comput. complex. **22** (2013), 159 – 189

1016-3328/13/010159-31
published online August 4, 2011
DOI 10.1007/s00037-011-0017-1

computational complexity

QUERY-EFFICIENT LOCALLY DECODABLE CODES OF SUBEXPONENTIAL LENGTH

Yeow Meng Chee, Tao Feng, San Ling, Huaxiong Wang, and Liang Feng Zhang

Abstract. A k-query locally decodable code (LDC) $\mathbf{C} : \Sigma^n \to \Gamma^N$ encodes each message x into a codeword $\mathbf{C}(x)$ such that each symbol of x can be probabilistically recovered by querying only k coordinates of $\mathbf{C}(x)$, even after a constant fraction of the coordinates has been corrupted. Yekhanin (in J ACM 55:1–16, 2008) constructed a 3-query LDC of subexponential length, $N = \exp(\exp(O(\log n/ \log\log n)))$, under the assumption that there are infinitely many Mersenne primes. Efremenko (in Proceedings of the 41st annual ACM symposium on theory of computing, ACM, New York, 2009) constructed a 3-query LDC of length $N_2 = \exp(\exp(O(\sqrt{\log n \log\log n})))$ with no assumption, and a 2^r-query LDC of length $N_r = \exp(\exp(O(\sqrt[r]{\log n(\log\log n)^{r-1}})))$, for every integer $r \geq 2$. Itoh and Suzuki (in IEICE Trans Inform Syst E93-D 2:263–270, 2010) gave a composition method in Efremenko's framework and constructed a $3 \cdot 2^{r-2}$-query LDC of length N_r, for every integer $r \geq 4$, which improved the query complexity of Efremenko's LDC of the same length by a factor of 3/4. The main ingredient of Efremenko's construction is the Grolmusz construction for super-polynomial size set-systems with restricted intersections, over $\mathbb{Z}_m$, where m possesses a certain "good" algebraic property (related to the "algebraic niceness" property of Yekhanin in J ACM 55:1–16, 2008). Efremenko constructed a 3-query LDC based on $m = 511$ and left as an open problem to find other numbers that offer the same property for LDC constructions.
In this paper, we develop the algebraic theory behind the constructions of Yekhanin (in J ACM 55:1–16, 2008) and Efremenko (in Proceedings of the 41st annual ACM symposium on theory of computing, ACM, New York, 2009), in an attempt to understand the "algebraic niceness"

phenomenon in $\mathbb{Z}_m$. We show that every integer $m = pq = 2^t - 1$, where p, q, and t are prime, possesses the same good algebraic property as $m = 511$ that allows savings in query complexity. We identify 50 numbers of this form by computer search, which together with 511, are then applied to gain improvements on query complexity via Itoh and Suzuki's composition method. More precisely, we construct a $3^{\lceil r/2 \rceil}$-query LDC for every positive integer $r < 104$ and a $\lfloor (3/4)^{51} \cdot 2^r \rfloor$-query LDC for every integer $r \geq 104$, both of length N_r, improving the 2^r queries used by Efremenko (in Proceedings of the 41st annual ACM symposium on theory of computing, ACM, New York, 2009) and $3 \cdot 2^{r-2}$ queries used by Itoh and Suzuki (in IEICE Trans Inform Syst E93-D 2:263–270, 2010).
We also obtain new efficient private information retrieval (PIR) schemes from the new query-efficient LDCs.

Keywords. Locally decodable codes, Mersenne numbers, private information retrieval.

Subject classification. 20C05, 94B60.

1. Introduction

A classical error-correcting code $\mathbf{C} : \Sigma^n \to \Gamma^N$ allows one to encode a message x into a codeword $\mathbf{C}(x)$ such that x can be recovered even if $\mathbf{C}(x)$ gets corrupted in a number of coordinates. However, to recover even a small portion of the message x, one has to consider all or most of the coordinates of the received (possibly corrupted) codeword. Katz & Trevisan (2000) considered error-correcting codes where each symbol of the message can be probabilistically recovered by looking at a limited number of coordinates of a corrupted encoding. Such codes are known as *locally decodable codes* (LDCs). Informally, a (k, δ, ϵ)-LDC $\mathbf{C} : \Sigma^n \to \Gamma^N$ encodes a message x into a codeword $\mathbf{C}(x)$ such that each symbol x_i of the message can be recovered with probability at least $1 - \epsilon$, by a probabilistic decoding algorithm that makes at most k queries, even if the codeword is corrupted in up to δN locations. LDCs have many applications in cryptography and complexity theory (see, for example, Gasarch (2004); Trevisan (2004)) and have attracted a considerable amount of attention (Deshpande *et al.* 2002; Obata 2002; Kerenidis & de Wolf 2004; Dvir & Shpilka 2005; Wehner &

de Wolf 2005; Goldreich *et al.* 2006; Shiowattana & Lokam 2006; Raghavendra 2007; Woodruff 2007; Kedlaya & Yekhanin 2008; Yekhanin 2008; Efremenko 2009; Gopalan 2009; Itoh & Suzuki 2010) since their formal introduction by Katz & Trevisan (2000).

For constant δ and ϵ, the efficiency of a (k, δ, ϵ)-LDC $\mathbf{C} : \Sigma^n \to \Gamma^N$ is measured by its *length* N and *query complexity* k. Ideally, we want both N and k to be as small as possible. Katz & Trevisan (2000) proved that there do not exist families of 1-query LDCs. Goldreich *et al.* (2006) obtained an exponential lower bound of $\exp(\Omega(n))$ on the length of 2-query *linear* LDCs. Kerenidis & de Wolf (2004) showed that the optimal length of *any* 2-query LDCs is $\exp(O(n))$ via a quantum argument. For a k-query ($k \geq 3$) LDC, Woodruff (2007) obtained a super-linear lower bound of $\Omega(n^{(k+1)/(k-1)}/\log n)$ on its length. Other lower bounds have been obtained by Deshpande *et al.* (2002); Obata (2002); Dvir & Shpilka (2005); Wehner & de Wolf (2005), and Shiowattana & Lokam (2006).

It has been conjectured for a long time that the length N of any constant-query LDC should have an exponential dependence on its message length n. This conjecture was disproved by Yekhanin (2008), who constructed a 3-query LDC of length $\exp(\exp(O(\log n/\log\log n)))$ under the assumption that there are infinitely many *Mersenne primes* (primes of the form $M_t = 2^t - 1$, where t is prime). Subsequently, Yekhanin's construction was nicely reformulated by Raghavendra (2007) using group homomorphism. Inspired by this, Efremenko (2009) generalized Yekhanin's construction and established a framework for constructing LDCs in which the above assumption on Mersenne primes is no longer necessary. Efremenko (2009) constructed a k_r-query ($k_r \leq 2^r$) LDC of length $N_r = \exp(\exp(O(\sqrt[r]{\log n(\log\log n)^{r-1}})))$ for every integer $r \geq 2$, and in particular, a 3-query ($k_2 = 3$) LDC of length $N_2 = \exp(\exp(O(\sqrt{\log n \log\log n})))$ for $r = 2$. The main ingredient of Efremenko's construction is a construction of Grolmusz (2000) for super-polynomial size set-systems with restricted intersections. Each of these set-systems is over a certain composite number, which has significant impact on the query complexity (the value of k_r) of the resulting LDC. Efremenko (2009) showed that the composite number 511 can result in a 3-query LDC of length N_2 and left as an open problem to find other suitable composite numbers.

Recently, Itoh & Suzuki (2010) developed a composition method in Efremenko's framework. This method allows one to compose, in an appropriate way, Efremenko's k_r-query ($k_r \leq 2^r$) LDC of length N_r and k_l-query ($k_l \leq 2^l$) LDC of length N_l to obtain a k-query LDC of length N_{r+l} such that $k \leq k_r k_l$. For every integer $r \geq 4$, taking Efremenko's 3-query LDC and k_{r-2}-query LDC as building blocks, the composition method yields a k-query LDC of length N_r in which $k \leq 3 \cdot 2^{r-2}$, improving the query complexity of Efremenko's LDC of the same length by a factor of 3/4. We stress that this improvement is due to the first building block, that is, the 3-query LDC. Hence, it is of great interest to obtain as many such 3-query LDCs as possible, or equivalently, as many new composite numbers as possible, which can result in 3-query LDCs of length N_2 in Efremenko's construction.

1.1. Our Results. In this paper, we study the algebraic properties of *good* composite numbers, which yield 3-query LDCs in Efremenko's construction. We give a characterization of such composite numbers and show that every Mersenne number that is a product of two primes is good. Consequently, we obtain a number of good composite numbers. These new good numbers, together with 511, are then applied to achieve improvements on the query complexity in Efremenko's framework.

