Locally decodable codes: a brief survey

Sergey Yekhanin

Microsoft Research Silicon Valley
yekhanin@microsoft.com

Abstract. Locally decodable codes are error correcting codes that simultaneously provide efficient random-access to encoded data and high noise resilience by allowing reliable reconstruction of an arbitrary data bit from looking at only a small number of randomly chosen codeword bits. Local decodability comes at the price of certain loss in terms of code efficiency. Specifically, locally decodable codes require longer codeword lengths than their classical counterparts. In this work we briefly survey the recent progress in constructions of locally decodable codes.

1 Introduction

Locally Decodable Codes (LDCs) are a special kind of error-correcting codes. Error-correcting codes are used to ensure reliable transmission of information over noisy channels as well as to ensure reliable storage of information on a medium that may be partially corrupted over time (or whose reading device is subject to errors). In both of these applications the message is typically partitioned into small blocks and then each block is encoded separately. Such encoding strategy allows efficient random-access retrieval of the information, since one needs to decode only the portion of data one is interested in. Unfortunately, this strategy yields very poor noise resilience, since in case even a single block (out of possibly tens of thousands) is completely corrupted some information is lost. In view of this limitation it would seem preferable to encode the whole message into a single codeword of an error-correcting code. Such solution clearly improves the robustness to noise, but is also hardly satisfactory, since one now needs to look at the whole codeword in order to recover any particular bit of the message (at least in the case when classical error-correcting codes are used). Such decoding complexity is prohibitive for modern massive data-sets.

Locally decodable codes are error-correcting codes that avoid the problem mentioned above by having extremely efficient *sublinear-time* decoding algorithms. More formally, an r-query locally decodable code C encodes k-symbol messages $\mathbf{x}$ in such a way that one can probabilistically recover any symbol $\mathbf{x}(i)$ of the message by querying only r symbols of the (possibly corrupted) codeword $C(\mathbf{x})$, where r can be as small as 2.

Hadamard code. The classical Hadamard code [MS] encoding k-bit messages to 2^k-bit codewords provides the simplest nontrivial example of locally decodable codes. In what follows, let $[k]$ denote the set $\{1, \ldots, k\}$. Every coordinate in the Hadamard code corresponds to one (of 2^k) subsets of $[k]$ and stores the XOR of the corresponding bits of the message $\mathbf{x}$. Let $\mathbf{y}$ be an (adversarially corrupted) encoding of $\mathbf{x}$. Given an index $i \in [k]$ and $\mathbf{y}$, the Hadamard decoder picks a set S in $[k]$ uniformly at random and outputs the XOR of the two coordinates of $\mathbf{y}$ corresponding to sets S and $S \triangle \{i\}$. (Here, $\triangle$ denotes the symmetric difference of sets such as $\{1,4,5\}\triangle\{4\} = \{1,5\}$, and $\{1,4,5\}\triangle\{2\} = \{1,2,4,5\}$). It is not difficult to verify that if $\mathbf{y}$ differs from the correct encoding of $\mathbf{x}$ in at most δ fraction of coordinates then with probability $1-2\delta$ both decoder's queries go to uncorrupted locations. In such case, the decoder correctly recovers the i-th bit of $\mathbf{x}$. The Hadamard code allows for a super-fast recovery of the message bits (such as, given a codeword corrupted in 0.1 fraction of coordinates, one is able to recover any bit of the message with probability 0.8 by reading only two codeword bits).

The main parameters of interest in locally decodable codes are the codeword length and the query complexity. The length of the code measures the amount of redundancy that is introduced into the message by the encoder. The query complexity counts the number of bits that need to be read from the (corrupted) codeword in order to recover a single bit of the message. Ideally, one would like to have both of these parameters as small as possible. One however can not minimize the length and the query complexity simultaneously. There is a trade-off. On one end of the spectrum we have LDCs with the codeword length close to the message length, decodable with somewhat large query complexity. Such codes are useful for data storage and transmission. On the other end we have LDCs where the query complexity is a small constant but the codeword length is large compared to the message length. Such codes find applications in complexity theory and cryptography. The true shape of the trade-off between the codeword length and the query complexity of LDCs is not known. Determining it is a major open problem.

