mathbf{u}_j,\mathbf{w})} \right) y^s. \tag{2.37}$$

It is evident from the above formula that

$$\mathbf{x}(i) = f(0)/g^{(\mathbf{u}_i,\mathbf{w})}. \tag{2.38}$$

Therefore all the decoder needs to do is recover the unique univariate polynomial $f(y) \in \mathbb{F}_r[y]$, of degree up to $r-1$, whose C_b-evaluation agrees with the (observed) $M_{\mathbf{w},\mathbf{v}_i}$-evaluation of F, and return $f(0)/g^{(\mathbf{u}_i,\mathbf{w})}$.

To estimate the probability of a decoding error, we note that each individual query of the decoder goes to a uniformly random location, and apply the union bound. □

By applying Proposition 5 to the currently largest known family of matching vectors [49], one gets the locally decodable codes of [38].

Springer

http://www.springer.com/978-3-642-14357-1

Locally Decodable Codes and Private Information
Retrieval Schemes
Yekhanin, S.
2010, XII, 82 p., Hardcover
ISBN: 978-3-642-14357-1