Let $\mathbb{M}_2$ be the set of composite numbers, each of which is the product of two distinct odd primes and good (i.e., can yield a 3-query LDC of length N_2 in Efremenko's construction). We characterize numbers in $\mathbb{M}_2$ and show that the subset of *Mersenne numbers* (numbers of the form $M_t = 2^t - 1$, where t is prime)

$$\mathbb{M}_{2,\mathrm{Mersenne}} = \{m : m = 2^t - 1 = pq, \text{ where } p, q \text{ and } t \text{ are primes}\}$$

is contained in $\mathbb{M}_2$. Note that the number $511 = 2^9 - 1 = 7 \times 73$, suggested by Efremenko (2009), is in $\mathbb{M}_2$ but not in $\mathbb{M}_{2,\mathrm{Mersenne}}$. On the other hand, the number $15 = 3 \times 5$, the smallest possible candidate for $\mathbb{M}_2$, is not in $\mathbb{M}_2$, checked via exhaustive search by Itoh & Suzuki (2010). We identify 50 numbers in $\mathbb{M}_{2,\mathrm{Mersenne}}$ and hence 50 new numbers in $\mathbb{M}_2$, which answers open problems raised by Efremenko (2009) and Itoh & Suzuki (2010). Furthermore, we show that

(a) For every integer $r, 1 \leq r \leq 103$, there is a k-query linear LDC of length N_r for which

$$k \leq \begin{cases} (\sqrt{3})^r, & \text{if } r \text{ is even} \\ 8 \cdot (\sqrt{3})^{r-3}, & \text{if } r \text{ is odd.} \end{cases}$$

(b) For every integer $r \geq 104$, there is a k-query linear LDC of length N_r for which $k \leq (3/4)^{51} \cdot 2^r$.

(c) If $|\mathbb{M}_{2,\text{Mersenne}}| = \infty$, then for every integer $r \geq 1$, there is a k-query linear LDC of length N_r for which k is the same as that in (a).

The notion of LDCs is closely related to the notion of information-theoretic private information retrieval (PIR) schemes. It is well-known that LDCs with perfectly smooth decoders imply PIR schemes, and there is a generic transformation from LDCs to PIR schemes (Katz & Trevisan 2000). As with the LDCs of Efremenko (2009) and Itoh & Suzuki (2010), the query-efficient LDCs obtained in this paper also have perfectly smooth decoders.[1] This in turn gives new PIR schemes with smaller communication complexity. For instance, the LDCs from (a) above imply PIR schemes with communication complexity $\exp(O(\sqrt[r]{\log n(\log\log n)^{r-1}}))$ for $3^{r/2}$ servers. Compared with the best known PIR schemes of Itoh & Suzuki (2010) with the same communication complexity for $3 \cdot 2^{r-2}$ servers, where $r < 104$ is even, our new schemes require fewer servers.

We are able to identify only 50 numbers in $\mathbb{M}_{2,\text{Mersenne}}$ by computer search with the largest one being $M_{7331} = 2^{7331} - 1$. We believe that the search for more numbers in $\mathbb{M}_{2,\text{Mersenne}}$ is of independent interest. In particular, it is an interesting open problem to determine how many numbers $\mathbb{M}_{2,\text{Mersenne}}$ contains. Compared with Mersenne primes, it seems reasonable to conjecture that $|\mathbb{M}_{2,\text{Mersenne}}| = \infty$.

[1] Note that the decoders for the LDCs of Yekhanin (2008) are not smooth.

1.2. Organization. This paper is organized as follows. In Section 2, we review Efremenko's framework and the composition method of Itoh & Suzuki (2010). In Section 3, we prove that all Mersenne numbers, which are products of two primes, belong to $\mathbb{M}_2$ and introduce the family $\mathbb{M}_{2,\mathrm{Mersenne}}$. We also characterize the numbers in $\mathbb{M}_2$ and discuss how to prove that a given number is not in $\mathbb{M}_2$. In Section 4, we obtain new query-efficient LDCs using the family $\mathbb{M}_{2,\mathrm{Mersenne}}$. This also gives new efficient PIR schemes with fewer servers. We conclude the paper in Section 5.

2. Preliminaries

We briefly review Efremenko's framework (2009) and the composition method of Itoh & Suzuki (2010).

Let m and h be positive integers. The ring $\mathbb{Z}/m\mathbb{Z}$ is denoted $\mathbb{Z}_m$. The set $\{1, 2, \ldots, m\}$ is denoted $[m]$. The mod m *inner product* of two vectors $x = (x_1, \ldots, x_h), y = (y_1, \ldots, y_h) \in \mathbb{Z}_m^h$ is defined to be $\langle x, y\rangle_m \equiv \sum_{i=1}^h x_i y_i \bmod m$. The *Hamming distance* between x and y is denoted $d_H(x, y)$.

Definition 2.1 (Locally Decodable Code). *Let $k, n,$ and N be positive integers, and $0 < \delta, \epsilon < 1$. A code $\mathbf{C} : \Sigma^n \to \Gamma^N$ is said to be (k, δ, ϵ)-locally decodable if there is a probabilistic decoding algorithm $\mathcal{D}$ such that*

(i) *For every $x \in \Sigma^n, i \in [n]$, and $y \in \Gamma^N$ such that $d_H(y, \mathbf{C}(x)) \leq \delta N$, we have $\Pr[\mathcal{D}^y(i) = x_i] \geq 1-\epsilon$, where $\mathcal{D}^y$ means that $\mathcal{D}$ makes oracle access to y, and the probability is taken over the internal coin tosses of $\mathcal{D}$.*

(ii) *In every invocation, $\mathcal{D}$ makes at most k queries to y.*

The algorithm $\mathcal{D}$ is called a (k, δ, ϵ)-*local decoding algorithm* for $\mathbf{C}$. Parameters k and N are called the *query complexity* and *length* of $\mathbf{C}$, respectively. The *alphabets* Σ and Γ are often taken to be a finite field $\mathbb{F}_q$, where q is a prime power. A k-query LDC $\mathbf{C} : \mathbb{F}_q^n \to \mathbb{F}_q^N$ is *linear* if it is a linear transformation and *nonadaptive* if in every invocation, $\mathcal{D}$ makes all queries simultaneously. All the LDCs in this paper are linear and nonadaptive.

2.1. Efremenko's Framework. Efremenko's framework (2009) for constructing LDCs is essentially a generalization of the work of Yekhanin (2008). Let $m = p_1 p_2 \cdots p_r$ be a product of $r \geq 2$ distinct odd primes $p_1, p_2, \ldots, p_r$. Let $S \subseteq \mathbb{Z}_m \backslash \{0\}$ and h be a positive integer. Let t be the multiplicative order of $2 \in \mathbb{Z}_m^*$, and let $\gamma_m \in \mathbb{F}_{2^t}^*$ be a primitive m-th root of unity. The building blocks of Efremenko's framework for constructing LDCs include both an *S-matching family* and an *S-decoding polynomial*, which are defined as follows:

DEFINITION 2.2 (S-Matching Family). *For $S \subseteq \mathbb{Z}_m \backslash \{0\}$, a family of vectors $\{u_i\}_{i=1}^n \subseteq \mathbb{Z}_m^h$ is called an S-matching family if:*

(i) $\langle u_i, u_i \rangle_m = 0$, *for $i \in [n]$; and*

(ii) $\langle u_i, u_j \rangle_m \in S$, *for distinct $i, j \in [n]$.*

DEFINITION 2.3 (S-Decoding Polynomial). *For $S \subseteq \mathbb{Z}_m \backslash \{0\}$, a polynomial $P(X) \in \mathbb{F}_{2^t}[X]$ is called an S-decoding polynomial if:*

(i) $P(\gamma_m^s) = 0$, *for $s \in S$; and*

(ii) $P(\gamma_m^0) = P(1) = 1$.

For any subset $S \subseteq \mathbb{Z}_m \backslash \{0\}$, an S-matching family and the corresponding S-decoding polynomial yield a linear LDC immediately.

THEOREM 2.4 (Efremenko 2009). *Let $\{u_i\}_{i=1}^n \subseteq \mathbb{Z}_m^h$ be an S-matching family and $P(X) = a_0 + a_1 X^{b_1} + \cdots + a_{k-1} X^{b_{k-1}} \in \mathbb{F}_{2^t}[X]$ be an S-decoding polynomial with k monomials. Then, there is a k-query linear LDC $\mathbf{C} : \mathbb{F}_{2^t}^n \to \mathbb{F}_{2^t}^{m^h}$ with encoding and decoding algorithms as in Figure 2.1.*

Theorem 2.4 shows that for any $S \subseteq \mathbb{Z}_m \backslash \{0\}$, an S-matching family of size n and an S-decoding polynomial with k monomials yield a k-query LDC, which encodes each message of length n into a codeword of length m^h. Once m and h are fixed, the length N is inversely proportional to n. Hence, ideally, n should be large and k small. To have a large S-matching family, the set S is usually taken to be S_m, the *canonical set* of m, which is defined as follows:

Encoding

Let $e_j \in \mathbb{F}_{2^t}^n$ denote the j-th unit vector for $j \in [n]$. The coordinates of a codeword $\mathbf{C}(x)$ are indexed by vectors in $\mathbb{Z}_m^h$, where $x \in \mathbb{F}_{2^t}^n$. The encoding algorithm works as follows:

1. for $j \in [n]$ and $v \in \mathbb{Z}_m^h$, $\mathbf{C}(e_j)_v = \gamma_m^{\langle u_j, v\rangle_m}$;
2. for $x = (x_1, \ldots, x_n) \in \mathbb{F}_{2^t}^n$, we have $\mathbf{C}(x) = \sum_{j=1}^n x_j \cdot \mathbf{C}(e_j)$.

