Electronic Colloquium on Computational Complexity, Report No. 16 (2007)

A Note on Yekhanin's Locally Decodable Codes

PRASAD RAGHAVENDRA*

Department of Computer Science and Engineering
University of Washington
Seattle, WA 98195

Abstract

Locally Decodable codes(LDC) support decoding of any particular symbol of the input message by reading constant number of symbols of the codeword, even in presence of constant fraction of errors.

In a recent breakthrough [9], Yekhanin constructed 3-query LDCs that hugely improve over earlier constructions. Specifically, for a Mersenne prime $p = 2^t - 1$, binary LDCs of length $2^{O(n^{1/t})}$ for infinitely many n were obtained. Using the largest known Mersenne prime, this implies LDCs of length less than $2^{O(n^{10^{-7}})}$. Assuming infinitude of Mersenne primes, the construction yields LDCs of length $2^{O(n^{1/\log\log n})}$ for infinitely many n.

Inspired by [9], we construct 3-query binary LDCs with same parameters from Mersenne primes. While all the main technical tools are borrowed from [9], we give a self-contained simple construction of LDCs. Our bounds do not improve over [9], and have worse soundness of the decoder. However the LDCs are simpler and generalize naturally to prime fields other than $\mathbb{F}_2 = \{0, 1\}$.

The LDCs presented also translate directly in to three server Private Information Retrieval(PIR) protocols with communication complexities $O(n^{1/t})$ for a database of size n, starting with a Mersenne prime $p = 2^t - 1$.

1 Introduction

The problem of recovering the original message from an erroneous received codeword is central algorithmic problem in coding theory. Using a classical error correcting code, it is possible to encode a message x of n symbols in to $C(x)$ such that it is possible to efficiently retrieve x from an erroneous copy of $C(x)$. Locally decodable codes are error correcting codes with the following additional property : there is a probabilistic procedure to retrieve any particular symbol of the original message by reading only a few symbols from the received message. In other words, local decodability allows one to recover any small part of the original message without having to read the entire received message. This ability for efficient partial recovery has found several applications in complexity theory such as worst case to average case reductions.

Formally, a binary code C is said to be (q, δ, ϵ) locally decodable if the following holds : For any message x, it is possible to recover any bit of x with probability at least $1 - \epsilon$, by making at most q queries to an erroneous copy of $C(x)$ having at most δ fraction of errors. Although the notion of locally decodable codes has been around for more than a decade[1, 7, 6], the formal definition of LDC was first given by Katz and Trevisan [4]. For two queries, the work of Kerendis and de Wolf [5] settled the optimal length to be $2^{\theta(n)}$. The Hadamard code is a 2-query locally decodable code of length 2^n. However with three queries, the best LDCs known were of length $2^{O(n^{1/2})}$ due to Beimel and Ishai [2] while the best known lower bound is $\tilde{\Omega}(n^2)$ [5]. For general q,

*Research supported in part by NSF CCF-0343672

ISSN 1433-8092

the best upper bound was $exp(n^{O(\log\log q/(q\log q))})$ due to Beimel et al[3] and the best lower bound known is $\tilde{\Omega}(n^{1+1/(\lceil q/2\rceil-1)})$ [5]. We refer the reader to [8] for a more detailed survey.

In a recent breakthrough [9], Yekhanin obtained 3-query binary LDCs of length $2^{O(n^{1/t})}$ given a Mersenne prime $p = 2^t - 1$. Using the largest known Mersenne prime, this leads to a huge improvement in the length of LDC from $2^{O(n^{1/2})}$ to $2^{O(n^{10^{-7}})}$. Under the conjecture of infinitude of Mersenne primes, Yekhanin's construction yields 3-query LDCs of length $exp(n^{1/\log\log n})$ for infinitely many values of n.

In this work, we give a self-contained constructions of LDCs that achieve similar parameters as [9]. We stress here that the LDCs in this paper are inspired by Yekhanin's construction, and borrow most of the technical tools from it. Further the bounds we obtain do not improve over [9] in any of the parameters. Our construction has a poorer dependence of soundness of the decoder $(1-\epsilon)$ on the fraction δ of errors.

