Towards 3-Query Locally Decodable Codes of Subexponential Length

SERGEY YEKHANIN

MIT, Cambridge, Massachusetts

Abstract. A q-query Locally Decodable Code (LDC) encodes an n-bit message x as an N-bit codeword $C(x)$, such that one can probabilistically recover any bit x_i of the message by querying only q bits of the codeword $C(x)$, even after some constant fraction of codeword bits has been corrupted.

We give new constructions of three query LDCs of vastly shorter length than that of previous constructions. Specifically, given any Mersenne prime $p = 2^t - 1$, we design three query LDCs of length $N = \exp(O(n^{1/t}))$, for every n. Based on the largest known Mersenne prime, this translates to a length of less than $\exp(O(n^{10^{-7}}))$, compared to $\exp(O(n^{1/2}))$ in the previous constructions. It has often been conjectured that there are infinitely many Mersenne primes. Under this conjecture, our constructions yield three query locally decodable codes of length $N = \exp(n^{O(\frac{1}{\log\log n})})$ for infinitely many n.

We also obtain analogous improvements for Private Information Retrieval (PIR) schemes. We give 3-server PIR schemes with communication complexity of $O(n^{10^{-7}})$ to access an n-bit database, compared to the previous best scheme with complexity $O(n^{1/5.25})$. Assuming again that there are infinitely many Mersenne primes, we get 3-server PIR schemes of communication complexity $n^{O(\frac{1}{\log\log n})}$ for infinitely many n.

Previous families of LDCs and PIR schemes were based on the properties of low-degree multivariate polynomials over finite fields. Our constructions are completely different and are obtained by constructing a large number of vectors in a small dimensional vector space whose inner products are restricted to lie in an algebraically nice set.

Categories and Subject Descriptors: E.4 [**Coding and Information Theory**]—*Error control codes*

General Terms: Theory

Additional Key Words and Phrases: Locally decodable codes, private information retrieval, Mersenne primes

ACM Reference Format:

Yekhanin, S. 2008. Towards 3-query locally decodable codes of subexponential length. J. ACM 55, 1, Article 1, (February 2008), 16 pages. DOI = 10.1145/1326554.1326555 http://doi.acm.org/10.1145/1326554.1326555

Preliminary versions of this article appeared in *Electronic Colloquium on Computational Complexity (ECCC)*. Tech. Rep. TR06-127, 2006 and in *Proceedings of the 39th ACM Symposium on Theory of Computing (STOC)*, ACM, New York, 2007, pp. 266–274.

Author's present address: Institute for Advanced Study (IAS), Princeton, NJ 08540; e-mail: yekhanin@ias.edu.

1. *Introduction*

Classical error-correcting codes allow one to encode an n-bit string x into in N-bit codeword $C(x)$, in such a way that x can still be recovered even if $C(x)$ gets corrupted in a number of coordinates. For instance, codewords $C(x)$ of length $N = O(n)$ already suffice to correct errors in up to δN locations of $C(x)$ for any constant $\delta < 1/4$. The disadvantage of classical error-correction is that one needs to consider all or most of the (corrupted) codeword to recover anything about x. Now suppose that one is only interested in recovering one or a few bits of x. In such case more efficient schemes are possible. Such schemes are known as locally decodable codes (LDCs). Locally decodable codes allow reconstruction of an arbitrary bit x_i, from looking only at q randomly chosen (correlated) coordinates of $C(x)$, where q can be as small as 2. Locally decodable codes have found numerous applications in complexity theory and cryptography. See Trevisan [2004] and Gasarch [2004] for a survey. Below is a slightly informal definition of LDCs:

A (q, δ, ε)-locally decodable code encodes n-bit strings to N-bit codewords $C(x)$, such that for every $i \in [n]$, the bit x_i can be recovered with probability $1 - \varepsilon$, by a randomized decoding procedure that makes only q queries, even if the codeword $C(x)$ is corrupted in up to δN locations.

One should think of $\delta > 0$ and $\varepsilon < 1/2$ as constants. The main parameters of interest in LDCs are the length N and the query complexity q. Ideally we would like to have both of them as small as possible. The notion of locally decodable codes was explicitly discussed in various places in the early 1990s, most notably in Babai et al. [1991], Sudan [1992], and Polishchuk and Spielman [1994]. Katz and Trevisan [2000] were the first to provide a formal definition of LDCs and prove lower bounds on their length. Further work on locally decodable codes includes Beimel et al. [2005, 2002], Deshpande et al. [2002], Obata [2002], Kerenidis and de Wolf [2003], and Wehner and de Wolf [1997]. The length of optimal 2-query LDCs was settled by Kerenidis and de Wolf [2003] and is $\exp(n)$.[1] The length of optimal 3-query LDCs is unknown. The best upper bound prior to our work was $\exp(n^{1/2})$ due to Beimel et al. [2005], and the best lower bound is $\tilde{\Omega}(n^2)$ [Kerenidis and de Wolf 2003; Woodruff 2007]. For general (constant) q the best upper bound was $\exp(n^{O(\log\log q/(q\log q))})$ due to Beimel et al. [2002] and the best lower bound is $\tilde{\Omega}(n^{1+1/(\lceil q/2\rceil-1)})$ [Kerenidis and de Wolf 2003; Woodruff 2007].

The current state of knowledge raises a natural question: Is the poor rate of known constructions an inherent property of locally decodable codes? Indeed, Gasarch [2004, Section 9] and Goldreich [2005, conjecture 4.4] conjecture that the exponential dependence on n, that is, the dependence of the form $N = \exp\left(n^{\Omega(1)}\right)$, is unavoidable for any constant number of queries. As our results suggest, such behavior may well not be inherent.

1.1. Our Results.

We give new families of locally decodable codes whose length is vastly shorter than that of previous constructions. We show that every Mersenne prime p (i.e., a prime of the form $p = 2^t - 1$) yields a family of three query locally decodable codes of length $\exp(n^{1/t})$. The largest Mersenne prime known currently has $t = 32,582,657 > 10^7$. Substituting this prime into our

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

theorem we conclude that for every n there exists a three query locally decodable code of length $\exp(n^{1/32,582,657})$.

It has often been conjectured that the number of Mersenne primes is infinite. If indeed this conjecture holds, our constructions yield three query locally decodable codes of length $N = \exp(n^{O(\frac{1}{\log\log n})})$ *for infinitely many* n. Finally, assuming that the conjecture of Lenstra [1980], Pomerance [1980/81], and Wagstaff [1983, p. 388] regarding the density of Mersenne primes holds, our constructions yield three query locally decodable codes of length $N = \exp(n^{O(\frac{1}{\log^{1-\varepsilon}\log n})})$ *for all* n, for every $\varepsilon > 0$.