Decoding

To recover x_i from a possibly corrupted codeword $y \in \mathbb{F}_{2^t}^{m^h}$ of any message x,we

1. choose a vector $v \in \mathbb{Z}_m^h$ uniformly and query the coordinates $y_v, y_{v+b_1 u_i}, \ldots, y_{v+b_{k-1}u_i}$;
2. output $\gamma_m^{-\langle u_i, v\rangle_m} \cdot (a_0 \cdot y_v + a_1 \cdot y_{v+b_1 u_i} + \ldots + a_{k-1} \cdot y_{v+b_{k-1}u_i})$.

Figure 2.1: Efremenko's framework for constructing LDCs.

Definition 2.5 (Canonical Set). *Let $m = p_1 p_2 \cdots p_r$ be the product of $r \geq 2$ distinct odd primes $p_1, p_2, \ldots, p_r$. The canonical set of m is defined to be*

$$S_m = \{s_\sigma \in \mathbb{Z}_m : \ \sigma \in \{0,1\}^r \setminus \{\mathbf{0}\} \textit{ and } s_\sigma \equiv \sigma_i \bmod p_i, \textit{ for } i \in [r]\}.$$

For every integer $r \geq 2$, Efremenko (2009) proved that there exists an S_m-matching family of super-polynomial size and an S_m-decoding polynomial with at most 2^r monomials.

Proposition 2.6 (Efremenko 2009). *Let $m = p_1 p_2 \cdots p_r$ be the product of $r \geq 2$ distinct odd primes $p_1, p_2, \ldots, p_r$.*

(i) *There is a positive constant c, depending only on m, such that for every integer $h > 0$, there is an S_m-matching family $\{u_i\}_{i=1}^n \subseteq \mathbb{Z}_m^h$ of size $n \geq \exp\left(c(\log h)^r/(\log\log h)^{r-1}\right)$.*

(ii) *There is an S_m-decoding polynomial with at most 2^r monomials.*

Efremenko's linear LDCs of subexponential length now immediately follow from Theorem 2.4 and Proposition 2.6.

THEOREM 2.7 (Efremenko 2009). *For every integer $r \geq 2$, there is a linear $(k_r, \delta, k_r\delta)$-LDC of length $N_r = \exp(\exp(O(\sqrt[r]{\log n (\log\log n)^{r-1}})))$ for which $k_r \leq 2^r$. In particular, when $r = 2$, there is a linear $(3, \delta, 3\delta)$-LDC of length $N_2 = \exp(\exp(O(\sqrt{\log n \log\log n})))$.*

2.2. The Composition Method. For every integer $r \geq 2$, there is a k_r-query linear LDC of subexponential length N_r by Theorem 2.7, but its query complexity k_r is only upper bounded by 2^r. It is attractive to improve the query complexity. This is the motivation for Itoh and Suzuki's composition method.

Let $m_1 = p_1p_2\cdots p_r$ be the product of r distinct odd primes $p_1, p_2, \ldots, p_r$, and $m_2 = q_1q_2\cdots q_l$, the product of l distinct odd primes $q_1, q_2, \ldots, q_l$, where $r, l \geq 2$. Suppose $\gcd(m_1, m_2) = 1$. Let $m = m_1m_2$, and t_1, t_2, and t be the multiplicative orders of 2 in $\mathbb{Z}^*_{m_1}, \mathbb{Z}^*_{m_2}$, and $\mathbb{Z}^*_m$, respectively. By Theorem 2.4 and Theorem 2.7, there are linear LDCs $\mathbf{C}_r : \mathbb{F}^n_{2^{t_1}} \to \mathbb{F}^{N_r}_{2^{t_1}}, \mathbf{C}_l : \mathbb{F}^n_{2^{t_2}} \to \mathbb{F}^{N_l}_{2^{t_2}}$ and $\mathbf{C}_{r+l} : \mathbb{F}^n_{2^t} \to \mathbb{F}^{N_{r+l}}_{2^t}$ of query complexities $k_r \leq 2^r, k_l \leq 2^l$, and $k_{r+l} \leq 2^{r+l}$, respectively. Let $P_1(X) \in \mathbb{F}_{2^{t_1}}[X]$ and $P_2(X) \in \mathbb{F}_{2^{t_2}}[X]$ be the S_{m_1}-decoding polynomial for $\mathbf{C}_r$ and S_{m_2}-decoding polynomial for $\mathbf{C}_l$, respectively. Let $\gamma_{m_1}, \gamma_{m_2}$, and γ_m be the primitive m_1-th, m_2-th, and m-th roots of unity used in the encoding algorithms of $\mathbf{C}_r, \mathbf{C}_l$, and $\mathbf{C}_{r+l}$, respectively. It is not hard to see that there are integers μ and ν such that $\gamma_{m_1} = \gamma_m^{\mu m_2}$ and $\gamma_{m_2} = \gamma_m^{\nu m_1}$. Itoh & Suzuki (2010) proved that $P(X) = P_1(X^{\mu m_2})P_2(X^{\nu m_1}) \in \mathbb{F}_{2^t}[X]$ is an S_m-decoding polynomial for $\mathbf{C}_{r+l}$. Obviously, $P(X)$ contains at most k_rk_l monomials. Hence, the composition theorem below follows.

THEOREM 2.8 (Itoh & Suzuki 2010). *With notations as above, there is a k-query linear LDC $\mathbf{C} : \mathbb{F}^n_{2^t} \to \mathbb{F}^{N_{r+l}}_{2^t}$ for which $k \leq k_rk_l$.*

Theorem 2.8 shows that Efremenko's LDC $\mathbf{C}_{r+l}$ essentially has a local decoding algorithm that makes at most k_rk_l queries. The key idea of the composition method is as follows: if we choose the building blocks $\mathbf{C}_r$ and $\mathbf{C}_l$ in such a way that either $k_r < 2^r$ or $k_l < 2^l$, then a local decoding algorithm for $\mathbf{C}_{r+l}$ that makes less than 2^{r+l} queries follows. For every integer $r \geq 4$, applying

Theorem 2.8 to Efremenko's 3-query LDC $\mathbf{C}_2$ (based on $m_1 = 511$) of length N_2 and k_{r-2}-query LDC $\mathbf{C}_{r-2}$ (based on $m_2 = q_1, \ldots, q_{r-2}$ such that $\gcd(m_1, m_2) = 1$) of length N_{r-2} gives:

COROLLARY 2.9 (Itoh & Suzuki 2010). *For every integer $r \geq 4$, there is a k-query linear LDC $\mathbf{C}$ of length N_r in which $k \leq 3 \cdot 2^{r-2}$.*

We note that Efremenko's 3-query linear LDC is crucial to the improvement provided by Corollary 2.9. The existence of this code depends on a carefully chosen good composite number $m_1 = 511$. It is natural to ask whether there are good composite numbers other than 511 based on which a 3-query linear LDC of length N_2 can be obtained from Efremenko's construction.

For every positive integer $r \geq 2$, we denote by $\mathbb{M}_r$ the set of integers, each of which is a product of r distinct odd primes and can yield a k-query linear LDC of length N_r for which $k < 2^r$ in Efremenko's construction. Efremenko (2009) showed that $511 \in \mathbb{M}_2$ and built their 3-query LDC on this number. Itoh & Suzuki (2010) proved that $15 \notin \mathbb{M}_2$ by exhaustive search. Both Efremenko (2009) and Itoh & Suzuki (2010) left as an open problem to find elements of $\mathbb{M}_2$ other than 511. We provide an answer to this problem in the next section.

We end this section with some algebra required to establish our results.