However the LDCs in this paper are simpler, and generalize more easily. For instance, LDCs using extension fields of characteristic > 2 follow immediately from our construction. Our codes immediately imply 3-server Private Information Retrieval schemes with corresponding communication complexity. The reduction from LDCs to PIR is more direct than in [9], since the queries of the decoder are smooth.

Our presentation brings to fore what we believe is the central theme in [9] : using homomorphisms to construct LDCs. Specifically the idea is to encode the input message in the local structure of a function. Now if the function is a homomorphism, then its local structure translates to all points in the domain. Hence even if the function is corrupt, the message can be retrieved by observing its local structure at a random point.

2 Preliminaries

For a finite field $\mathbb{F}$, a *linear code* over $\mathbb{F}$ is a subspace $C \subset \mathbb{F}^N$. The number of input symbols n that can be encoded by a codeword is equal to the dimension of the subspace C. The *block length* of the code C is N.

Definition 2.1. *A code $C : \Sigma^n \to \Sigma^N$ is said to be (q, δ, ϵ)-locally decodable if there exists a randomized algorithm $\mathcal{A}$ such that*

- *For all $x \in \Sigma^n, i \in [n]$ and $y \in \{0,1\}^N$ such that $d_H(C(x), y) \leqslant \delta N$: $\Pr[\mathcal{A}^y(i) = x_i] \geqslant 1 - \epsilon$ where the probability is over the random choices of the algorithm $\mathcal{A}$.*
- *$\mathcal{A}$ makes at most q queries to y.*

A simple example of 2-query locally decodable codes are Hadamard codes defined below:

Definition 2.2. *For a vector $a \in \mathbb{F}_2^t$, the corresponding Hadamard codeword H_a is a 2^t long binary vector which represents the function $H_a(x) = a \cdot x$.*

As in [9], our construction also relies crucially on sets of vectors U, V with some special properties. Hence we make the following definition:

Definition 2.3. *Two families of vectors $U = \{u_1, \ldots, u_n\}$ and $V = \{v_1, \ldots v_n\}$ in $\mathbb{F}_p^m$ are said to be matching if*

- *For all $i \in [n]$, $u_i \cdot v_i = 0$.*
- *For all $i, j \in [n]$ such that $i \neq j$, $u_j \cdot v_i = 2^{r_{ij}} \mod p$ for some integer r_{ij}.*

The following lemma is implied from the results in [9], we state it here in our notation for the sake of completeness.

Lemma 2.4. *Let p be a prime and let t be the order of $2 \mod p$. Let $n = \binom{M}{p-1}$ and $m = \binom{M-1+\frac{p-1}{t}}{\frac{p-1}{t}}$ for some integer $M > p-1$. Then there are explicit families of vectors $U = \{u_1, \ldots, u_n\}$ and $V = \{v_1, \ldots, v_n\}$ in $\mathbb{F}_p^m$ that form a matching family.*

Proof: Let $e \in \mathbb{F}_p^M$ be the vector that contains 1 in all its coordinates. Let $\{u_i'\}$ be the incidence vectors of all possible $\binom{M}{p-1}$ subsets of $[M]$ of cardinality $(p-1)$. For every i, define $v_i' = e - u_i'$. It is easy to see that $u_i' \cdot v_j' = 0$ if and only if $i = j$. Observe that elements $G = \{1, 2, \ldots, 2^{t-1}\}$ form a subgroup of $\mathbb{F}_p^*$. Let $l = \frac{p-1}{t}$. Define $u_i = u_i'^{\otimes l}$, where $u_i'^{\otimes l}$ is the l^{th} tensor product of u_i'.

$$u_i \cdot v_j = u_i'^{\otimes l} \cdot v_j'^{\otimes l} = (u_i' \cdot v_j')^l$$

For $i \neq j$, $u_i \cdot v_j$ is a l^{th} power and hence an element of G. Further $u_i \cdot v_i = 0$ for all i. Thus the set of vectors u_i, v_i already form a *matching* family. Observe that the dimension of the vectors u_i, v_i is $M^{\frac{p-1}{t}}$. This construction is sufficient to obtain LDCs and PIRs with required parameters up to logarithmic factors in the exponent.