1.2. Application to Private Information Retrieval. A q-server *private information retrieval* (PIR) scheme allows a user to retrieve the ith bit of an n-bit string x replicated between q servers while each server individually learns no information about i. The main parameter of interest in a PIR scheme is its communication complexity $C_q(n)$, namely the number of bits exchanged by the user and the servers. Private information retrieval schemes were introduced by Chor et al. [1995]. Further work on PIRs includes Ambainis [1997], Mann [1998], Itoh [1999, 2001], Beimel et al. [2005], Goldreich et al. [2006], Kerenidis and de Wolf [2003], Beigel et al. [2006], Woodruff and Yekhanin [2007], and Razborov and Yekhanin [2006]. Below is a brief summary of known bounds for $C_q(n)$.

The best upper bound for $C_2(n)$ is $O(n^{1/3})$ due to Chor et al. [1995]. The best upper bounds for larger values of q are $C_q(n) \leq n^{O(\log\log q/(q\log q))}$ due to Beimel et al. [2002]. In particular, Beimel et al. [2002] show that $C_3(n) \leq O(n^{1/5.25})$, $C_4(n) \leq O(n^{1/7.87})$ and $C_5(n) \leq O(n^{1/10.83})$. The best lower bound for $C_2(n)$ is $5\log n$ due to Wehner and de Wolf [1997].

Private information retrieval schemes are closely related to locally decodable codes. In particular, our constructions of LDCs yield three server private information retrieval schemes with small communication complexity. We show that every Mersenne prime $p = 2^t - 1$ yields $C_3(n) \leq O(n^{1/(t+1)})$. Instantiating this with the largest known Mersenne prime we get $C_3(n) \leq O(n^{1/32,582,658})$. Assuming that the number of Mersenne primes is infinite our bound goes further down to $n^{O(\frac{1}{\log\log n})}$ *for infinitely many* n. Finally, assuming the density conjecture of Lenstra [1980], Pomerance [1980/81], and Wagstaff [1983, p. 388], we get $C_3(n) \leq n^{O(\frac{1}{\log^{1-\varepsilon}\log n})}$ *for all* n, for every $\varepsilon > 0$.

1.3. Our Technique. All previously known constructions of locally decodable codes and private information retrieval schemes are (implicitly or explicitly) centered around the idea of representing a message x by an evaluation of a certain low degree polynomial over a finite field. Our constructions take a completely different approach. We start by reducing the problem of constructing locally decodable codes to the problem of designing certain families of sets with restricted intersections. We use elementary algebra over finite fields to design such families.

The heart of our construction is the design of a set $S \subseteq \mathbb{F}_p^*$ for a prime p that simultaneously satisfies two properties: (1) There exist two large sequences of vectors $u_1, \ldots, u_n, v_1, \ldots, v_n$ in some low dimensional space $\mathbb{F}_p^m$, such that the dot products $(u_i, v_i) = 0$ for all i, and the dot products $(u_j, v_i) \in S$ for all $i \neq j$. We refer to this property as the combinatorial niceness of S; (2) For a small integer q there exists a q-sparse polynomial $\phi(x) \in \mathbb{F}_2[x]$ such that the common GCD of

all polynomials of the form $\phi(x^\beta)$, $\beta \in S$ and the polynomial $x^p - 1$ is non-trivial. We refer to this property as the algebraic niceness of S. Our notion of combinatorial niceness is related to the notion of set families with restricted intersections in Babai and Frank [1998].

Our construction of locally decodable codes thus comes in three steps: First we show that a set S exhibiting both combinatorial and algebraic niceness leads to good locally decodable codes. In particular the length n of the sequences $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ corresponds to the number of message bits we can encode, while the length of the codewords we build is $N = p^m$. So the longer the sequence and the smaller the dimension the better. The query complexity of our codes is given by the parameter q from the definition of algebraic niceness of S. This step of our construction is quite general and applies to vectors $u_1, \ldots, v_n$ and subsets S over any field. It leads us to the task of identifying good sets that are both combinatorially and algebraically nice, and these tasks narrow our choice of fields. As our second step, we focus on combinatorial niceness. In general big sets tend to be "nicer" (allow longer sequences) than small ones. We show that every multiplicative subgroup of a prime field is combinatorially as nice as its cardinality would allow. This still leaves us with a variety of fields and subsets to work with. Finally, as the last step, we attempt to understand the algebraic niceness of sets. We focus on the very narrow case of Mersenne primes p and the subgroup generated by the element 2 in $\mathbb{F}_p^*$. We manage to show that this subgroup is nice enough to get 3-query locally decodable codes, leading to our final result.

1.4. OUTLINE. In Section 3, we formally define locally decodable codes and introduce certain combinatorial objects that we call regular intersecting families of sets. Those objects later serve as our tool to construct binary LDCs. In Section 4, 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 5 builds upon the construction of Section 4. We formally introduce combinatorial and algebraic niceness and show how the interplay between these two notions yields new LDCs. The last subsection of Section 5 and Section 6 contain our main results for LDCs and private information retrieval schemes.

2. *Notation*

We use the following standard mathematical notation:

—$[s] = \{1, \ldots, s\}$;

—$\mathbb{F}_q$ is a finite field of q elements;

—$\mathbb{F}_q^*$ is the multiplicative group of $\mathbb{F}_q$;

—$d_H(x, y)$ denotes the Hamming distance between binary vectors x and y;

—(u, v) stands for the dot product of vectors u and v.

—For a linear space $L \subseteq \mathbb{F}_2^m$, $L^\perp$ denotes the *dual* space. That is, $L^\perp = \{u \in \mathbb{F}_2^m \mid \forall v \in L, (u, v) = 0\}$.

3. *A Combinatorial Approach to Locally Decodable Codes*

In this section, we formally define locally decodable codes and introduce certain combinatorial objects that we call *regular intersecting families* of sets. We show that regular intersecting families of sets yield binary LDCs.

Definition 3.1. A binary code $C : \{0,1\}^n \to \{0,1\}^N$ is said to be (q, δ, ε)-locally decodable if there exists a randomized decoding algorithm $\mathcal{A}$ such that

(1) For all $x \in \{0,1\}^n$, $i \in [n]$ and $y \in \{0,1\}^N$ such that $d_H(C(x), y) \leq \delta N$: $\Pr[\mathcal{A}^y(i) = x_i] \geq 1 - \varepsilon$,[2] where the probability is taken over the random coin tosses of the algorithm $\mathcal{A}$.