2.3. Group Rings, Characters and Cyclotomic Cosets. Let G be a finite multiplicative abelian group. The *group ring*

$$\mathbb{Z}[G] = \left\{ \sum_{g \in G} a_g g : a_g \in \mathbb{Z} \right\}$$

is a ring of formal sums, in which addition and multiplication are defined as follows:

$$A + B = \sum_{g \in G} (a_g + b_g) g,$$

$$A \cdot B = \sum_{g \in G} \sum_{h \in G} a_g b_h gh,$$

where $A = \sum_{g\in G} a_g g, B = \sum_{g\in G} b_g g \in \mathbb{Z}[G]$. The following are standard notations:

$$A^{(j)} = \sum_{g\in G} a_g g^j, \quad \forall j \in \mathbb{Z},$$
$$D = \sum_{g\in D} g, \quad \forall D \subseteq G.$$

Let $\mathbb{C}$ be the field of complex numbers and $\mathbb{C}^*$ its multiplicative group. Any group homomorphism $\chi : G \to \mathbb{C}^*$ is called a *character* of G. If $|G| = n$, then it has exactly n distinct characters. Let $\widehat{G}$ be the set of all characters of G. Then, $\widehat{G}$ is a multiplicative group in which $\chi_1\chi_2(g) = \chi_1(g)\chi_2(g)$ for all $\chi_1, \chi_2 \in \widehat{G}, g \in G$. The identity χ_0 of $\widehat{G}$, called the *principal character*, maps every $g \in G$ to $1 \in \mathbb{C}^*$. For every $\chi \in \widehat{G}$, the *order* of χ is defined to be the least positive integer l such that $\chi^l = \chi_0$. Every $\chi \in \widehat{G}$ can be easily extended to $\mathbb{Z}[G]$ linearly: $\chi(A) = \sum_{g\in G} a_g\chi(g)$. The following properties are well-known:

1. If $|G| = n < \infty$, then for any $\chi \in \widehat{G}$ and $g \in G, \chi(g)^n = 1$.

2. If $\chi \in \widehat{G}\backslash\{\chi_0\}$, then $\sum_{g\in G} \chi(g) = 0$.

3. $\chi(A^{(-1)}) = \overline{\chi(A)}$, for every $\chi \in \widehat{G}, A \in \mathbb{Z}[G]$.

Let p be a prime or prime power and $m \in \mathbb{Z}^+$ such that $\gcd(p, m) = 1$. For every $s \in \mathbb{Z}_m$, the *cyclotomic coset of p modulo m containing s* is defined to be the following set

$$E_s = \{(sp^l \bmod m) \in \mathbb{Z}_m : l = 0, 1, \ldots\},$$

where s is called *coset representative* of E_s. We always suppose that s is smallest in E_s. It is well-known that all distinct cyclotomic cosets of p modulo m form a partition of $\mathbb{Z}_m$.

The interested reader is referred to Curtis & Reiner (2006); Washington (1997); MacWilliams & Sloane (1977); McDonald (1974) for more information.

3. Mersenne Numbers which are Products of Two Primes Belong to $\mathbb{M}_2$

In this section, we answer the open problem raised by Efremenko (2009) and Itoh & Suzuki (2010) by proving that any Mersenne number, which is the product of two primes, belongs to $\mathbb{M}_2$. This result allows us to obtain a family of numbers in $\mathbb{M}_2$. Furthermore, we also give characterizations of numbers in $\mathbb{M}_2$, which turn out to be helpful for deciding whether a given number is in $\mathbb{M}_2$.

Let $m = pq$ be the product of two distinct odd primes p and q. Let t be the multiplicative order of 2 in $\mathbb{Z}_m^*$, and let $\gamma_m \in \mathbb{F}_{2^t}^*$ be a primitive m-th root of unity. Let $S_m = \{s_{11} = 1, s_{01}, s_{10}\}$ be the canonical set of m. Then, the set of S_m-decoding polynomials is

$$\mathcal{F} = \{f(X) \in \mathbb{F}_{2^t}[X] : f(\gamma_m) = f(\gamma_m^{s_{01}}) = f(\gamma_m^{s_{10}}) = 0 \text{ and } f(1) = 1\}.$$

By Lagrange interpolation, there exists $f \in \mathcal{F}$ that contains at most four monomials. On the other hand, we have the following proposition.

Proposition 3.1. *Let $m = pq$ be the product of two distinct odd primes. Then, any S_m-decoding polynomial contains at least three monomials.*

Proof. Suppose $f(X) = ax^u + bx^v \in \mathcal{F}$ is an S_m-decoding polynomial with less than three monomials. Then, $a\gamma_m^u + b\gamma_m^v = a\gamma_m^{us_{01}} + b\gamma_m^{vs_{01}} = a\gamma_m^{us_{10}} + b\gamma_m^{vs_{10}} = 0$ and $a + b = 1$. It follows that $a\gamma_m^{u-v} = a\gamma_m^{(u-v)s_{01}} = a\gamma_m^{(u-v)s_{10}} = 1 + a$. Obviously, $a \neq 0$ and therefore $\gamma_m^{u-v} = \gamma_m^{(u-v)s_{01}} = \gamma_m^{(u-v)s_{10}}$. This implies that $m \mid \gcd((u-v)(s_{01}-1), (u-v)(s_{10}-1), (u-v)(s_{10}-s_{01}))$. Since $\gcd(m, s_{10} - s_{01}) = 1$, we have $m \mid (u - v)$. Hence, $a = a\gamma_m^{u-v} = a\gamma_m^{(u-v)s_{01}} = a\gamma_m^{(u-v)s_{10}} = 1 + a$, which is a contradiction. □

Proposition 3.1 shows that for $m = pq$, the best we can expect is to have an S_m-decoding polynomial with exactly three monomials. Let

$$\mathcal{G} = \{g(X) \in \mathbb{F}_{2^t}[X] : \; g(\gamma_m) = g(\gamma_m^{s_{01}}) = g(\gamma_m^{s_{10}}) = 0 \text{ and } g(1) \neq 0\}.$$

Then, we have the following result.

Table 3.1: New elements m determined to be in $\mathbb{M}_2$

m	$M_{11} = 2^{11} - 1 = 2047$	$M_{23} = 2^{23} - 1 = 8388607$
$\mathbb{F}_{2^t}$	$\mathbb{F}_{2^{11}} = \mathbb{F}_2[\gamma]/(\gamma^{11} + \gamma^2 + 1)$	$\mathbb{F}_{2^{23}} = \mathbb{F}_2[\gamma]/(\gamma^{23} + \gamma^5 + 1)$
S_m	$\{s_{11} = 1, s_{01} = 713, s_{10} = 1335\}$	$\{s_{11} = 1, s_{01} = 5711393, s_{10} = 2677215\}$
$f(X)$	$\gamma^{1485}X^{29} + \gamma^{694}X^{27} + \gamma^{118}$	$\gamma^{6526329}X^{3526} + \gamma^{7574532}X^{3363} + \gamma^{2861754}$

PROPOSITION 3.2. *There is an S_m-decoding polynomial $f \in \mathcal{F}$ with three monomials if and only if there is a polynomial $g \in \mathcal{G}$ with three monomials.*

PROOF. The forward implication is trivial, since $\mathcal{F} \subseteq \mathcal{G}$. Let $g \in \mathcal{G}$ have exactly three monomials. Then, $f(X) = g(X)/g(1) \in \mathcal{F}$ contains the same number of monomials as $g(X)$, namely three. □

By Proposition 3.2, finding an S_m-decoding polynomial with exactly three monomials is equivalent to finding a polynomial $g(X) \in \mathcal{G}$ with exactly three monomials. Let $g(X) \in \mathcal{G}$ be such a polynomial. Since $\mathcal{G}$ is closed under multiplication by elements of $\mathbb{F}_{2^t} \backslash \{0\}$, we may suppose, without loss of generality, that $g(X) = X^u + aX^v + b \in \mathbb{F}_{2^t}[X]$ for some distinct $u, v \in \mathbb{Z}_m \backslash \{0\}$ (only $g(1), g(\gamma_m), g(\gamma_m^{s_{01}})$ and $g(\gamma_m^{s_{10}})$ are concerned) and $a, b \in \mathbb{F}_{2^t} \backslash \{0\}$. By the definition of $\mathcal{G}$, the following conditions hold simultaneously:

$$\begin{pmatrix} \gamma_m^{us_{01}} & \gamma_m^{vs_{01}} & 1 \\ \gamma_m^{us_{10}} & \gamma_m^{vs_{10}} & 1 \\ \gamma_m^{u} & \gamma_m^{v} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{3.3}$$

$$1 + a + b \neq 0. \tag{3.4}$$

Conditions (3.3) and (3.4) shed much light on how to determine elements of $\mathbb{M}_2$. A computer search based on these conditions shows that the Mersenne numbers $M_{11} = 2^{11} - 1 = 2047$ and $M_{23} = 2^{23} - 1 = 8388607$ both belong to $\mathbb{M}_2$ (see Table 3.1 for the corresponding S_m-decoding polynomials).

Theorem 2.8 shows that the more numbers in $\mathbb{M}_2$ we find, the more improvements we get on the query complexity within Efremenko's framework. This motivates the consideration of numbers taking the form of M_{11} and M_{23} and to understand why they yield better local decoding algorithms within Efremenko's framework. We note that M_{11} and M_{23} are both Mersenne numbers and each a product of two primes. This begs the question: do all numbers of this form belong to $\mathbb{M}_2$ and do they intrinsically yield better local decoding algorithms in Efremenko's framework? For the remaining of this section, we provide an affirmative answer to this question. More precisely, we prove the following theorem.