Currently there are three known families of LDCs: classical Reed Muller codes [MS], multiplicity codes [KSY11], and matching vector codes [Yek08,Efr09]. In this brief survey we give a high level review of each of these families. We focus on the main ideas underlying the codes and omit many details. In section 3 we review Reed Muller codes. In section 4 we review multiplicity codes, and in section 5 we review matching vector codes. A detailed survey of a large body of work on LDCs (including a detailed treatment of the constructions, lower bounds, and applications) can be found in [Yek10].

2 Preliminaries

We now set up the necessary notation and formally define LDCs.

- $[k] = \{1, \ldots, k\}$;
- $\mathbb{F}_q$ is a finite field of q elements;
- $\mathbb{F}_q^*$ is the multiplicative group of $\mathbb{F}_q$;
- $(\mathbf{x}, \mathbf{y})$ stands for the dot product of vectors $\mathbf{x}$ and $\mathbf{y}$;
- $d(\mathbf{x}, \mathbf{y})$ denotes the Hamming distance between vectors $\mathbf{x}$ and $\mathbf{y}$, i.e., the number of coordinates where $\mathbf{x}$ and $\mathbf{y}$ differ;

- For $\mathbf{w} \in \mathbb{F}_q^n$ and an integer $l \in [n]$, $\mathbf{w}(l)$ denotes the l-th coordinate of $\mathbf{w}$;
- A D-evaluation of a function h defined over a domain D, is a vector of values of h at all points of D;
- With a slight abuse of terminology we often refer to a dimension n of a vector $\mathbf{x} \in \mathbb{F}_q^n$ as its *length.*

A q-ary LDC encoding k-long messages to N-long codewords has three parameters: r, δ, and ϵ. Informally an (r, δ, ϵ)-locally decodable code encodes k-long messages $\mathbf{x}$ to N-long codewords $C(\mathbf{x})$, such that for every $i \in [k]$, the coordinate value $\mathbf{x}_i$ can be recovered with probability $1 - \epsilon$, by a randomized decoding procedure that makes only r queries, even if the codeword $C(x)$ is corrupted in up to δN locations. Formally [KT00,STV99],

Definition 1. *A q-ary code $C : \mathbb{F}_q^k \to \mathbb{F}_q^N$ is said to be (r, δ, ϵ)-locally decodable if there exists a randomized decoding algorithm $\mathcal{A}$ such that*

1. *For all $\mathbf{x} \in \mathbb{F}_q^k$, $i \in [k]$ and all vectors $\mathbf{y} \in \mathbb{F}_q^N$ such that $d(C(\mathbf{x}), \mathbf{y}) \leq \delta N$:*
$$Pr[\mathcal{A}^{\mathbf{y}}(i) = \mathbf{x}(i)] \geq 1 - \epsilon,$$
where the probability is taken over the random coin tosses of $\mathcal{A}$.
2. *$\mathcal{A}$ makes at most r queries to $\mathbf{y}$.*

An LDC is called *linear* if C is a linear transformation over $\mathbb{F}_q$. A locally decodable code allows to probabilistically decode any coordinate of a message by probing only few coordinates of its corrupted encoding. A stronger property that is sometimes desirable is that of local correctability, allowing to efficiently recover not only coordinates of the message but also all other coordinates of the encoding. We now formally define Locally Correctable Codes (LCCs).