(2) $\mathcal{A}$ makes at most q queries to y.

A locally decodable code is called linear if C is a linear transformation over $\mathbb{F}_2$. A locally decodable code is called nonadaptive if $\mathcal{A}$ makes all its queries simultaneously. Our constructions of locally decodable codes are linear and nonadaptive. They are obtained by viewing the basis elements of the code and the decoding sets of the code as specifying a set system (where a vector corresponds to the set of coordinates on which it is non-zero), with some special intersection properties. We define these properties next.

Definition 3.2. Let N, R and n be positive integers. For $i \in [n]$, $r \in [R]$ let T_i and Q_{ir}, be subsets of $[N]$. We say that subsets T_i and Q_{ir} form a (q, n, N, R, s) regular intersecting family if the following conditions are satisfied:

(1) q is odd;

(2) For all $i \in [n]$, $|T_i| = s$;

(3) For all $i \in [n]$ and $r \in [R]$, $|Q_{ir}| = q$;

(4) For all $i \in [n]$ and $r \in [R]$, $Q_{ir} \subseteq T_i$;

(5) For all $i \in [n]$ and $w \in T_i$, $|\{r \in [R] \mid w \in Q_{ir}\}| = (Rq)/s$, (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 $i \neq j$, $\left|Q_{ir} \cap T_j\right| \equiv 0 \mod (2)$.

The following proposition shows that regular intersecting families imply locally decodable codes.

PROPOSITION 3.3. *A (q, n, N, R, s) regular intersecting family yields a binary linear code encoding n bits to N bits that is $(q, \delta, \delta Nq/s)$ 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 linear code C via its generator matrix $G \in \{0,1\}^{n \times N}$. For $i \in [n]$, we set the ith row of G to be the incidence vector of the set T_i. Below is the description of the decoding algorithm $\mathcal{A}$. Given oracle access to y and input $i \in [n]$, the algorithm $\mathcal{A}$

(1) picks $r \in [R]$ uniformly at random;

(2) outputs the dot product $(y, I(Q_{ir}))$ over $\mathbb{F}_2$.

[2] We remark that many earlier papers about LDCs used the parameter ε in a different way. They required $\Pr[\mathcal{A}^y(i) = x_i] \geq 1/2 + \varepsilon$, rather than $\Pr[\mathcal{A}^y(i) = x_i] \geq 1 - \varepsilon$. We choose to break with this tradition.

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

$$(xG, I(Q_{ir})) = \sum_{j=1}^{n} x_j(I(T_j), I(Q_{ir})) = x_i(I(T_i), I(Q_{ir})) = x_i. \qquad (1)$$

The second equality in formula (1) follows from part (6) of Definition 3.2 and the last equality follows from parts (1), (3), and (4) of Definition 3.2. Now assume that up to δN bits of the encoding xG have been corrupted. Part (5) of Definition 3.2 implies that there are at most $(\delta N R q)/s$ sets Q_{ir} that contain at least one corrupted location. Thus, with probability at least $1 - (\delta N q)/s$, the algorithm $\mathcal{A}$ outputs the correct value. □

4. *Basic Construction*

In this section, we present our basic construction of regular intersecting families that yields q-query locally decodable codes of length $\exp(n^{1/(q-1)})$ for prime values of $q \geq 3$. We choose sets T_i to be unions of cosets of certain hyperplanes and sets Q_{ir} to be lines. We argue the intersection properties based on elementary linear algebra. Let p be an odd prime and $m \geq p - 1$ be an integer.

LEMMA 4.1. *Let* $n = \binom{m}{p-1}$. *There exist two families of vectors* $\{u_1, \ldots, u_n\}$ *and* $\{v_1, \ldots, v_n\}$ *in* $\mathbb{F}_p^m$, *such that*

—*For all* $i \in [n]$, $(u_i, v_i) = 0$;
—*For all* $i, j \in [n]$ *such that* $i \neq j$, $(u_j, v_i) \neq 0$.

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

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

$$T_i = \left\{ x \in \mathbb{F}_p^m \;\middle|\; (u_i, x) \in \mathbb{F}_p^* \right\}.$$

Set $R = s = (p-1) \cdot p^{m-1}$. For each $i \in [n]$ assume some bijection between points of T_i and elements of $[R]$. For $i \in [n]$ and $r \in [R]$, let w_{ir} be the rth point of T_i. Set

$$Q_{ir} = \{w_{ir} + \lambda v_i \mid \lambda \in \mathbb{F}_p\}.^3$$

LEMMA 4.2. *For* $i \in [n]$ *and* $r \in [R]$ *sets* T_i *and* Q_{ir} *form a* (p, n, N, R, s) *regular intersecting family.*

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

PROOF. We simply need to verify that all 6 conditions listed in Definition 3.2 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 $(u_i, w_{ir}) \in \mathbb{F}_p^*$ let us show that $Q_{ir} \subseteq T_i$. By Lemma 4.1 $(u_i, v_i) = 0$. Thus for every $\lambda \in \mathbb{F}_p : (u_i, w_{ir} + \lambda v_i) = (u_i, w_{ir})$. Condition 4 follows.
(5) Fix $i \in [n]$ and $w \in T_i$. Note that
$$|\{r \in [R] \mid w \in Q_{ir}\}| = |\{w_{ir} \in T_i \mid \exists \lambda \in \mathbb{F}_p, w = w_{ir} + \lambda v_i\}| =$$
$$|\{w_{ir} \in T_i \mid \exists \lambda \in \mathbb{F}_p, w_{ir} = w - \lambda v_i\}| = p.$$
It remains to notice that $Rp/s = p$. Condition 5 follows.
(6) Fix $i, j \in [n]$ and $r \in [R]$ such that $i \neq j$. Note that
$$|Q_{ir} \cap T_j| = |\{\lambda \in \mathbb{F}_p \mid (u_j, w_{ir} + \lambda v_i) \in \mathbb{F}_p^*\}| =$$
$$|\{\lambda \in \mathbb{F}_p \mid ((u_j, w_{ir}) + \lambda(u_j, v_i)) \in \mathbb{F}_p^*\}| = p - 1.$$
The last equality follows from the fact that $(u_j, v_i) \neq 0$, and therefore the univariate linear function $(u_j, w_{ir}) + \lambda(u_j, v_i)$ takes every value in $\mathbb{F}_p$ exactly once. It remains to notice that $p - 1$ is even. Condition 6 follows. □

Combining Lemma 4.2 and Proposition 3.3, we get

COROLLARY 4.3. *Let p be an odd prime and $m \geq p - 1$ be an integer. There exists a binary linear code encoding $\binom{m}{p-1}$ bits to p^m bits that is $(p, \delta, \delta p^2/(p-1))$ 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 $\exp(n^{1/(p-1)})$.