THEOREM 3.5. *Let $m = 2^t - 1 = pq$ be a Mersenne number, where t, p, and q are primes. Then, $m \in \mathbb{M}_2$.*

The proof of Theorem 3.5 is based on analysis of conditions (3.3) and (3.4) and is an easy consequence of Proposition 3.6 and Proposition 3.10 below.

PROPOSITION 3.6. *Let $m = pq$ be the product of two distinct odd primes p and q. Let t be the multiplicative order of $2 \in \mathbb{Z}_m^*$, and let $\gamma_m \in \mathbb{F}_{2^t}^*$ be a primitive m-th root of unity. Define*

$$\mathcal{Z} = \left\{ \frac{z_1 + z_2}{z_1 z_2 + z_2} : \ z_1, z_2 \in \mathbb{F}_{2^t}^*, \mathrm{ord}(z_1) = p, \text{ and } \mathrm{ord}(z_2) = q \right\}. \tag{3.7}$$

If $\mathcal{Z}$ is a multiset containing an element of multiplicity greater than one, then $m \in \mathbb{M}_2$.

PROOF. Suppose $\mathcal{Z}$ contains an element of multiplicity greater than one. Then, there exists $z_1, z_2, z_1', z_2' \in \mathbb{F}_{2^t}^*$ such that the following hold:

(i) $\mathrm{ord}(z_1) = \mathrm{ord}(z_1') = p$,

(ii) $\mathrm{ord}(z_2) = \mathrm{ord}(z_2') = q$,

(iii) $(z_1, z_2) \neq (z_1', z_2')$,

(iv) $\dfrac{z_1 + z_2}{z_1 z_2 + z_2} = \dfrac{z_1' + z_2'}{z_1' z_2' + z_2'}$.

Obviously, we have $\mathrm{ord}(\gamma_m^{s_{10}}) = p$ and $\mathrm{ord}(\gamma_m^{s_{01}}) = q$. It follows that there are integers $u_1, v_1 \in \mathbb{Z}_p \backslash \{0\}$ and $u_2, v_2 \in \mathbb{Z}_q \backslash \{0\}$ such that the following hold:

(v) $z_1 = (\gamma_m^{s_{10}})^{u_1} = \gamma_m^{u_1 s_{10}}$,

(vi) $z_2 = \gamma_m^{u_2 s_{01}}$,

(vii) $z'_1 = \gamma_m^{v_1 s_{10}}$,

(viii) $z'_2 = \gamma_m^{v_2 s_{01}}$.

Since p and q are distinct primes, the Chinese remainder theorem implies that there are unique numbers $u, v \in \mathbb{Z}_m \setminus \{0\}$ such that

(ix) $u \equiv u_1 \bmod p$ and $u \equiv u_2 \bmod q$,

(x) $v \equiv v_1 \bmod p$ and $v \equiv v_2 \bmod q$.

Combing the set of conditions (i)–(x), it is easy to verify that the numbers $u, v \in \mathbb{Z}_m \setminus \{0\}$ satisfy the following conditions

(xi) $z_1 = \gamma_m^{us_{10}}, z_2 = \gamma_m^{us_{01}}, z'_1 = \gamma_m^{vs_{10}}$, and $z'_2 = \gamma_m^{vs_{01}}$,

(xii) $u \neq v$,

(xiii) $\dfrac{\gamma_m^u + \gamma_m^{us_{01}}}{\gamma_m^u + \gamma_m^{us_{10}}} = \dfrac{\gamma_m^v + \gamma_m^{vs_{01}}}{\gamma_m^v + \gamma_m^{vs_{10}}}$.

The last condition (xiii) implies that the matrix

$$\Gamma_{u,v} = \begin{pmatrix} \gamma_m^{us_{01}} & \gamma_m^{vs_{01}} & 1 \\ \gamma_m^{us_{10}} & \gamma_m^{vs_{10}} & 1 \\ \gamma_m^u & \gamma_m^v & 1 \end{pmatrix} \tag{3.8}$$

has determinant zero. It follows that $\mathrm{rank}(\Gamma_{u,v}) = 1$ or 2. If $\mathrm{rank}(\Gamma_{u,v}) = 1$, then the rank of

$$\begin{pmatrix} \gamma_m^{us_{01}} + \gamma_m^u & \gamma_m^{vs_{01}} + \gamma_m^v & 0 \\ \gamma_m^{us_{10}} + \gamma_m^u & \gamma_m^{vs_{10}} + \gamma_m^v & 0 \\ \gamma_m^u & \gamma_m^v & 1 \end{pmatrix}$$

is also 1. Hence, $\gamma_m^{us_{01}} + \gamma_m^u = \gamma_m^{vs_{01}} + \gamma_m^v = \gamma_m^{us_{10}} + \gamma_m^u = \gamma_m^{vs_{10}} + \gamma_m^v = 0$, which in turn implies $\gamma_m^{us_{01}} = \gamma_m^{us_{10}}$ and $\gamma_m^{vs_{01}} = \gamma_m^{vs_{10}}$. Since γ_m is of order m and $\gcd(m, s_{01} - s_{10}) = 1$, we have $m | \gcd(u(s_{01} - s_{10}), v(s_{01} - s_{10}))$ and therefore $m | \gcd(u, v)$, which contradicts the fact that $u, v \in \mathbb{Z}_m \backslash \{0\}$. Consequently, $\operatorname{rank}(\Gamma_{u,v}) = 2$ and the Eq. (3.3) has a unique solution $(a, b) \in \mathbb{F}_{2^t}^2$.

Next, we show that both a and b are nonzero. If $a = 0$, then $b = \gamma_m^{us_{01}} = \gamma_m^{us_{10}} = \gamma_m^u$, which implies that $u \equiv 0 \bmod m$. If $b = 0$, then $a = \gamma_m^{(u-v)s_{01}} = \gamma_m^{(u-v)s_{10}} = \gamma_m^{u-v}$, which implies that $u \equiv v \bmod m$. Both cases yield contradictions, since $u, v \in \mathbb{Z}_m \backslash \{0\}$ are distinct.

Let $g(X) = X^u + aX^v + b \in \mathbb{F}_{2^t}[X]$. Then, $g(X)$ contains three monomials since $u, v \in \mathbb{Z}_m \setminus \{0\}$ are distinct and $a, b \in \mathbb{F}_{2^t} \setminus \{0\}$. Furthermore, we have $g(\gamma_m) = g(\gamma_m^{s_{01}}) = g(\gamma_m^{s_{10}}) = 0$ since (a, b) satisfies (3.3).

As the last step, we claim that $g(1) \neq 0$, for otherwise the vector $(1, 1, 1)$ is necessarily a linear combination of the rows of $\Gamma_{u,v}$, since $(1, a, b) \neq (0, 0, 0)$, and thereby

$$\begin{pmatrix} \gamma_m^{us_{01}} & \gamma_m^{vs_{01}} & 1 \\ \gamma_m^{us_{10}} & \gamma_m^{vs_{10}} & 1 \\ \gamma_m^u & \gamma_m^v & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

has rank two. Applying elementary row operations (adding the third row to each of the first three rows) to the above matrix gives

$$\frac{1 + \gamma_m^u}{1 + \gamma_m^v} = \frac{1 + \gamma_m^{us_{10}}}{1 + \gamma_m^{vs_{10}}} = \frac{1 + \gamma_m^{us_{01}}}{1 + \gamma_m^{vs_{01}}}. \tag{3.9}$$

Condition (xiii) and (3.9) now jointly yield $\gamma_m^{(u-v)s_{01}} = \gamma_m^{(u-v)s_{10}}$, which in turn implies that $u = v$. This is a contradiction.

We have actually shown that $g(X) \in \mathcal{G}$ and contains exactly three monomials. By Proposition 3.2, there is an S_m-decoding polynomial $f(X) \in \mathcal{F}$, which also contains exactly three monomials. Hence, $m \in \mathbb{M}_2$. $\square$

Proposition 3.10. *Let $m = 2^t - 1 = pq$ be a Mersenne number, where t, p, and q are all primes, $p \neq q$. Then, $\mathcal{Z}$ (as defined in Proposition 3.6) is a multiset containing an element of multiplicity greater than one.*

PROOF. Obviously, $\mathcal{Z}$ has at most $(p-1)(q-1)$ distinct elements. Suppose $\mathcal{Z}$ is a set of cardinality $(p-1)(q-1)$. For every $z_1, z_2 \in \mathbb{F}_{2^t}^*$ such that $\mathrm{ord}(z_1) = p$ and $\mathrm{ord}(z_2) = q$, we have $(z_1 + z_2)/(z_1 z_2 + z_2) = 1 + (1 + z_2^{-1})/(1 + z_1^{-1})$. Hence,

$$S = \{(1 + z_2)/(1 + z_1) : z_1, z_2 \in \mathbb{F}_{2^t}^*, \mathrm{ord}(z_1) = p, \text{ and } \mathrm{ord}(z_2) = q\} \tag{3.11}$$

is also a set of cardinality $(p-1)(q-1)$. Let $G = \mathbb{F}_{2^t}^*$ and 1_G its identity. Consider the group ring $\mathbb{Z}[G]$. We identify the two subsets of G,

$$A = \{1 + z_1 : z_1 \in \mathbb{F}_{2^t}^* \text{ and } \mathrm{ord}(z_1) = p\}, \tag{3.12}$$

$$B = \{1 + z_2 : z_2 \in \mathbb{F}_{2^t}^* \text{ and } \mathrm{ord}(z_2) = q\}, \tag{3.13}$$

with two elements of $\mathbb{Z}[G]$.