Now we will decrease their dimension from M^l to $\binom{M-1+l}{l}$ by slightly modifying the construction. Towards this, we observe that for an arbitrary vector $w \in \mathbb{F}_p^M$ the value of $w_{i_1,\ldots,i_l}^{\otimes l}$ depends only on the multiset of indices $\{i_1, \ldots i_l\}$. Therefore we reduce the dimension by combining many of these identical(redundant) coordinates in to a single coordinate. Let $F(M, l)$ denote the family of all multi-subsets of $[M]$ of cardinality l. Note that $|F(M, l)| = \binom{M-1+l}{l} = m$. 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 vectors u_i, v_i in $\mathbb{F}_p^m$. Coordinates of u_i, v_i are indexed by multisets $\sigma \in F(M, l)$. For all $i \in [n]$ and $\sigma \in F(m, l)$ we set

$$(u_i)_\sigma = c(\sigma)(u_i'^{\otimes l})_\sigma \text{ and } (v_i)_\sigma = (v_i'^{\otimes l})_\sigma$$

It is easy to verify that for all $i, j \in [n]$, $u_j \cdot v_i = u_j'^{\otimes l} \cdot v_i'^{\otimes l}$. Hence the vectors u_i, v_i for $i \in [n]$ form a *matching* family of vectors with the desired dimension. □

Observe that $n = \binom{M-1}{p-1} \geqslant \left(\frac{M}{p}\right)^{p-1}$. Hence for a fixed $p = 2^t - 1$, we get $m = \binom{M-1+l}{l} \leqslant \left(\frac{e(M-1+l)}{l}\right)^l = O(n^{\frac{1}{t}})$. Further using $M = O(p)$ we get $m < n^{1/\log\log n}$.

3 A Simple Construction

In this section, we present a construction of LDCs over a large alphabet. Let $p = 2^t - 1$ be a Mersenne prime. Let g be a generator of the multiplicative group $\mathbb{F}_{2^t}^*$. Hence clearly we have $g^p = 1$. Further there exists an integer γ such that

$$1 + g + g^\gamma = 0$$

Let $U = \{u_1, \ldots, u_n\}$ and $V = \{v_1, \ldots, v_n\}$ be *matching* families of vectors. For all $i \in [n]$ define a homomorphism $f_i : \mathbb{F}_p^m \to \mathbb{F}_{2^t}^*$ as follows:

$$f_i(x) = g^{u_i \cdot x}$$

It is evident from the definition that f_i is a homomorphism from the additive group of $\mathbb{F}_p^m$ to the multiplicative group $\mathbb{F}_{2^t}^*$. Specifically we have $f_i(x + y) = f_i(x)f_i(y)$, for all $x, y \in \mathbb{F}_p^m$.

Observation 3.1. *For all $j, i \in [n]$ and $x \in \mathbb{F}_p^m$, we have*

$$f_j(x) + f_j(x+v_i) + f_j(x+\gamma v_i) = \begin{cases} 0 & \text{if } j \neq i \\ g^{u_j \cdot x} & \text{if } j = i \end{cases}$$

Proof: By definition,

$$\begin{aligned} f_j(x) + f_j(x+v_i) + f_j(x+\gamma v_i) &= g^{u_j \cdot x} + g^{u_j \cdot (x+v_i)} + g^{u_j \cdot (x+\gamma v_i)} \\ &= g^{u_j \cdot x}(1 + g^{u_j \cdot v_i} + g^{\gamma u_j \cdot v_i}) \end{aligned}$$