Definition 2. *A code (set) C in the space $\mathbb{F}_q^N$ is (r, δ, ϵ)-locally correctable if there exists a randomized correcting algorithm $\mathcal{A}$ such that*

1. *For all $\mathbf{c} \in C$, $i \in [N]$ and all vectors $\mathbf{y} \in \mathbb{F}_q^N$ such that $d(\mathbf{c}, \mathbf{y}) \leq \delta N$:*
$$Pr[\mathcal{A}^{\mathbf{y}}(i) = \mathbf{c}(i)] \geq 1 - \epsilon,$$
where the probability is taken over the random coin tosses of $\mathcal{A}$.
2. *$\mathcal{A}$ makes at most r queries to $\mathbf{y}$.*

It is not hard to show [Yek10] that any linear (r, δ, ϵ)-locally correctable code of dimension k in $\mathbb{F}_q^N$ can be turned into a linear (r, δ, ϵ)-locally decodable code encoding k-long q-ary messages to N-long q-ary codewords.

3 Reed Muller codes

Generalized Reed Muller (RM) codes are the oldest family LDCs. All later families can be seen their generalizations. RM codes are named after their discoverers, Reed and Muller. Muller discovered the codes [Mul54] in the 1950s, and Reed proposed the majority logic decoding [Ree54]. The key idea behind these codes is that of polynomial interpolation. Messages are encoded by complete evaluations of low degree multivariate polynomials over a finite field. Local decodability is achieved through reliance on the rich structure of short local dependencies between such evaluations at multiple points.

A Reed Muller code is specified by three integer parameters. Namely, a prime power (alphabet size) q, number of variables n, and a degree $d < q-1$. The q-ary code consists of $\mathbb{F}_q^n$-evaluations of all polynomials of total degree at most d in the ring $\mathbb{F}_q[z_1, \ldots, z_n]$. Such code encodes $k = \binom{n+d}{d}$-long messages over $\mathbb{F}_q$ to q^n-long codewords.

We now show that RM codes are LCCs, presenting the simplest local corrector from [BF90,Lip90]. To recover the value of a degree d polynomial $F \in \mathbb{F}_q[z_1, \ldots, z_n]$ at a point $\mathbf{w} \in \mathbb{F}_q^n$ our local corrector shoots a random affine line through $\mathbf{w}$ and then relies on the local dependency between the values of F at some $d+1$ points along the line.

Proposition 1. *Let n and d be positive integers. Let q be a prime power, $d < q-1$; then there exists a linear code of dimension $k = \binom{n+d}{d}$ in $\mathbb{F}_q^N$, $N = q^n$, that is $(d+1, \delta, (d+1)\delta)$-locally correctable for all δ.*

Proof. The code consists of $\mathbb{F}_q^n$-evaluations of all polynomials of total degree at most d in the ring $\mathbb{F}_q[z_1, \ldots, z_n]$. The local correction procedure is the following. Given an evaluation of a polynomial F corrupted in up to δ fraction of coordinates and a point $\mathbf{w} \in \mathbb{F}_q^n$ the local corrector picks a vector $\mathbf{v} \in \mathbb{F}_q^n$ uniformly at random and considers a line

$$L = \{\mathbf{w} + \lambda \mathbf{v} \mid \lambda \in \mathbb{F}_q\}$$

through $\mathbf{w}$. Let S be an arbitrary subset of $\mathbb{F}_q^*$, $|S| = d+1$. The corrector queries coordinates of the evaluation vector corresponding to points $\mathbf{w} + \lambda \mathbf{v}$, $\lambda \in S$ to obtain values $\{e_\lambda\}$. Next, it recovers the unique univariate polynomial h, $\deg h \leq d$, such that $h(\lambda) = e_\lambda$, for all $\lambda \in S$, and outputs $h(0)$.

Note that in case all queries of our corrector go to uncorrupted locations h is the restriction of F to L, and $h(0) = F(\mathbf{w})$. It remains to note that since each individual query of the corrector goes to a uniformly random location, with probability at least $1 - (d+1)\delta$, it never query a corrupted coordinate.