We claim that

$$S \cup A^{(-1)} \cup B \cup \{1_G\} = G. \tag{3.14}$$

Indeed, since $S \cup A^{(-1)} \cup B \cup \{1_G\} \subseteq G$ and $|S| + |A^{-1}| + |B| + |\{1_G\}| = |G|$, it suffices to show that $S, A^{(-1)}, B$, and $\{1_G\}$ are pairwise disjoint. It is obvious that $1_G \notin S \cup A^{(-1)} \cup B$. If $S \cap A^{(-1)} \neq \varnothing$, then there exists $z_1, z_1', z_2 \in \mathbb{F}_{2^t}^*$ such that $(1 + z_2)/(1 + z_1) = 1/(1 + z_1')$, where $\mathrm{ord}(z_1) = \mathrm{ord}(z_1') = p$ and $\mathrm{ord}(z_2) = q$. It follows that $(1 + z_2^2)/(1 + z_1) = (1 + z_2)/(1 + z_1')$, which contradicts our assumption that S is a set of cardinality $(p-1)(q-1)$. Similarly, we have $S \cap B = A^{(-1)} \cap B = \varnothing$.

From (3.14), we derive

$$(A + 1_G)^{(-1)}(B + 1_G) = G. \tag{3.15}$$

Let $\gamma_p, \gamma_q \in G$ be some primitive p-th and q-th roots of unity, respectively. We claim that there exists a permutation $a : \mathbb{Z}_p^* \to \mathbb{Z}_p^*$ and a mapping $b : \mathbb{Z}_p^* \to \mathbb{Z}_q$ such that for every $i \in \mathbb{Z}_p^*$,

$$1 + \gamma_p^i = \gamma_p^{a(i)} \gamma_q^{b(i)}. \tag{3.16}$$

Let $\theta_p, \theta_q \in \mathbb{C}$ be some complex primitive p-th and q-th roots of unity, respectively, where $\mathbb{C}$ is the field of complex numbers. Let χ_p

be a multiplicative character of order p of the group G, such that $\chi_p(\gamma_p) = \theta_p$. The identity $\chi_p((A+1_G)^{(-1)})\chi_p(B+1_G) = \chi_p(G) = 0$ implies that either $\chi_p((A+1_G)^{(-1)}) = 0$ or $\chi_p(B+1_G) = 0$. If $\chi_p(B+1_G) = 0$, then $q \equiv \chi_p(B+1_G) \equiv 0 \bmod (1-\theta_p)$ and therefore $q \in (1-\theta_p)\mathbb{Z}[\theta_p]$. On the other hand, $p = \Pi_{i=1}^{p-1}(1-\theta_p^i) \in (1-\theta_p)\mathbb{Z}[\theta_p]$. Since $\gcd(p,q) = 1$, there are rational integers α, β such that $\alpha p + \beta q = 1$. It follows that $1 \in (1-\theta_p)\mathbb{Z}[\theta_p]$, which contradicts the well-known fact that $(1-\theta_p)\mathbb{Z}[\theta_p]$ is a prime ideal in $\mathbb{Z}[\theta_p]$ (cf. Washington (1997, Lemma 1.4)). Hence, we have $\chi_p((A+1_G)^{(-1)}) = 0$ and $\chi_p(A+1_G) = \overline{\chi_p((A+1_G)^{(-1)})} = 0$, giving $\sum_{i=1}^{p-1}\chi_p(1+\gamma_p^i) + 1 = 0$. Clearly, there is a mapping $a : \mathbb{Z}_p^* \to \mathbb{Z}_p$ such that $\chi_p(1+\gamma_p^i) = \theta_p^{a(i)}$ for all $i \in \mathbb{Z}_p^*$. Hence, $\sum_{i=1}^{p-1}\theta_p^{a(i)} + 1 = 0$. Since any $p-1$ elements of $\{1, \theta_p, \ldots, \theta_p^{p-1}\}$ form an integral basis of $\mathbb{Z}[\theta_p]$ over $\mathbb{Z}, a$ must be a permutation of $\mathbb{Z}_p^*$. Since $G = \{\gamma_p^\alpha\gamma_q^\beta : \alpha \in \mathbb{Z}_p, \beta \in \mathbb{Z}_q\}$, there are two mappings $\alpha : \mathbb{Z}_p^* \to \mathbb{Z}_p$ and $\beta : \mathbb{Z}_p^* \to \mathbb{Z}_q$ such that $1+\gamma_p^i = \gamma_p^{\alpha(i)}\gamma_q^{\beta(i)}$ for all $i \in \mathbb{Z}_p^*$. It follows that $\theta_p^{a(i)} = \chi_p(1+\gamma_p^i) = \chi_p(\gamma_p^{\alpha(i)})\chi_p(\gamma_q^{\beta(i)}) = \theta_p^{\alpha(i)}\chi_p(\gamma_q)^{\beta(i)}$. Obviously, $\chi_p(\gamma_q)^p = \chi_p(\gamma_q)^q = 1$ and so $\chi_p(\gamma_q) = 1$. Therefore, $\theta_p^{a(i)} = \theta_p^{\alpha(i)}$, which implies $\alpha = a$. We identify β with b and obtain (3.16).

Similarly, there exists a permutation $c : \mathbb{Z}_q^* \to \mathbb{Z}_q^*$ and a mapping $d : \mathbb{Z}_q^* \to \mathbb{Z}_p$ such that, for every $j \in \mathbb{Z}_q^*$,

$$1+\gamma_q^j = \gamma_q^{c(j)}\gamma_p^{d(j)}. \tag{3.17}$$

Let χ_m be a multiplicative character of order m of G. Without loss of generality, we suppose that $\chi_m(\gamma_p) = \theta_p$ and $\chi_m(\gamma_q) = \theta_q$. Applying χ_m to (3.15), we have $\chi_m((A+1_G)^{(-1)})\chi_m(B+1_G) = \chi_m(G) = 0$, which implies either $\chi_m(A+1_G) = 0$ or $\chi_m(B+1_G) = 0$. If $\chi_m(A+1_G) = 0$, then $0 = \sum_{i=1}^{p-1}\chi_m(1+\gamma_p^i)+1 = \sum_{i=1}^{p-1}\theta_p^{a(i)}\theta_q^{b(i)} + 1 = \sum_{i=1}^{p-1}\theta_p^{a(i)}(\theta_q^{b(i)}-1)$. Since $\{\theta_p, \ldots, \theta_p^{p-1}\}$ is an integral basis of $\mathbb{Z}[\theta_p, \theta_q]$ over $\mathbb{Z}[\theta_q]$, we have $\theta_q^{b(i)} - 1 = 0$ for every $i \in \mathbb{Z}_p^*$. It follows that $1+\gamma_p^i = \gamma_p^{a(i)}$ for every $i \in \mathbb{Z}_p^*$. Hence, $\{0, 1, \gamma_p, \ldots, \gamma_p^{p-1}\}$ is a subfield of $\mathbb{F}_{2^t}$. However, the only subfields of $\mathbb{F}_{2^t}$ are $\mathbb{F}_2$ and $\mathbb{F}_{2^t}$. Hence, either $p+1 = 2$ or $p+1 = 2^t$, that is, either $p = 1$ or $q = 1$, which is a contradiction.

Similarly, if $\chi_m(B+1_G) = 0$, then we conclude that $\{0, 1, \gamma_q, \ldots, \gamma_q^{q-1}\}$ is a subfield of $\mathbb{F}_{2^t}$, which yields the same contradiction.

Hence, our assumption that $\mathcal{Z}$ is a set of cardinality $(p-1)(q-1)$ is wrong and the proposition is established. $\square$

We are now ready to proof Theorem 3.5.