THEOREM 4.4. *Let p be a fixed odd prime. For every positive integer n there exists a code of length* $\exp(n^{1/(p-1)})$ *that is $(p, \delta, \delta p^2/(p-1))$ 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 x of length n, we pad it with zeros to length n' and use the code from Corollary 4.3 encoding x with a codeword of length $p^m = \exp(n^{1/(p-1)})$. □

5. *Main Construction*

In the previous section, we presented our basic linear algebraic construction of regular intersecting families. We chose sets T_i to be unions of cosets of certain hyperplanes. We chose sets Q_{ir} to be lines. The high-level idea behind our main construction, is to reduce the number of codeword locations queried by choosing sets Q_{ir} to be *proper subsets of lines* rather than whole lines. Before we proceed to our main construction, we introduce two central technical concepts of our article, namely *combinatorial* and *algebraic niceness*. Let p be an odd prime.

Definition 5.1. A set $S \subseteq \mathbb{F}_p^*$ is called (m, n)-combinatorially nice if there exist two families of vectors $\{u_1, \ldots, u_n\}$ and $\{v_1, \ldots, v_n\}$ in $\mathbb{F}_p^m$, such that

—For all $i \in [n]$, $(u_i, v_i) = 0$;
—For all $i, j \in [n]$ such that $i \neq j$, $(u_j, v_i) \in S$.

Remark 5.2. Note that, in Lemma 4.1, we established that the set $S = \mathbb{F}_p^*$ is $(m, \binom{m}{p-1})$-combinatorially nice for every integer $m \geq p - 1$.

Definition 5.3. A set $S \subseteq \mathbb{F}_p^*$ is called q-algebraically nice if q is odd and there exist two sets $S_0, S_1 \subseteq \mathbb{F}_p$ such that

—S_0 is not empty;
—$|S_1| = q$;
—For all $\alpha \in \mathbb{F}_p$ and $\beta \in S : |S_0 \cap (\alpha + \beta S_1)| \equiv 0 \mod (2)$.

Remark 5.4. It is easy to verify that the set $S = \mathbb{F}_p^*$ is p-algebraically nice. Simply pick $S_1 = \mathbb{F}_p$ and $S_0 = \mathbb{F}_p^*$.

5.1. Removing Points from Lines. The next proposition shows how an interplay between combinatorial and algebraic niceness yields regular intersecting families. It is the core of our construction.

Proposition 5.5. *Assume $S \subseteq \mathbb{F}_p^*$ is simultaneously (m, n)-combinatorially nice and q-algebraically nice. Let S_0 be the set from the definition of algebraic niceness of S. The set S yields a $(q, n, p^m, |S_0|p^{m-1}, |S_0|p^{m-1})$ regular intersecting family.*

Proof. For $i \in [n]$ let u_i, v_i be the vectors from 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

$$T_i = \left\{ x \in \mathbb{F}_p^m \;\middle|\; (u_i, x) \in S_0 \right\}.$$

For each $i \in [n]$ assume some bijection between $[R]$ and T_i. Let w_{ir} denote the rth point of T_i. Set

$$Q_{ir} = \{ w_{ir} + \lambda v_i \mid \lambda \in S_1 \}.$$

It remains to verify that all six conditions listed in Definition 3.2 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 $(u_i, w_{ir}) \in S_0$ let us show that $Q_{ir} \subseteq T_i$. Definition 5.1 implies that $(u_i, v_i) = 0$. Thus for every $\lambda \in S_1$: $(u_i, w_{ir} + \lambda v_i) = (u_i, w_{ir})$. Condition 4 follows.
(5) Fix $i \in [n]$ and $w \in T_i$. Note that

$$|\{r \in [R] \mid w \in Q_{ir}\}| = |\{w_{ir} \in T_i \mid \exists \lambda \in S_1, w = w_{ir} + \lambda v_i\}| =$$
$$|\{w_{ir} \in T_i \mid \exists \lambda \in S_1, w_{ir} = w - \lambda v_i\}| = |S_1| = q.$$

It remains to notice that $Rq/s = q$. Condition 5 follows.

(6) Fix $i, j \in [n]$ and $r \in [R]$ such that $i \neq j$. Note that

$$\begin{aligned}\left|Q_{ir} \cap T_j\right| &= \left|\left\{\lambda \in S_1 \mid \left(u_j, w_{ir} + \lambda v_i\right) \in S_0\right\}\right| \\ &= \left|\left\{\lambda \in S_1 \mid \left(\left(u_j, w_{ir}\right) + \lambda(u_j, v_i)\right) \in S_0\right\}\right| \\ &= \left|S_0 \cap \left(\left(u_j, w_{ir}\right) + (u_j, v_i)S_1\right)\right| \equiv 0 \mod (2).\end{aligned}$$

The last equality follows from the fact that $(u_j, v_i) \in S$, and Definition 5.3. Condition 6 follows. □

Observe that one can derive a regular intersecting family with parameters from Lemma 4.2 using Proposition 5.5 in combination with remarks 5.2 and 5.4.

5.2. On Combinatorially Nice Subsets of $\mathbb{F}_p^*$. For $w \in \mathbb{F}_p^d$ and a positive integer l, let $w^{\otimes l} \in \mathbb{F}_p^{d^l}$ denote the lth tensor power of w. Coordinates of $w^{\otimes l}$ are labelled by all possible sequences in $[d]^l$ and $w^{\otimes l}_{i_1,\ldots,i_l} = \prod_{j=1}^{l} w_{i_j}$. The goal of this section is to establish the following:

Lemma 5.6. *Let p be an odd prime and $d \geq p-1$ be an integer. Suppose S is a subgroup of $\mathbb{F}_p^*$; then S is $\left(\binom{d-1+(p-1)/|S|}{(p-1)/|S|}, \binom{d}{p-1}\right)$-combinatorially nice.*

Proof. Let $n = \binom{d}{p-1}$. For $i \in [n]$ let vectors u_i'' and v_i'' in $\mathbb{F}_p^d$ be the same as vectors u_i, v_i in the proof of Lemma 4.1, that is, vectors u_i'' are incidence vectors of all possible subsets of $[d]$ of cardinality $(p-1)$ and vectors v_i'' are their complements. Recall that

—For all $i \in [n]$, $(u_i'', v_i'') = 0$;
—For all $i, j \in [n]$ such that $i \neq j$, $(u_j'', v_i'') \neq 0$.