Recall that for $j = i$, we have $u_j \cdot v_i = 0$. Hence the above expression reduces to $g^{u_j \cdot x}(1+1+1) = g^{u_j \cdot x}$. For $j \neq i$ we have $u_j \cdot v_i = 2^r \mod p$ for some integer r. Substituting we get

$$\begin{aligned} f_j(x) + f_j(x+v_i) + f_j(x+\gamma v_i) &= g^{u_j \cdot x}(1 + g^{2^r} + g^{\gamma 2^r}) \\ &= g^{u_j \cdot x}(1 + g + g^{\gamma})^{2^r} = 0 \end{aligned}$$

□

Each homomorphism f_i can be thought of as a table of p^m values. Given a vector $\mathbf{a} \in \mathbb{F}_{2^t}^n$, its encoding is the function $C : \mathbb{F}_p^m \to \mathbb{F}_{2^t}$ defined as follows:

$$C(x) = \sum_{i=1}^{n} a_i f_i(x)$$

In other words, the codewords consist of a p^m-long vector of values from $\mathbb{F}_{2^t}$. Towards locally decoding an input symbol a_i, the decoder does the following:

Decoding Algorithm

- Pick a random $x \in \mathbb{F}_p^m$, and query $C(x), C(x+v_i), C(x+\gamma v_i)$.
- Output $a_i = g^{-u_i \cdot x}\left(C(x) + C(x+v_i) + C(x+\gamma v_i)\right)$

We wish to draw an analogy with the 2-query local decoding of the Hadamard codes. Given a Hadamard code word C_a, the i^{th} bit of message a is decoded as $a_i = C_a(x) + C_a(x+v_i)$, where $v_i = e_i$ the i^{th} basis vector.

Theorem 3.2. *Let $p = 2^t - 1$ be a fixed Mersenne prime. There exist linear codes of dimension n over $\mathbb{F}_{2^t}$ with block length at most $2^{O(n^{1/t})}$ that are $(3, \delta, 3\delta)$ locally decodable.*

Proof: First we show that the decoder succeeds with probability at least $1 - 3\delta$. Suppose the values $C(x), C(x+v_i), C(x+\gamma v_i)$ do not have any errors. By definition,

$$g^{-u_i \cdot x}\left(C(x) + C(x+v_i) + C(x+\gamma v_i)\right) = g^{-u_i \cdot x} \sum_{j=1}^{n} a_i \left(f_j(x) + f_j(x+v_i) + f_j(x+\gamma v_i)\right)$$

Using observation 3.1 in the above expression,

$$\begin{aligned} g^{-u_i \cdot x}\left(C(x) + C(x+v_i) + C(x+\gamma v_i)\right) &= g^{-u_i \cdot x} \cdot g^{u_i \cdot x} a_i \\ &= a_i \end{aligned}$$

Hence if all the three values read by the decoder have no noise, then the output is equal to a_i. Observe that for random choice of $x \in \mathbb{F}_p^m$, each of the three query locations are uniformly distributed over $\mathbb{F}_p^m$. In particular, for each of $C(x), C(x+v_i), C(x+\gamma v_i)$ the probability that the value is erroneous is at most δ. Hence with probability at least $1-3\delta$ all the three values are correct, and the decoder outputs a_i.

From Lemma 2.4, there is a *matching* family of vectors with $n = \binom{M}{p-1}$ and $m = \binom{M-1+\frac{p-1}{t}}{\frac{p-1}{t}}$ for all integers $M \geqslant p-1$. The code encodes n symbols over $\mathbb{F}_{2^t}$ in to p^m long vector over the same field. Hence the length of the code is $2^{O(n^{1/t})}$. □

4 Constructing Binary Codes

Towards constructing locally decodable codes over the binary alphabet $\mathbb{F}_2$, we use concatenation with Hadamard code. For a field element $z \in \mathbb{F}_{2^t}$, we will use $[z] \in \mathbb{F}_2^t$ to denote the t-dimensional vector corresponding to z for some fixed representation of the field $\mathbb{F}_{2^t}$.