3.1 Summary of the parameters

The method behind Reed Muller codes is simple and general. It yields codes for all possible values of query complexity r, i.e., one can set r to be an arbitrary function of the message length k by specifying an appropriate relation between the number of variables and the degree of polynomials and letting these parameters grow to infinity. Increasing the degree relative to the number of variables yields shorter codes of larger query complexity. We now specify the parameters of Reed Muller locally decodable codes in the two regimes that are of primary interest to applications, namely, the regime of positive rate (i.e., a setting where the ratio of codeword length to message length stays above some constant), and the regime of constant query complexity:

- For every constant $\epsilon > 0$, Reed Muller codes yield $O(k^{\epsilon})$-query LDCs of rate $\epsilon^{\Omega(1/\epsilon)}$. However, for codes of rate above $1/2$, no nontrivial locality is achieved.
- For every constant $r \geq 2$, Reed Muller codes yield r-query LDCs of length $\exp\left(k^{1/(r-1)}\right)$.[1]

4 Multiplicity codes

Multiplicity codes [KSY11], are the youngest family of LDCs. These codes generalize Reed Muller codes and greatly improve upon them in the regime of high rate. Note that with Reed Muller codes, for the code to have any distance, the degrees of the polynomials need to be smaller than the field size. Multiplicity codes, however, use much higher degree polynomials (and thus have significantly improved rates), and compensate for the loss in distance by evaluating polynomials together with their *partial derivatives.*

In section 3.1 we noted that Reed Muller based LDCs cannot have rate above $1/2$. One can however get $O\left(\sqrt{k}\right)$-query LDCs of rate arbitrary close to half by setting $n = 2$, and $d = (1-\tau)q$ in proposition 1 and letting q grow to infinity.

In what follows, rather than treating multiplicity codes in full generality, we present the most basic family of such codes that have query complexity $O\left(\sqrt{k}\right)$ and rate close to $2/3$. Later, we explain how general multiplicity codes are defined.

Proposition 2. *Let q be a prime power, $\tau > 0$, and $d \leq 2(1-\tau)(q-1)-2$ be an integer; then there exists a linear code of dimension $k = \binom{d+2}{2}$ in $\mathbb{F}_q^N$, $N = 3q^2$, that is $(2(q-1), \delta, 12\delta/\tau + 2/q)$-locally correctable for all δ.*

Proof. Codewords of the multiplicity code correspond to polynomials F in the ring $\mathbb{F}_q[z_1, z_2]$ of total degree up to d. Coordinates are organized in triples indexed by elements of $\mathbf{w} \in \mathbb{F}_q^2$. A triple corresponding to a point $\mathbf{w}$ stores the values

$$F(\mathbf{w}), \left.\frac{\partial F}{\partial z_1}\right|_{\mathbf{w}}, \left.\frac{\partial F}{\partial z_2}\right|_{\mathbf{w}}. \tag{1}$$

We omit a simple proof [KSY11] that distinct polynomials F yield distinct codewords. Given a δ-corrupted codeword corresponding to a polynomial F and a point $\mathbf{w} \in \mathbb{F}_q^2$ the local corrector needs to recover the triple (1).

[1] Throughout the paper we use the standard notation $\exp(x) = 2^{O(x)}$.