Proof of Theorem 3.5. To apply Proposition 3.6 and Proposition 3.10, we need to show that p and q are odd and distinct. Since $pq = m = 2^t - 1$ is odd, it suffices to show that p and q are distinct. Suppose $p = q$, then $pq \equiv p^2 \equiv 1 \bmod 4$ and $pq \equiv m \equiv 2^t - 1 \equiv -1 \bmod 4$, which is a contradiction. $\square$

Theorem 3.5 provides a general method of obtaining new numbers in $\mathbb{M}_2$ and motivates the following definition of a subset of $\mathbb{M}_2$:

$$\mathbb{M}_{2,\text{Mersenne}} = \left\{ m : m = 2^t - 1 = pq, \text{ where } t, p, \text{ and } q \text{ are primes} \right\}.$$

It is an interesting open problem to determine the cardinality of $\mathbb{M}_{2,\text{Mersenne}}$. A similar but much more well-known problem in number theory is determining the number of Mersenne primes. Although it is generally believed that there are infinitely many Mersenne primes, no proof or disproof is known. It seems that our question on the cardinality of $\mathbb{M}_{2,\text{Mersenne}}$ is also difficult to answer. We have, however, determined 50 elements of $\mathbb{M}_{2,\text{Mersenne}}$ by computer search. These fifty numbers $M_t = 2^t - 1 = pq \in \mathbb{M}_{2,\text{Mersenne}}$ with their smaller prime divisors p are listed in Table 3.2. The first 33 numbers in $\mathbb{M}_{2,\text{Mersenne}}$ are $M_{11}, M_{23}, \ldots, M_{809}$. However, we do not know whether M_{881} is the 34th number in $\mathbb{M}_{2,\text{Mersenne}}$ or not.

We summarize our results below.

Proposition 3.18. $|\mathbb{M}_{2,\text{Mersenne}}| \geq 50$.

It seems reasonable to conjecture that $|\mathbb{M}_{2,\text{Mersenne}}| = \infty$.

The set $\mathbb{M}_{2,\text{Mersenne}}$ does enable us to improve query complexity in Efremenko's framework through Itoh and Suzuki's composition method (Theorem 2.8). However, to apply this method, we have to make sure that the elements of $\mathbb{M}_{2,\text{Mersenne}}$ are pairwise relatively prime.

Table 3.2: Fifty elements in $\mathbb{M}_{2,\mathrm{Mersenne}}$

m	p	m	p
M_{11}	23	M_{373}	25569151
M_{23}	47	M_{379}	180818808679
M_{37}	223	M_{421}	614002928307599
M_{41}	13367	M_{457}	150327409
M_{59}	179951	M_{487}	4871
M_{67}	193707721	M_{523}	160188778313202118610543685368878688932828701136501444932217468039063
M_{83}	167	M_{727}	17606291711815434037934881872331611670777491166445300472749449436575622328171096762265466521858927
M_{97}	11447	M_{809}	4148386731260605647525186547488842396461625774241327567978137
M_{101}	7432339208719	M_{881}	26431
M_{103}	2550183799	M_{971}	23917104973173909566916321016011885041962486321502513
M_{109}	745988807	M_{983}	1808226257914551209964473260866417929207023
M_{131}	263	M_{997}	167560816514084819488737767976263150405095191554732902607
M_{137}	32032215596496435569	M_{1063}	1485761479
M_{139}	5625767248687	M_{1427}	19054580564725546974193126830978590503
M_{149}	86656268566282183151	M_{1487}	24464753918382797416777
M_{167}	2349023	M_{1637}	81679753
M_{197}	7487	M_{2927}	1217183584262023230020873
M_{199}	164504919713	M_{3079}	25324846649810648887383180721
M_{227}	26986333437777017	M_{3259}	21926805872270062496819221124452121
M_{241}	22000409	M_{3359}	6719
M_{269}	13822297	M_{4243}	101833
M_{271}	15242475217	M_{4729}	61944189981415866671112479477273
M_{281}	80929	M_{5689}	919724609777
M_{293}	40122362455616221971122353	M_{6043}	11155520642419038056369903183
M_{347}	14143189112952632419639	M_{7331}	458072843161

Proposition 3.19. (a) *Any two distinct elements in* $\mathbb{M}_{2,\mathrm{Mersenne}}$ *are relatively prime.*

(b) *Elements in* $\mathbb{M}_{2,\mathrm{Mersenne}}$ *are relatively prime to* 511.

Proof. (a) Let $M_t = 2^t - 1 = pq \in \mathbb{M}_{2,\mathrm{Mersenne}}$ and let t_1 and t_2 be the multiplicative orders of 2 in $\mathbb{Z}_p^*$ and $\mathbb{Z}_q^*$, respectively. Then, $t_1|t$ and $t_2|t$, which in turn implies $t_1 = t_2 = t$ since t is prime and $t_1, t_2 > 1$. Suppose, there are two distinct numbers $M_t, M_{t'} \in \mathbb{M}_{2,\mathrm{Mersenne}}$ such that $\gcd(M_t, M_{t'}) > 1$. Then, M_t and $M_{t'}$ have a common prime factor, say p. It follows that $t = t' = \mathrm{ord}_p(2)$, the multiplicative order of $2 \in \mathbb{Z}_p^*$. Hence, we have $M_t = M_{t'}$, which is a contradiction.

(b) Suppose that $M_t = 2^t - 1 \in \mathbb{M}_{2,\mathrm{Mersenne}}$ is such that $\gcd(M_t, 511) > 1$. Then, either $7|M_t$ or $73|M_t$. The multiplicative orders of 2 in $\mathbb{Z}_7^*$ and $\mathbb{Z}_{73}^*$ are 3 and 9, respectively. Hence, $3|t$ or $9|t$. However, t is prime and greater than 9, which yields a contradiction. □

The result below follows from Proposition 3.18 and Proposition 3.19.

Corollary 3.20. *There are at least 51 elements in* $\mathbb{M}_2$*, which are pairwise relatively prime.*

Although Theorem 3.5 provides a rather general method of finding new elements in $\mathbb{M}_2$ (since $\mathbb{M}_{2,\mathrm{Mersenne}} \subset \mathbb{M}_2$), it does not provide a way for disproving membership in $\mathbb{M}_2$ that is easier than exhaustive search. Itoh & Suzuki (2010) showed that $15 \notin \mathbb{M}_2$ by exhaustive search. The next result shows that it is possible to avoid exhaustive search in proving that $15 \notin \mathbb{M}_2$.

Proposition 3.21. *Let* p, q, m, t, γ_m*, and* $\mathcal{Z}$ *be as defined in Proposition 3.6. Then,* $m \in \mathbb{M}_2$ *if and only if there are cyclotomic cosets* E_α *and* E_β *of 2 modulo* m $(\alpha, \beta \in \mathbb{Z}_m)$ *such that* $E_\alpha \cup E_\beta$ *does not contain any multiples of* p *or* q *and nonnegative integers* $c, d < t$ *such that*

$$(\alpha, c) \neq (\beta, d), \tag{3.22}$$

$$\left(\frac{\gamma_m^{\alpha} + \gamma_m^{\alpha s_{01}}}{\gamma_m^{\alpha} + \gamma_m^{\alpha s_{10}}}\right)^{2^c} = \left(\frac{\gamma_m^{\beta} + \gamma_m^{\beta s_{01}}}{\gamma_m^{\beta} + \gamma_m^{\beta s_{10}}}\right)^{2^d}. \tag{3.23}$$

Proof. Suppose $m \in \mathbb{M}_2$. By Proposition 3.1, there is an S_m-decoding polynomial $f(X) \in \mathcal{F}$ with exactly three monomials. By Proposition 3.2, there is a $g(X) \in \mathcal{G}$ with exactly three monomials. Without loss of generality, let $u, v \in \mathbb{Z}_m \backslash \{0\}$ be distinct and $a, b \in \mathbb{F}_{2^t} \backslash \{0\}$ be such that $g(X) = X^u + aX^v + b \in \mathbb{F}_{2^t}[X]$. It follows that (3.3) and (3.4) hold, and therefore $\det(\Gamma_{u,v}) = 0$, which in turn implies the following identity

$$(\gamma_m^u + \gamma_m^{us_{01}})(\gamma_m^v + \gamma_m^{vs_{10}}) = (\gamma_m^u + \gamma_m^{us_{10}})(\gamma_m^v + \gamma_m^{vs_{01}}). \tag{3.24}$$

Since all cyclotomic cosets of 2 modulo m form a partition of $\mathbb{Z}_m$, there exists $\alpha, \beta \in \mathbb{Z}_m$ such that $u \in E_\alpha$ and $v \in E_\beta$, where E_α and E_β are cyclotomic cosets of 2 modulo m with representatives α and β, respectively.

Suppose that $hp \in E_\alpha$ for some integer h. Then, $q \nmid h$, for otherwise $\alpha = 0$ and therefore $u = 0$, which is a contradiction. Since $u \in E_\alpha$, there is an integer l such that $u \equiv 2^l hp \bmod m$. It follows that $\gamma_m^u + \gamma_m^{us_{01}} = (\gamma_m^{hp} + \gamma_m^{hps_{01}})^{2^l} = 0$ since $hps_{01} \equiv hp \bmod m$. By identity (3.24), we have $(\gamma_m^u + \gamma_m^{us_{10}})(\gamma_m^v + \gamma_m^{vs_{01}}) = 0$. Since $hps_{10} \neq hp \bmod m$, we have $\gamma_m^u + \gamma_m^{us_{10}} = (\gamma_m^{hp} + \gamma_m^{hps_{10}})^{2^l} \neq 0$, which in turn implies that $\gamma_m^v + \gamma_m^{vs_{01}} = 0$ and therefore $p|v$. Thus, $\gamma_m^{us_{10}} = \gamma_m^{2^l hps_{10}} = (\gamma_m^{hps_{10}})^{2^l} = 1$ and $\gamma_m^{vs_{10}} = (\gamma_m^{ps_{10}})^{v/p} = 1$. In other words, the second row of $\Gamma_{u,v}$ is $(1, 1, 1)$, which implies $1 + a + b = 0$ by (3.3), contradicting (3.4). Hence, E_α does not contain any multiples of p. Similarly, E_α does not contain any multiples of q and E_β does not contain any multiples of p or q.