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

$$\begin{aligned}\left(u^{\otimes l}, v^{\otimes l}\right) &= \sum_{(i_1,\ldots,i_l)\in[d]^l} \left(\prod_{j=1}^{l} u_{i_j} \prod_{j=1}^{l} v_{i_j}\right) \\ = \sum_{(i_1,\ldots,i_l)\in[d]^l} \left(\prod_{j=1}^{l} u_{i_j} v_{i_j}\right) &= \left(\sum_{i_1\in[d]} u_{i_1} v_{i_1}\right) \ldots \left(\sum_{i_l\in[d]} u_{i_l} v_{i_l}\right) = (u, v)^l.\end{aligned} \tag{2}$$

Let $l = (p-1)/|S|$. Lagrange's theorem implies that l is an integer. For $i \in [n]$, set $u_i' = u_i''^{\otimes l}$ and $v_i' = v_i''^{\otimes l}$. Formula (2) and cyclicity of $\mathbb{F}_p^*$ yield

—For all $i \in [n]$, $(u_i', v_i') = 0$;
—For all $i, j \in [n]$ such that $i \neq j$, $(u_j', v_i') \in S$.

Note that vectors u_i' and v_i' are of dimension $d^{(p-1)/|S|}$. Therefore, at this point, we have already shown that the set S is $\left(d^{(p-1)/|S|}, \binom{d}{p-1}\right)$-combinatorially nice.

Let w be an arbitrary vector in $\mathbb{F}_p^d$. Note that the value of $w^{\otimes l}_{i_1,\ldots,i_l}$ depends on the *multi-set* $\{i_1, \ldots, i_l\}$ rather than the sequence $i_1, \ldots, i_l$. Thus many coordinates of $w^{\otimes l}$ contain identical (and therefore redundant) values. We are going to reduce the dimension of the vectors u_i' and v_i' using this observation. Let $F(d, l)$ denote the family of all multi-subsets of $[d]$ of cardinality l. Note that $|F(d, l)| = \binom{d-1+l}{l}$.

For a multi-set $\sigma \in F(d, l)$, let $c(\sigma)$ denote the number of sequences in $[d]^l$ that represent σ. Now we are ready to define vectors u_i and v_i in $\mathbb{F}_p^{|F(d,l)|}$. The coordinates of the vectors u_i and v_i are labelled by multi-sets $\sigma \in F(d, l)$. For all $i \in [n]$ and $\sigma \in F(d, l)$ we set

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

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

5.3. On Algebraically Nice Subsets of $\mathbb{F}_p^*$. In this section we construct 3-algebraically nice subsets of $\mathbb{F}_p^*$, for primes p that have the form $p = 2^t - 1$. Such primes are known as *Mersenne* primes. Our construction relies on some basic properties of finite fields [Lidl and Niederreiter 1983]. Consider a natural one to one correspondence between subsets S_1 of $\mathbb{F}_p$ and polynomials $\phi_{S_1}(x)$ in the ring $\mathbb{F}_2[x]/(x^p - 1)$:

$$\phi_{S_1}(x) = \sum_{s \in S_1} x^s.$$

It is immediate to verify that for all sets $S_1 \subseteq \mathbb{F}_p$ and all $\alpha, \beta \in \mathbb{F}_p$, such that $\beta \neq 0$:

$$\phi_{\alpha+\beta S_1}(x) = x^\alpha \phi_{S_1}(x^\beta). \tag{3}$$

Lemma 5.7. *Let $p = 2^t - 1$ be a Mersenne prime. The set $S = \{1, 2, 4, 8, \ldots, 2^{t-1}\} \subseteq \mathbb{F}_p^*$ is three algebraically nice.*

Proof. Recall that the multiplicative group $\mathbb{F}_{2^t}^*$ is cyclic. Thus, for every $x \in \mathbb{F}_{2^t}^*$, we have $x^{2^t-1} - 1 = x^p - 1 = 0$. Let g be a generator of $\mathbb{F}_{2^t}^*$. Fix γ such that $1 + g + g^\gamma = 0$. Set $S_1 = \{0, 1, \gamma\}$.

Let α be a variable ranging over $\mathbb{F}_p$ and β be a variable ranging over S. We are going to argue the existence of a set S_0 that has even intersections with all sets of the form $\alpha + \beta S_1$, by showing that all polynomials $\phi_{\alpha+\beta S_1}$ belong to a certain linear space $L \in \mathbb{F}_2[x]/(x^p - 1)$ of dimension less than p. In this case, any nonempty set $T \subseteq \mathbb{F}_p$ such that $\phi_T \in L^\perp$ can be used as the set 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 $1 + x + x^\gamma$. Let L be the space of polynomials in $\mathbb{F}_2[x]/(x^p - 1)$ that are multiples of $\tau(x)$. Clearly, $\dim L = p - \deg \tau$. Fix some $\alpha \in \mathbb{F}_p$ and $\beta \in S$. Let us prove that $\phi_{\alpha+\beta S_1}(x)$ is in L :

$$\phi_{\alpha+\beta S_1}(x) = x^\alpha \phi_{S_1}(x^\beta) = x^\alpha (\phi_{S_1}(x))^\beta.$$

The last identity above follows from the fact that for any polynomial $f \in \mathbb{F}_2[x]$ and any integer i : $f(x^{2^i}) = (f(x))^{2^i}$ and our choice of the set S. □

The parameters of a regular intersecting family that one gets by applying Proposition 5.5 to a certain (nice) set S depend on the size of the set S_0 from the definition of algebraic niceness of S. The next lemma shows that one can always pick the set S_0 to be large.

LEMMA 5.8. *Let $S \subseteq \mathbb{F}_p^*$ be a q-algebraically nice set. Let $S_0, S_1 \subseteq \mathbb{F}_p$ be sets from the definition of algebraic niceness of S. One can always redefine the set S_0 to satisfy $|S_0| \geq \lceil p/2 \rceil$.*

PROOF. Let $L \subset \mathbb{F}_2[x]/(x^p - 1)$ be the linear space spanned by polynomials of the form $\phi_{\alpha+\beta S_1}(x)$, for $\alpha \in \mathbb{F}_p$ and $\beta \in S$. Clearly, the space L is closed under cyclic shifts. This implies that the space $L^\perp$ is also closed under cyclic shifts. Note that $L^\perp$ has positive dimension since $\phi_{S_0}(x) \in L^\perp$. The last two observations imply that $L^\perp$ has *full support,* that is, for every coordinate i there exists a vector $\phi \in L^\perp$ such that $\phi_i \neq 0$. It is easy to verify that any linear subspace of $\mathbb{F}_2^p$ that has full support contains a vector of Hamming weight at least $\lceil p/2 \rceil$. Let $\phi_T(x) \in L^\perp$ be such a vector. Redefining the set S_0 to be the set T, we conclude the proof. □