1. The corrector starts by picking a vector $\mathbf{v}_1 \in \mathbb{F}_q^2$ uniformly at random and considering a line
$$L_1 = \{\mathbf{w} + \lambda \mathbf{v}_1 \mid \lambda \in \mathbb{F}_q\}$$
through $\mathbf{w}$. The goal of the corrector here is to recover the univariate restriction $f_1(\lambda) = F(\mathbf{w} + \lambda \mathbf{v}_1) \in \mathbb{F}_q[\lambda]$. To this end the corrector queries $3(q-1)$ codeword coordinates corresponding to points $\{\mathbf{w} + \lambda \mathbf{v}_1\}_{\lambda \neq 0}$, to obtain the (possibly corrupted) values of F and partial derivatives of F. The corrector then uses these values to recover the (possibly corrupted) values $\{\mathrm{val}_\lambda, \mathrm{der}_\lambda\}_{\lambda \neq 0}$ of f_1 and the derivative of f_1 via the chain rule
$$f_1'(\lambda) = \left.\frac{\partial F}{\partial z_1}\right|_{\mathbf{w}+\lambda\mathbf{v}_1} \mathbf{v}_1(1) + \left.\frac{\partial F}{\partial z_2}\right|_{\mathbf{w}+\lambda\mathbf{v}_1} \mathbf{v}_1(2). \qquad (2)$$
Next, the corrector recovers the unique univariate polynomial f_1, $\deg f_1 \leq d$, such that
$$f_1(\lambda) = \mathrm{val}_\lambda \quad \text{and} \quad f_1'(\lambda) = \mathrm{der}_\lambda,$$
for all but at most $\lfloor \tau(q-1)/2 \rfloor$ values of $\lambda \in \mathbb{F}_q^*$. The uniqueness of f_1 if it exists follows from the fact that a degree d nonzero univariate polynomial cannot vanish together with its derivative at more than $d/2$ points. If a polynomial f_1 does not exist; the corrector halts with an arbitrary output. Note that since each individual query of the corrector goes to a uniformly random location, by Markov's inequality the probability that $\tau(q-1)/2$ or more of the queries go to corrupted locations is at most $6\delta/\tau$. Therefore with probability at least $1 - 6\delta/\tau$, the recovered polynomial f_1 is indeed the restriction of F to the line L_1. Thus $f_1(0) = F(\mathbf{w})$, and $f_1'(0)$ is the derivative of F in direction $\mathbf{v}_1$.
2. It is not hard to see that knowing the polynomial f_1 is not sufficient to recover (1). Thus on the second step the corrector again picks a uniformly random vector $\mathbf{v}_2 \in \mathbb{F}_q^2$, considers the line L_2 through $\mathbf{w}$ in direction $\mathbf{v}_2$, and recovers the restriction f_2 of F to line L_2, to obtain the value of the directional derivative $f_2'(0)$ of F in direction $\mathbf{v}_2$.
3. Finally, on the last step, the corrector combines directional derivatives of F in directions $\mathbf{v}_1$ and $\mathbf{v}_2$ to recover the partial derivatives of F at $\mathbf{w}$. It is not hard to show [KSY11] that such a recovery is always possible whenever $\mathbf{v}_1$ and $\mathbf{v}_2$ are not collinear, which happens with probability at least $1 - 2/q$.

Proposition 2 yields $O\left(\sqrt{k}\right)$-query codes of rate arbitrarily close to $2/3$ by evaluating bivariate polynomials together with their first partial derivatives. General multiplicity codes are obtained by evaluating n-variate polynomials together with all their mixed partial derivatives of order up to s, for arbitrary positive integers n and s. Increasing n reduces the query complexity; increasing s yields codes of larger rate.

4.1 Summary of the parameters

Setting the number of variables, and the order of derivatives appropriately one can for arbitrary constants $\alpha, \epsilon > 0$ get multiplicity codes of rate $1 - \alpha$ and query complexity $O\left(k^{\epsilon}\right)$. Multiplicity codes also have respectable concrete parameters [KSY11], and thus are potentially useful in practice.

5 Matching vector codes

Matching Vector (MV) locally decodable codes were developed in a sequence of works [Yek08,Rag07,Efr09,IS10,DGY10,BET10a,MFL+10,BET10b,SY11]. In the setting of a constant number of queries these codes have dramatically better parameters than Reed Muller based LDCs. Our presentation of matching vector codes follows the "polynomial-centric" view developed in [DGY10].