Recall that the decoder described in Section 3, used the following equation:

$$a_i = g^{-u_i \cdot x} C(x) + g^{-u_i \cdot x} C(x+v_i) + g^{-u_i \cdot x} C(x+\gamma v_i)$$

Each of the elements $C(x), C(x+v_i), C(x+\gamma v_i) \in \mathbb{F}_{2^t}$ are represented as a t-dimensional binary vectors. Multiplication by $g^{-u_i \cdot x}$ is equivalent to a linear transformation on t-dimensional binary vectors. Let $(a_i)_j$ denote the j^{th} bit of a_i. Then for each j, there is a vector $w_j \in \mathbb{F}_2^t$ such that $(a_i)_j = w_j \cdot [C(x)] + w_j \cdot [C(x+v_i)] + w_j \cdot [C(x+\gamma v_i)]$. By definition, $w_j \cdot [C(x)]$ is part of the Hadamard codeword corresponding to $C(x)$. Suppose the codeword contained the Hadamard codeword for each $C(x)$. In case of no errors, using the Hadamard codes corresponding to $C(x), C(x+v_i), C(x+\gamma v_i)$ any bit of a_i can be retrieved by 3- bit queries. In case of errors, it is possible to locally decode with 6 bit queries using the 2-query local decodability of Hadamard codes.

Instead we slightly modify the original codes to construct locally decodable codes with exactly 3 bit queries. Specifically, we restrict the a_i to be $\{0,1\}$ instead of any element from $\mathbb{F}_{2^t}$. The details of the construction and local decoding are described below.

As in Section3, for each $i \in [n]$ we have a homomorphism $f_i : \mathbb{F}_p^m \to \mathbb{F}_{2^t}$. For a vector $\mathbf{a} \in \mathbb{F}_2^n$, the corresponding codeword $H : \mathbb{F}_p^m \times \mathbb{F}_2^t \to \mathbb{F}_2$ is a function defined as follows

$$\begin{aligned} H(x,w) &= w \cdot [C(x)] \\ C(x) &= \sum_{i=1}^{n} a_i f_i(x) \end{aligned}$$

Decoding Algorithm

- Pick $x \in \mathbb{F}_p^m$ uniformly at random.
- Pick $w \in \mathbb{F}_2^t$ such that
$$w \cdot [g^{u_i \cdot x}] = 1$$
- Output $a_i = H(x,w) + H(x+v_i,w) + H(x+\gamma v_i, w)$

Theorem 4.1. *Let $p = 2^t - 1$ be a fixed Mersenne prime. There exist binary linear codes of dimension n with block length at most $2^{O(n^{1/t})}$ that are $(3, \delta, 9\sqrt{\delta})$ locally decodable.*

Proof : To begin with, we analyze the case in which there are no errors. In this case,

$$\begin{aligned} H(x,w) + H(x+v_i,w) + H(x+\gamma v_i,w) &= w \cdot [C(x)] + w \cdot [C(x+v_i)] + w \cdot [C(x+\gamma v_i)] \\ &= w \cdot [C(x) + C(x+v_i) + C(x+\gamma v_i)] \end{aligned}$$

As seen earlier, $C(x) + C(x+v_i) + C(x+\gamma v_i) = a_i g^{u_i \cdot x}$. Substituting

$$\begin{aligned} H(x,w) + H(x+v_i,w) + H(x+\gamma v_i,w) &= w \cdot [a_i g^{u_i \cdot x}] = w \cdot a_i [g^{u_i \cdot x}] \\ &= a_i (w \cdot [g^{u_i \cdot x}]) = a_i \end{aligned}$$

Suppose the codeword has at most δ fraction of errors. Let us call an $x \in \mathbb{F}_p^m$ to be *bad* if for more than $\sqrt{\delta}$-fraction of $w \in \mathbb{F}_2^t$, the value $H(x,w)$ is erroneous. Clearly there cannot be more than $\sqrt{\delta}$ fraction of *bad* vectors x. Since $x, x+v_i, x+\gamma v_i$ are all uniformly distributed, with probability at least $1 - 3\sqrt{\delta}$, none of $x, x+v_i, x+\gamma v_i$ are *bad*. Suppose none of $x, x+v_i, x+\gamma v_i$ are *bad*. For at most $3\sqrt{\delta}2^t$ choices of w, one of $H(x,w), H(x+v_i,w), H(x+\gamma v_i,w)$ is erroneous. As $g^{u_i \cdot x}$ is nonzero, there are exactly 2^{t-1} possible choices for w which satisfy $w \cdot [g^{u_i \cdot x}] = 1$. Consequently with probability at least $(\frac{1}{2} - 3\sqrt{\delta})/\frac{1}{2}$ over the choice of w all the values queried by the decoder are correct. Hence the decoder outputs the correct value of a_i with probability at least