5.4. RESULTS. Let $p = 2^t - 1$ be a Mersenne prime. Note that the set $S = \{1, 2, 4, 8, \ldots, 2^{t-1}\}$ is a multiplicative subgroup of $\mathbb{F}_p^*$. Combining Proposition 5.5 with Lemmas 5.6, 5.7, and 5.8, we conclude

LEMMA 5.9. *Let $p = 2^t - 1$ be a Mersenne prime and $d \geq p - 1$ be an integer. Let $m = \binom{d-1+(p-1)/t}{(p-1)/t}$. For some integer $z \geq \lceil p/2 \rceil$ there exists a regular intersecting family with parameters*

$$\left(3, \binom{d}{p-1}, p^m, zp^{m-1}, zp^{m-1}\right).$$

Combining Lemma 5.9 with Proposition 3.3, we obtain the key lemma of this article.

LEMMA 5.10. *Let $p = 2^t - 1$ be a Mersenne prime and $d \geq p - 1$ be an integer. Let $m = \binom{d-1+(p-1)/t}{(p-1)/t}$. There exists a binary linear code encoding $n = \binom{d}{p-1}$ bits to p^m bits that is $(3, \delta, 6\delta)$ locally decodable code for all δ.*

For every fixed Mersenne prime $p = 2^t - 1$, we get a family of 3-query LDCs of length $\exp(n^{1/t})$. We omit the proof since it is essentially identical to the proof of Theorem 4.4.

THEOREM 5.11. *Let $p = 2^t - 1$ be a fixed Mersenne prime. For every positive integer n there exists a code of length $\exp(n^{1/t})$ 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. It is still unknown whether the number of Mersenne primes is infinite. There has been a large amount of effort and computational power invested in search for large Mersenne primes. The largest currently known Mersenne prime (as of November 19, 2007) is $p = 2^{32,582,657} - 1$. It was discovered by C. Cooper and S. Boone [2006] on September 4, 2006. Plugging p into Theorem 5.11 we get

THEOREM 5.12. *For every integer n, there exists a code of length $\exp(n^{1/32,582,657})$ 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.

THEOREM 5.13. *Suppose that the number of Mersenne primes is infinite; then for infinitely many values of n there exists a code of length* $\exp(n^{O(\frac{1}{\log\log n})})$*, that is,* $(3, \delta, 6\delta)$ *locally decodable for all* δ.

PROOF. Given a Mersenne prime p, set $d = 2^p$. Substituting d and p into Lemma 5.10 and making some basic manipulations we conclude that there exists a $(3, \delta, 6\delta)$ locally decodable code encoding $n = d^{\Theta(\log d)}$ bits to $N = \exp(d^{O(\frac{\log d}{\log\log d})})$ bits. An observation that $\log\log n = \Theta(\log\log d)$ completes the proof. □

Lenstra [1980], Pomerance [1980/81], and Wagstaff [1983, p. 388] have made the following conjecture regarding the density of Mersenne primes.

CONJECTURE 5.14. *Let* $M(t)$ *be the number of Mersenne primes that are less than or equal to* $2^t - 1$; *then*

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

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

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

THEOREM 5.15. *Let* ε *be a positive constant. Suppose the conjecture* 5.14 *holds; then for all values of n there exists a code of length* $\exp(n^{O(\frac{1}{\log^{1-\varepsilon}\log n})})$ *that is* $(3, \delta, 6\delta)$ *locally decodable for all* δ.

PROOF. Conjecture 5.14 implies that for all sufficiently large integers z there is a Mersenne prime between $2^{\log^{1-\varepsilon} z}$ and z. Assume n is sufficiently large. Pick a Mersenne prime p from the interval $[2^{\log^{1-\varepsilon}\sqrt{\log n}}, \sqrt{\log n}]$. Let d be the smallest integer such that $n \le \binom{d}{p-1}$. Note that $d = pn^{\Theta(1/p)}$. Given an n-bit message x, we pad it with zeros to length $\binom{d}{p-1}$ and use the code from Lemma 5.10 to encode x into a codeword of length p^m for $m = (n^{1/p}\log p)^{O(p/\log p)}$. It remains to notice that $\log m = O(\frac{\log n}{\log p} + \frac{p\log\log p}{\log p}) = O(\frac{\log n}{\log^{1-\varepsilon}\log n})$. □

6. *Application to Private Information Retrieval*

Private information retrieval (PIR) schemes are cryptographic protocols developed in order to protect the privacy of the user's query, when accessing a public database. In such schemes a database (modelled by an n-bit string x) is replicated between few noncommunicating servers. The user holds an index i and is interested in obtaining the value of the bit x_i. To achieve this goal, the user queries each of the servers and gets replies from which the desired bit x_i can be computed. The query to each server is distributed independently of i and therefore each server gets no information about what the user is after. Private information retrieval schemes were introduced by Chor et al. [1995] and received a lot of attention in the subsequent work [Ambainis 1997; Mann 1998; Itoh 1999, 2001; Beimel et al. 2005; Goldreich et al. 2006; Kerenidis and de Wolf 2003; Beigel et al. 2006; Woodruff and Yekhanin 2007; Razborov and Yekhanin 2006].

The main parameters of interest in a PIR scheme are the number of servers involved, and the communication complexity, namely the number of bits exchanged

by the user accessing an n-bit database and the servers. Ideally one would like to keep both of these parameters low. In what follows we show how our results from previous sections yield improved upper bounds for communication complexity of three server PIR schemes. A detailed comparison of our results with the earlier work is given in Section 1.2.

We start with a formal definition of a three server PIR protocol. Let $x \in \{0,1\}^n$ be the database.

Definition 6.1. A three server PIR protocol is a triplet of nonuniform 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 a triple of queries (que_1, que_2, que_3). For $j \in [3]$, $\mathcal{U}$ sends que_j to $\mathcal{S}_j$. Each server $\mathcal{S}_j$ responds with an answer $ans_j = \mathcal{A}(j, x, que_j)$. (We can assume without loss of generality that servers are deterministic; hence, each answer is a function of a 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, $x \in \{0,1\}^n$ and $i \in [n]$, the user outputs the correct value of x_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 jth 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.

There are known generic procedures [Katz and Trevisan 2000] to convert q-query LDCs into q-server PIR schemes. However a simple application of such a procedure to our LDCs will either yield a PIR protocol with perfect privacy, but small probability of error, or a PIR protocol with perfect correctness and some slight privacy leakage. Fortunately, it is possible to achieve both perfect privacy and perfect correctness simultaneously via a specially designed reduction.