An MV code consists of a linear subspace of polynomials in $\mathbb{F}_q[z_1, \ldots, z_n]$, evaluated at all points of $\mathbb{C}_m^n$, where $\mathbb{C}_m$ is a certain multiplicative subgroup of $\mathbb{F}_q^*$. The decoding algorithm is similar to traditional local decoders for RM codes given by proposition 1. The decoder shoots a line in a certain direction and decodes along it. The difference is that the monomials which are used are not of low-degree, they are chosen according to a matching family of vectors. Further, the lines for decoding are *multiplicative*, a notion that we define shortly. In what follows let $\mathbb{Z}_m$ denote the ring of integers modulo an integer m.

Definition 3. *Let $S \subseteq \mathbb{Z}_m \setminus \{0\}$. We say that families $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ and $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ of vectors in $\mathbb{Z}_m^n$ form an S-matching family if the following two conditions are satisfied:*

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

We now show how one can obtain an MV code out of a matching family. We start with some notation.

- We assume that q is a prime power, m divides $q-1$, and denote a subgroup of $\mathbb{F}_q^*$ of order m by $\mathbb{C}_m$;
- We fix some generator g of $\mathbb{C}_m$;
- For $\mathbf{w} \in \mathbb{Z}_m^n$, we define $g^{\mathbf{w}} \in \mathbb{C}_m^n$ by $\left(g^{\mathbf{w}(1)}, \ldots, g^{\mathbf{w}(n)}\right)$;
- For $\mathbf{w}, \mathbf{v} \in \mathbb{Z}_m^n$ we define the multiplicative line $M_{\mathbf{w},\mathbf{v}}$ through $\mathbf{w}$ in direction $\mathbf{v}$ to be the multi-set

$$M_{\mathbf{w},\mathbf{v}} = \left\{g^{\mathbf{w}+\lambda\mathbf{v}} \mid \lambda \in \mathbb{Z}_m\right\}; \tag{3}$$

- For $\mathbf{u} \in \mathbb{Z}_m^n$, we define the monomial $\mathrm{mon}_{\mathbf{u}} \in \mathrm{F_q[z_1, \ldots, z_n]}$ by

$$\mathrm{mon}_{\mathbf{u}}(\mathrm{z}_1, \ldots, \mathrm{z}_\mathrm{n}) = \prod_{\ell \in [\mathrm{n}]} \mathrm{z}_\ell^{\mathbf{u}(\ell)}. \tag{4}$$

We now outline the encoding/decoding framework for matching vector codes. Observe that for any $\mathbf{w}, \mathbf{u}, \mathbf{v} \in \mathbb{Z}_m^n$ and $\lambda \in \mathbb{Z}_m$ we have

$$\mathrm{mon}_{\mathbf{u}}\left(g^{\mathbf{w}+\lambda\mathbf{v}}\right) = g^{(\mathbf{u},\mathbf{w})}\left(g^{\lambda}\right)^{(\mathbf{u},\mathbf{v})}. \tag{5}$$

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

$$g^{(\mathbf{u},\mathbf{w})} y^{(\mathbf{u},\mathbf{v})} \in \mathbb{F}_q[y]. \tag{6}$$

This observation is the foundation of our local decoder. We now sketch encoding and decoding procedures. Let $\mathcal{U}, \mathcal{V}$ be an S-matching family in $\mathbb{Z}_m^n$.

Encoding: We encode a message $(\mathbf{x}(1), \ldots, \mathbf{x}(k)) \in \mathbb{F}_q^k$ by the $\mathbb{C}_m^n$-evaluation of the polynomial

$$F(z_1, \ldots, z_n) = \sum_{j=1}^{k} \mathbf{x}(j) \cdot \mathrm{mon}_{\mathrm{\mathbf{u}_j}}(\mathrm{z_1, \ldots, z_n}). \tag{7}$$

Decoding: The input to the decoder is a (corrupted) $\mathbb{C}_m^n$-evaluation of F and an index $i \in [k]$.

1. The decoder picks $\mathbf{w} \in \mathbb{Z}_m^n$ uniformly at random;
2. The decoder recovers the noiseless restriction of F to $M_{\mathbf{w},\mathbf{v}_i}$. To accomplish this the decoder queries the (corrupted) $M_{\mathbf{w},\mathbf{v}_i}$-evaluation of F at a certain number of locations.