$$\begin{aligned} \Pr[\text{Decoder succeeds}] &\geqslant (1 - 3\sqrt{\delta}) \frac{(\frac{1}{2} - 3\sqrt{\delta})}{\frac{1}{2}} \\ &\geqslant 1 - 9\sqrt{\delta} \end{aligned}$$

From Lemma 2.4, there is a *matching* family of vectors with $n = \binom{M}{p-1}$ and $m = \binom{M-1+\frac{p-1}{t}}{\frac{p-1}{t}}$ for all integers $M \geqslant p-1$. To encode n bits, the above code uses p^m Hadamard codewords each of length 2^t. Hence the length of the code is $2^t 2^{O(n^{1/t})} = 2^{O(n^{1/t})}$. □

Assuming the infinitude of Mersenne primes, the following result is implied:

Theorem 4.2. *Suppose there are infinitely many Mersenne primes, then for infinitely many positive integers n, there exists binary linear codes of dimension n with block length at most $2^{O(n^{1/\log\log n})}$ that are $(3, \delta, 9\sqrt{\delta})$ locally decodable.*

5 Private Information Retrieval

Let $D \in \{0,1\}^n$ be a database. Then a three server PIR protocol on x is defined as follows:

Definition 5.1. *A three server PIR protocol is a triplet of non-uniform algorithms $\mathcal{P} = (\mathcal{Q}, \mathcal{A}, \mathcal{C})$. We assume that each algorithm is given n as an advice. At the beginning of the protocol, the user $\mathcal{U}$ tosses random coins and obtains a random string r. Next $\mathcal{U}$ invokes $\mathcal{Q}(i,r)$ to generate triple of queries (que_1, que_2, que_3). For $i \in [3]$, $\mathcal{U}$ sends que_i to $\mathcal{S}_i$. Each server $\mathcal{S}_j$ responds with an answer $ans_j = \mathcal{A}(j, D, que_j)$(we can assume without loss of generality that servers are deterministic; hence each answer is a function of the query and a database). Finally $\mathcal{U}$ computes its output by applying the reconstruction algorithm $\mathcal{C}(ans_1, ans_2, ans_3, i, r)$. A protocol as above should satisfy the following requirements:*

- ***Correctness :** For any n, $D \in \{0,1\}^n$ and $i \in [n]$, the user outputs the correct value of D_i with probability 1(where the probability is over the random strings r).*

- ***Privacy :*** *Each server individually learns no information about i. To formalize this, let $\mathcal{Q}_j$ denote the j^{th} output of $\mathcal{Q}$. We require that for $j = 1, 2, 3$ and any $n, i_1, i_2 \in [n]$ the distributions $\mathcal{Q}_j(i_1, r)$ and $\mathcal{Q}_j(i_2, r)$ are identical.*

Lemma 5.2. *Let $p = 2^t - 1$ be a Mersenne prime and let $M > p - 1$ be an integer. Let $n = \binom{M}{p-1}$ and $m = M^{\frac{p-1}{t}}$. There exists a three server PIR protocol with question size $m \log p$ and answers of length t that can privately retrieve from a database of length n*

Proof : Encode the database $D \in \{0,1\}^n$ using the simple LDC construction over $\mathbb{F}_{2^t}$ described in Section 3. Specifically let the codeword C be obtained by using $a_i = D_i$ in the LDC. All the three servers have the codeword C and will return any symbol of the codeword on user's request. The user $\mathcal{U}$ runs the decoding algorithm and sends each query of the decoder to one server. Clearly the distribution of each of the queries is uniform. Hence we have a three server PIR protocol. Each query is a point $x \in \mathbb{F}_p^m$, and can be represented by $m \log p$ bits. The server's answer is an element from $\mathbb{F}_{2^t}$ and can be represented in t bits. □