LEMMA 6.2. *Let $p = 2^t - 1$ be a Mersenne prime and $d \geq p-1$ be an integer. Let $n = \binom{d}{p-1}$ and $m = \binom{d-1+(p-1)/t}{(p-1)/t}$. There exists a one-round three-server PIR protocol with questions of length $m \log p$ and answers of length p that allows private retrieval of bits from databases of length n.*

PROOF. In the preprocessing stage the servers encode the database x with a three query locally decodable code C from Lemma 5.10. We are going to use the notation from that lemma and Proposition 5.5. Recall that the coordinates of $C(x)$ are in one to one correspondence with points in $\mathbb{F}_p^m$. In order to decode x_i the user has to query three locations $\{w + \lambda v_i \mid \lambda \in S_1\}$ for some $w \in T_i$, where T_i is the union of certain cosets of the hyperplane $\{y \in \mathbb{F}_p^m \mid (u_i, y) = 0\}$. Unlike the LDC setup, in the PIR setup the user can not pick $w \in T_i$ uniformly at random and then query locations $\{w + \lambda v_i \mid \lambda \in S_1\}$ from three different servers, since in such case the servers would observe the uniform distribution on T_i rather than the uniform distribution on $\mathbb{F}_p^m$. Here is our way to go around this problem.

Let $e \in \mathbb{F}_p^m$ be the all-ones vector. The definition of vectors u_i in Lemma 5.6 implies that $(e, u_i) \neq 0 \mod (p)$ for all $i \in [n]$. Thus for every $i \in [n]$ and every

$w \in \mathbb{F}_p^m$ there is some $\gamma_0 \in \mathbb{F}_p$ such that $w + \gamma_0 e \in T_i$. The user picks $w \in \mathbb{F}_p^m$ uniformly at random and (simultaneously) asks p triples of queries of the form $\{w + \gamma e + \lambda v_i \mid \lambda \in S_1\}$ for all $\gamma \in \mathbb{F}_p$. For every triple, the first query always goes to server 1, the second to server 2 and the last to server 3. (Note that, in order to ask all those queries, the user needs to communicate only a single point in $\mathbb{F}_p^m$ to each of the servers.) It is easy to verify that in such case each server individually observes a uniform distribution independent of i, while the user always successfully reconstructs x_i from one of the triples.

Below is a step-by-step summary of our protocol. The protocol involves three servers $\{\mathcal{S}_h\}_{h\in[3]}$ and a user $\mathcal{U}$. The sets $S_1 = \{\lambda_1, \lambda_2, \lambda_3\} \subseteq \mathbb{F}_p$, $\{T_i \subseteq \mathbb{F}_p^m\}_{i\in[n]}$, and vectors $\{v_i \in \mathbb{F}_p^m\}_{i\in[n]}$ are those used in code construction in Lemma 5.10.

$\mathcal{S}_h$	:	Encodes the database x with the code C from Lemma 5.10.
$\mathcal{U}$	:	Picks $w \in \mathbb{F}_p^m$ uniformly at random.
$\mathcal{U} \to \mathcal{S}_h$	:	$w + \lambda_h v_i$
$\mathcal{U} \leftarrow \mathcal{S}_h$	:	$\{C(x)_{w+\gamma e+\lambda_h v_i}\}_{\gamma\in\mathbb{F}_p}$
$\mathcal{U}$	:	Picks $\gamma_0 \in \mathbb{F}_p$ such that $w + \gamma_0 e \in T_i$ and outputs $\sum_{h\in[3]} C(x)_{w+\gamma_0 e+\lambda_h v_i}$.

□

The next theorem captures the asymptotic behavior of our PIR schemes for a fixed Mersenne prime p. We omit the proof since it is essentially identical to the proof of Theorem 4.4.

THEOREM 6.3. *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 $O\left(n^{1/t}\right)$ and answers of length $O(1)$.*

A generic balancing technique of Chor et al. [1995, Section 4.3] allows to convert any PIR protocol with $O(n^{1/t})$ long queries and $O(1)$ long answers into a new PIR protocol with $O(n^{1/(t+1)})$ total communication. Such a conversion yields

THEOREM 6.4. *Let $p = 2^t - 1$ be a fixed Mersenne prime. For every positive integer n there exists a three server PIR protocol with $O(n^{1/(t+1)})$ communication.*

Plugging the value of the largest known Mersenne prime $p = 2^{32,582,657} - 1$ into Theorem 6.4, we conclude

THEOREM 6.5. *For every positive integer n, there exists a three-server PIR protocol with communication complexity of $O(n^{1/32,582,658})$.*

The next two theorems capture the asymptotic parameters of our PIR schemes under the number-theoretic assumptions. Both theorems follow immediately from Lemma 6.2 using the arguments that are essentially identical to the proofs of Theorems 5.13 and 5.15.

THEOREM 6.6. *Suppose that 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 $n^{O(\frac{1}{\log\log n})}$.*

THEOREM 6.7. *Let ε be a positive constant. Suppose the conjecture* 5.14 *holds; then for all values of n, there exists a three-server PIR protocol with communication complexity of $n^{O(\frac{1}{\log^{1-\varepsilon}\log n})}$.*

7. *Conclusion*

We presented a novel approach to constructing locally decodable codes and vastly improved the known upper bounds. However, the gap between the upper and lower bounds for LDCs still remains very large. It might be the case that the technique proposed in this paper has not yet been pushed to its limit and further improvements will be obtained in this way. In particular, Proposition 5.5 can be generalized to arbitrary finite fields (rather than just prime fields), and even finite commutative rings. It may happen that a clever choice of a ring R and a subset $S \subseteq F$ that is simultaneously combinatorially and algebraically nice will yield shorter LDCs.

After the preliminary version of this article [Yekhanin 2006], was published some progress towards understanding the limits of our technique has been made in Kedlaya and Yekhanin [2007]. Specifically, it was shown that our constructions do not necessarily require Mersenne primes. It suffices to have Mersenne numbers (i.e., integers of form $2^t - 1$) with large prime factors. It was also shown that obtaining locally decodable codes of constant query complexity and length $\exp\left(n^{o(1)}\right)$ through constructing nice subsets of finite fields would imply progress on an old number theory problem, and therefore seems unlikely in the near future.