To see that noiseless $M_{\mathbf{w},\mathbf{v}_i}$-evaluation of F uniquely determines $\mathbf{x}(i)$ note that by formulas (5), (6) and (7) the $M_{\mathbf{w},\mathbf{v}_i}$-evaluation of F is a $\mathbb{C}_m$-evaluation of a polynomial

$$f(y) = \sum_{j=1}^{k} \mathbf{x}(j) \cdot g^{(\mathbf{u}_j,\mathbf{w})} y^{(\mathbf{u}_j,\mathbf{v}_i)} \in \mathbb{F}_q[y]. \tag{8}$$

Further observe that properties of the S-matching family $\mathcal{U}, \mathcal{V}$ and (8) yield

$$f(y) = \mathbf{x}(i) \cdot g^{(\mathbf{u}_i,\mathbf{w})} + \sum_{s \in S} \left(\sum_{j \;:\; (\mathbf{u}_j,\mathbf{v}_i)=s} \mathbf{x}(j) \cdot g^{(\mathbf{u}_j,\mathbf{w})} \right) y^s. \tag{9}$$

It is evident from the above formula that the restriction of F to a multiplicative line $M_{\mathbf{w},\mathbf{v}_i}$ yields a univariate polynomial $f(y)$ such that the set of monomial degrees of f is in $S \cup \{0\}$ and

$$\mathbf{x}(i) = f(0)/g^{(\mathbf{u}_i,\mathbf{w})}. \tag{10}$$

Proposition 3. *Let $\mathcal{U}, \mathcal{V}$ be a family of S-matching vectors in $\mathbb{Z}_m^n$, $|\mathcal{U}| = |\mathcal{V}| = k$, $|S| = s$. Suppose $m \mid q-1$, where q is a prime power; then there exists a q-ary linear code encoding k-long messages to m^n-long codewords that is $(s+1, \delta, (s+1)\delta)$-locally decodable for all δ.*

Proof. The encoding procedure has already been specified by formula (7). To recover the value $\mathbf{x}(i)$

1. The decoder picks $\mathbf{w} \in \mathbb{Z}_m^n$ at random, and queries the (corrupted) $M_{\mathbf{w},\mathbf{v}_i}$-evaluation of F at $(s+1)$ consecutive locations $\left\{ g^{\mathbf{w}+\lambda \mathbf{v}_i} \mid \lambda \in \{0, \ldots, s\} \right\}$ to obtain values $c_0, \ldots, c_s$.
2. The decoder recovers the unique sparse univariate polynomial $h(y) \in \mathbb{F}_q[y]$ with supp(h) $\subseteq S \cup \{0\}$ such that for all $\lambda \in \{0, \ldots, s\}$, $h(g^\lambda) = c_\lambda$. (The uniqueness of $h(y)$ follows from standard properties of Vandermonde matrices. [LN83])
3. Following the formula (10) the decoder returns $h(0)/g^{(\mathbf{u}_i, \mathbf{w})}$.

The discussion above implies that if all $(s+1)$ locations queried by the decoder are not corrupted then $h(y)$ is indeed the noiseless restriction of F to $M_{\mathbf{w},\mathbf{v}_i}$, and decoder's output is correct. It remains to note that each individual query of the decoder goes to a uniformly random location and apply the union bound.

5.1 Summary of the parameters

Parameters of MV codes are determined by parameters of the underlying family of matching vectors. The largest currently know such families are based on Grolmusz's set systems with restricted intersections modulo composites [Gro00]. Plugging the parameters of these families into proposition 3 one gets 2^t-query LDCs of length $\exp\exp\left((\log k)^{1/t}(\log\log k)^{1-1/t}\right)$ for any constant $t \geq 2$. Further modest reductions in query complexity are possible [Efr09,IS10,MFL+10].