Restating the above lemma, for a fixed Mersenne prime:

Theorem 5.3. *Let $p = 2^t - 1$ be a fixed Mersenne prime. For every positive integer n there exists a three server PIR protocol with questions of length $O(n^{1/t})$ and answers of length t. Specifically for every positive integer n there exists a three server PIR protocol with communication complexity $O(n^{1/32582658})$.*

Under the assumption of existence of infinitely many Mersenne primes we get

Theorem 5.4. *Suppose the number of Mersenne primes is infinite, then for infinitely many values of n there exists a three server PIR protocol with communication complexity of $O(n^{1/\log\log n})$*

6 Generalizations

Let q be a prime. Let p be an odd prime and let t denote the order of q modulo p. By definition of t, p divides $q^t - 1$. Hence there exists an element g of $\mathbb{F}_{q^t}$ that satisfies $g^p = 1$. Let $Q \in \mathbb{F}_q[x]$ be a sparse polynomial with at most c nonzero coefficients satisfying

$$\begin{aligned} Q(g) &= 0 \\ Q(1) &\neq 0 \end{aligned}$$

Recall that in Section 3, we used $Q(g) = 1 + g + g^\gamma$ as the sparse polynomial.

Let $m' > p - 1$ be an integer, and let $n = \binom{M}{p-1}, m = \binom{M-1+\frac{p-1}{t}}{\frac{p-1}{t}}$. Using a construction similar to Lemma 2.4, we obtain families of vectors $U = \{u_1, u_2, \ldots, u_n\}$ and $V = \{v_1, \ldots, v_n\}$ that satisfy:

- For all $i \in [n]$, $u_i \cdot v_i = 0 \mod p$
- For all $i \neq j \in [n]$, $u_j \cdot v_i = q^{r_{ij}} \mod p$ for some integer r_{ij}

Define homomorphisms $f_i(x) = g^{u_i \cdot x}$ from $\mathbb{F}_p^m \to \mathbb{F}_{q^t}$, and construct LDCs similar to Section 3. The number of queries required to decode is equal to the number of nonzero entries in the polynomial Q. Hence we obtain generalized LDCs from primes p, q for which there exist a sparse polynomial Q satisfying $Q(g) = 0, Q(1) \neq 0$.

Acknowledgments: I would like to thank Venkatesan Guruswami for valuable comments on an earlier draft of the paper.

References

[1] Babai, Fortnow, Levin, and Szegedy. Checking computations in polylogarithmic time. In *STOC: ACM Symposium on Theory of Computing (STOC)*, 1991.

[2] Beimel and Ishai. Information-theoretic private information retrieval: A unified construction. In *ICALP: Annual International Colloquium on Automata, Languages and Programming*, 2001.

[3] Beimel, Ishai, Kushilevitz, and Raymond. Breaking the barrier for information-theoretic private information retrieval. In *FOCS: IEEE Symposium on Foundations of Computer Science (FOCS)*, 2002.

[4] J. Katz and L. Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *STOC*, pages 80–86, 2000.

[5] I. Kerenidis and R. de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. In *STOC*, pages 106–115. ACM, 2003.

[6] A. Polishchuk and D. A. Spielman. Nearly-linear size holographic proofs. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on the Theory of Computing*, pages 194–203, Montréal, Québec, Canada, 23–25 May 1994.

[7] M. Sudan. *Efficient Checking of Polynomials and Proofs anf the Hardness of Approximation Problems*, volume 1001 of *Lecture Notes in Computer Science*. Springer, 1995.

[8] Trevisan. Some applications of coding theory in computational complexity. In *ECCCTR: Electronic Colloquium on Computational Complexity, technical reports*, 2004.

[9] S. Yekhanin. New locally decodable codes and private information retrieval schemes. In *ECCCTR: Electronic Colloquium on Computational Complexity, TR06-127*, 2006.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/