ACKNOWLEDGMENTS. I am indebted to Madhu Sudan for expressing his optimism regarding viability of the approach taken in this paper at an early stage of the work. I would also like to thank him for many helpful in-depth technical discussions. Many thanks to Oded Goldreich, Nick Harvey, Kiran Kedlaya, Swastik Kopparty, Dieter van Melkebeek and David Woodruff for their valuable comments.

REFERENCES

AMBAINIS, A. 1997. Upper bound on the communication complexity of private information retrieval. In *Proceedings of the 32nd Annual International Colloquium on Antomata, Languages and Programming*. Lecture Notes in Computer Science, vol. 1256, Springer-Verlag, New York, pp. 401–407.

BABAI, L., FORTNOW, L., LEVIN, L., AND SZEGEDY, M. 1991. Checking computations in polylogarithmic time. In *Proceedings of the 23th ACM Symposium on Theory of Computing (STOC)*. ACM, New York, pp. 21–31.

BABAI, L., AND FRANKL, P. 1998. Linear algebra methods in combinatorics.

BEIGEL, R., FORTNOW, L., AND GASARCH, W. 2006. A tight lower bound for restricted PIR protocols. *Computat. Complex 15*, 82–91.

BEIMEL, A., ISHAI, Y., AND KUSHILEVITZ, E. 2005. General constructions for information-theoretic private information retrieval. *J. Comput. Syst. Sci. 71*, 213–247, (Preliminary versions in STOC 1999 and ICALP 2001.)

BEIMEL, A., ISHAI, Y., KUSHILEVITZ, E., AND RAYMOND, J. F. 2002. Breaking the $O(n^{1/(2k-1)})$ barrier for information-theoretic private information retrieval. In *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society Press, Los Alamitos, CA, pp. 261–270.

Chor, B., Goldreich, O., Kushilevitz, E., and Sudan, M. 1995. Private information retrieval. In *Proceedings of the 36rd IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society Press, Los Alamitos, CA, pp. 41–50. (Also, in *J. ACM 45*, 1998).

DESHPANDE, A., JAIN, R., KAVITHA, T., LOKAM, S., AND RADHAKRISHNAN, J. 2002. Better lower bounds for locally decodable codes. In *Proceedings of the 20th IEEE Computational Complexity Conference (CCC)*. IEEE Computer Society Press, Los Alamitos, CA, pp. 184–193.

COOPER, C., AND BOONE, S. 2006. http://www.mersenne.org/32582657.htm.

GASARCH, W. 2004. A survey on private information retrieval. *Bull. EATCS 82*, 72–107.

GOLDREICH, O. 2005. Short locally testable codes and proofs. Tech. Rep. TR05-014. In *Proceedings of the Electronic Colloquim on Computational Complexity* (ECCC).

GOLDREICH, O., KARLOFF, H., SCHULMAN, L., AND TREVISAN, L. 2006. Lower bounds for locally decodable codes and private information retrieval. *Computat. Complex. 15*, 3, 263–296.

ITOH, T. 1999. Efficient private information retrieval. *IEICE Trans. Fund. Electron. Commun. Comput. Sci. E82-A, 1*, 11–20.

ITOH, T. 2001. On lower bounds for the communication complexity of private information retrieval. *IEICE Trans. Fund. Electron. Commun. Comput. Sci. E84-A1*, 157–164.

KEDLAYA, K., AND YEKHANIN, S. 2007. Locally decodable codes from nice subsets of finite fields and prime factors of Mersenne numbers. In *Proceedings of the Electronic Colloquium on Computational Complexity,* Tech. rep. TR07-040.

KATZ, J., AND TREVISAN, L. 2000. On the efficiency of local decoding procedures for error-correcting codes. In *Proceedings of the 32th ACM Symposium on Theory of Computing (STOC)*. ACM, New York, pp. 80–86.

KERENIDIS, I., AND DE WOLF, R. 2003. Exponential lower bound for 2-query locally decodable codes via a quantum argument. *J. Comput. Syst. Sci. 69*, 3, 395–420. (Earlier version in STOC'03. quant-ph/0208062.)

LENSTRA, JR., H. W. 1980. Primarlity testing. In *Studieweek Getaltheorie en Computers* (Sept. 1–5). Stichting Math., Centrum, Amsterdam, The Netherlands.

LIDL, R., AND NIEDERREITER, H. 1983. *Finite Fields*. Cambridge University Press, Cambridge, MA.

MANN, E. 1998. Private access to distributed information. Master's thesis, Technion - Israel Institute of Technology, Haifa, Israel.

OBATA, K. 2002. Optimal lower bounds for 2-query locally decodable linear codes. In *Proceedings of the 6th International Workshop on Randomization Techniques*: Lecture Notes in Computer Science, vol. 2483. Springer-Verlag, New York, pp. 39–50.

POLISHCHUK, A., AND SPIELMAN, D. 1994. Nearly-linear size holographic proofs. In *Proceedings of the 26th ACM Symposium on Theory of Computing (STOC)*. ACM, New York, pp. 194–203.

POMERANCE, C. 1980/81. Recent developments in primality testing. *Math. Intell. 3*, 3, 97–105.

RAZBOROV, A., AND YEKHANIN, S. 2006. An $\Omega(n^{1/3})$ lower bound for bilinear group based private information retrieval. In *Proceedings of the 47rd IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society Press, Los Alamitos, CA, 739–748.

SUDAN, M. 1992. Efficient checking of polynomials and proofs and the hardness of approximation problems. Ph.D. dissertation, University of California at Berkeley, Berkeley, CA.

TREVISAN, L. 2004. Some applications of coding theory in computational complexity. *Quad. Matemat. 13*, 347–424.

WEHNER, S., AND DE WOLF, R. 1997. Improved lower bounds for locally decodable codes and private information retrieval. In *Proceedings of 32nd International Colloquium on Automata, Languages and Programming (ICALP'05)*. Lecture Notes in Computer Science, vol. 3580, Springer-Verlag, New York, pp. 1424–1436.

WAGSTAFF, S. 1983. Divisors of Mersenne numbers. *Math. Comput. 40*, 161, 385–397.

WOODRUFF, D. 2007. New lower bounds for general locally decodable codes. In *Proceedings of the Electronic Colloquium on Computational Complexity,* Tech. rep. TR07-006.

WOODRUFF, D., AND YEKHANIN, S. 2007. A geometric approach to information theoretic private information retrieval. *SIAM J. Comput. 37*, 4, 1046–1056.

YEKHANIN, S. 2006. New locally decodable codes and private infromation retrieval schemes. In *Electronic Colloquium on Computational Complexity,* Tech. rep. TR06-127.

RECEIVED MAY 2007; REVISED NOVEMBER 2007; ACCEPTED NOVEMBER 2007