6 Conclusions

In this paper we have briefly surveyed the three main families of locally decodable codes. Namely, classical Reed Muller codes; multiplicity codes that are the best LDCs in the regime of high rate; and matching vector codes that are the best LDCs in the regime of low query complexity. We focused on the key ideas underlying the constructions and omitted many details, such as alphabet reduction techniques, and extensions that allow to correct more errors. A longer survey of a large body of work on locally decodable codes (including a detailed treatment of the constructions, lower bounds, and applications) can be found in [Yek10].

References

[BET10a] Avraham Ben-Aroya, Klim Efremenko, and Amnon Ta-Shma. Local list decoding with a constant number of queries. In *51st IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 715–722, 2010.

[BET10b] Avraham Ben-Aroya, Klim Efremenko, and Amnon Ta-Shma. A note on amplifying the error-tolerance of locally decodable codes. In *Electronic Colloquium on Computational Complexity (ECCC)*, TR10-134, 2010.

[BF90] Donald Beaver and Joan Feigenbaum. Hiding instances in multioracle queries. In *7th International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 415 of Lecture Notes in Computer Science, pages 37–48. Springer, Berlin, Heidelberg, 1990.

[DGY10] Zeev Dvir, Parikshit Gopalan, and Sergey Yekhanin. Matching vector codes. In *51st IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 705–714, 2010.

[Efr09] Klim Efremenko. 3-query locally decodable codes of subexponential length. In *41st ACM Symposium on Theory of Computing (STOC)*, pages 39–44, 2009.

[Gro00] Vince Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit Ramsey graphs. *Combinatorica*, 20:71–86, 2000.

[IS10] Toshiya Itoh and Yasuhiro Suzuki. New constructions for query-efficient locally decodable codes of subexponential length. *IEICE Transactions on Information and Systems*, pages 263–270, 2010.

[KSY11] Swastik Kopparty, Shubhangi Saraf, and Sergey Yekhanin. High-rate codes with sublinear-time decoding. In *43nd ACM Symposium on Theory of Computing (STOC)*, 2011.

[KT00] Jonathan Katz and Luca Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *32nd ACM Symposium on Theory of Computing (STOC)*, pages 80–86, 2000.

[Lip90] Richard Lipton. Efficient checking of computations. In *7th International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 415 of Lecture Notes in Computer Science, pages 207–215. Springer, Berlin, Heidelberg, 1990.

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

[MFL$^+$10] Yeow Meng Chee, Tao Feng, San Ling, Huaxiong Wang, and Liangfeng Zhang. Query-efficient locally decodable codes of subexponential length. In *Electronic Colloquium on Computational Complexity (ECCC)*, TR10-173, 2010.

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

[Mul54] D. E. Muller. Application of boolean algebra to switching circuit design and to error detection. *IEEE Transactions on Computers*, 3:6–12, 1954.

[Rag07] Prasad Raghavendra. A note on Yekhanin's locally decodable codes. In *Electronic Colloquium on Computational Complexity (ECCC)*, TR07-016, 2007.

[Ree54] Irving S. Reed. A class of multiple-error-correcting codes and the decoding scheme. *IEEE Transactions on Information Theory*, 4:38–49, 1954.

[STV99] Madhu Sudan, Luca Trevisan, and Salil Vadhan. Pseudorandom generators without the XOR lemma. In *39th ACM Symposium on Theory of Computing (STOC)*, pages 537–546, 1999.

[SY11] Shubhangi Saraf and Sergey Yekhanin. Noisy interpolation of sparse polynomials, and applications. In *29th IEEE Computational Complexity Conference (CCC)*, 2011.

[Yek08] Sergey Yekhanin. Towards 3-query locally decodable codes of subexponential length. *Journal of the ACM*, 55:1–16, 2008.

[Yek10] Sergey Yekhanin. Locally decodable codes. *Foundations and trends in theoretical computer science*, 2010. to appear.