For $u \in E_\alpha$ and $v \in E_\beta$, there exists nonnegative integers $c, d < t$ such that $u \equiv 2^c\alpha \bmod m$ and $v \equiv 2^d\beta \bmod m$. The fact that $u \neq v$ implies $(\alpha, c) \neq (\beta, d)$. Let $u = 2^c\alpha$ and $v = 2^d\beta$ in (3.24). Then, (3.23) follows.

It remains to show that the converse is also true. Let $u \equiv 2^c\alpha \bmod m$ and $v \equiv 2^d\beta \bmod m$. Then, $u, v \in \mathbb{Z}_m$ are nonzero and distinct. Let $z_1 = \gamma_m^{us_{10}}, z_2 = \gamma_m^{us_{01}}, z_1' = \gamma_m^{vs_{10}}$, and $z_2' = \gamma_m^{vs_{01}}$. Then, it is easy to verify that $\operatorname{ord}(z_1) = \operatorname{ord}(z_1') = p, \operatorname{ord}(z_2) = \operatorname{ord}(z_2') = q$, and $(z_1, z_2) \neq (z_1', z_2')$. Then, (3.23) implies

$$(3.25) \qquad (z_1 + z_2)/(z_1 z_2 + z_2) = (z_1' + z_2')/(z_1' z_2' + z_2').$$

Note that (3.25) shows that $\mathcal{Z}$ is a multiset that contains an element of multiplicity greater than one. By Proposition 3.6, we have $m \in \mathbb{M}_2$, which completes the proof. □

Proposition 3.21 provides a rough characterization of elements in $\mathbb{M}_2$. However, it turns out to be helpful for proving that some integers are not in $\mathbb{M}_2$. In particular, we obtain a computer-free proof of the following result of Itoh & Suzuki (2010).

COROLLARY 3.26. $15 \notin \mathbb{M}_2$.

PROOF. The multiplicative order of $2 \in \mathbb{Z}_{15}^*$ is $t = 4$ and $S_{15} = \{1, 6, 10\}$. Let $\mathbb{F}_{2^4} = \mathbb{F}_2[\gamma]/(\gamma^4 + \gamma + 1)$ and let γ be a primitive 15th root of unity. The cyclotomic cosets of 2 modulo 15 are $E_0 = \{0\}, E_1 = \{1, 2, 4, 8\}, E_3 = \{3, 6, 9, 12\}, E_5 = \{5, 10\}$, and $E_7 = \{7, 14, 13, 11\}$. If $15 \in \mathbb{M}_2$, then by Proposition 3.21, there are cyclotomic cosets E_α and E_β such that $E_\alpha \cup E_\beta$ does not contain any multiples of three or five and nonnegative integers $c, d < 4$ such that (3.22) and (3.23) hold. It follows that $\{\alpha, \beta\} \subseteq \{1, 7\}$.

If $\alpha = \beta = 1$, then $((\gamma+\gamma^6)/(\gamma+\gamma^{10}))^{2^c} = ((\gamma+\gamma^6)/(\gamma+\gamma^{10}))^{2^d}$ by (3.23), that is, $\gamma^{3\cdot 2^c} = \gamma^{3\cdot 2^d}$. It follows that $c = d$ and therefore $(\alpha, c) = (\beta, d)$, which is a contradiction.

If $\alpha = \beta = 7$, then $((\gamma^7 + \gamma^{42})/(\gamma^7 + \gamma^{70}))^{2^c} = ((\gamma^7 + \gamma^{42})/(\gamma^7 + \gamma^{70}))^{2^d}$ by (3.23), that is, $\gamma^{11\cdot 2^c} = \gamma^{11\cdot 2^d}$. It follows that $c = d$ and thereby $(\alpha, c) = (\beta, d)$, which is a contradiction.

If $\{\alpha, \beta\} = \{1, 7\}$, then $((\gamma+\gamma^6)/(\gamma+\gamma^{10}))^{2^c} = ((\gamma^7+\gamma^{42})/(\gamma^7+\gamma^{70}))^{2^d}$ by (3.23), that is, $\gamma^{3\cdot 2^c} = \gamma^{11\cdot 2^d}$. Since $\gcd(2^c, 15) = \gcd(2^d, 15) = 1$, we have that $\mathrm{ord}(\gamma^3) = \mathrm{ord}(\gamma^{11})$. However, $\mathrm{ord}(\gamma^3) = 5 \neq 15 = \mathrm{ord}(\gamma^{11})$, which is a contradiction. □

4. Improved LDCs and PIR Schemes

In this section, we apply the set $\mathbb{M}_{2,\mathrm{Mersenne}}$ to the constructions of LDCs and information-theoretic PIR schemes. Consequently, we obtain a new family of query-efficient LDCs and a new family of PIR schemes with few servers. Compared with previous results of Efremenko (2009) and Itoh & Suzuki (2010), the new LDCs and

PIR schemes do achieve quantitative improvements of efficiency, which are considerable.

4.1. Query-Efficient Locally Decodable Codes. By Corollary 3.20, Theorem 2.7, Theorem 2.8 and Table 3.1, we have the following theorem:

THEOREM 4.1. *Let* $N_r = \exp(\exp(O(\sqrt[r]{\log n(\log\log n)^{r-1}})))$. *Then, the following statements hold:*

(a) *For every positive integer* $r \leq 103$, *there is a* k*-query linear LDC of length* N_r *for which*

$$k \leq \begin{cases} (\sqrt{3})^r, & \text{if } r \text{ is even} \\ 8 \cdot (\sqrt{3})^{r-3}, & \text{if } r \text{ is odd.} \end{cases}$$

(b) *For every integer* $r \geq 104$, *there is a* k*-query linear LDC of length* N_r *for which* $k \leq (3/4)^{51} \cdot 2^r$.

(c) *If* $|\mathbb{M}_{2,\text{Mersenne}}| = \infty$, *then for every integer* $r \geq 1$, *there is a* k*-query linear LDC of length* N_r *for which* k *is the same as in (a).*

PROOF. (a) Let $r \in [103]$ be even. By Corollary 3.20, we can take distinct $m_1, \ldots, m_{r/2} \in \mathbb{M}_2$ which are pairwise relatively prime. There is a 3-query linear LDC of length N_2 based on each of them by the definition of $\mathbb{M}_2$ and Theorem 2.7. Applying Theorem 2.8 $r/2-1$ times, we obtain a k-query linear LDC of length N_r for which $k \leq 3^{r/2}$, that is, $k \leq (\sqrt{3})^r$.

Let $r \in [103]$ be odd. If $r = 1$, then the Hadamard code is a 2-query linear LDC of length $N_1 = \exp(n)$ satisfying the required condition. If $r \geq 3$, then $r = 2 \cdot \frac{r-3}{2} + 3$ and we can take distinct $m_1, \ldots, m_{\frac{r-3}{2}} \in \mathbb{M}_2$, which are pairwise relatively prime. Since there are infinitely many primes, we can always take another $m_{\frac{r-1}{2}}$ to be a product of three distinct odd primes such that $m_{\frac{r-1}{2}}$ is relatively prime to all of $m_1, \ldots, m_{\frac{r-3}{2}}$. By Theorem 2.7, there are a 3-query linear LDC of length N_2 based on each of $m_1, \ldots, m_{\frac{r-3}{2}}$ and a

k_3-query linear LDC of length N_3 for which $k_3 \leq 2^3$. Applying Theorem 2.8 $(r-3)/2$ times gives a k-query linear LDC of length N_r for which $k \leq 3^{\frac{r-3}{2}} \cdot 8 = 8 \cdot (\sqrt{3})^{r-3}$.

(b) If $r \geq 104$, we take distinct $m_1, \ldots, m_{51} \in \mathbb{M}_2$ and m_{52} a product of $r-102$ distinct odd primes such that $\gcd(m_i, m_j) = 1$ for all distinct $i, j \in [52]$. By Theorem 2.7, there is a 3-query linear LDC of length N_2 based on each of $m_1, \ldots, m_{51}$ and a k_{r-102}-query linear LDC of length N_{r-102} based on m_{52}. Application of Theorem 2.8 gives a k-query linear LDC of length N_r for which $k \leq 3^{51} \cdot 2^{r-102} = (3/4)^{51} \cdot 2^r$.

(c) It suffices to prove the statement for $r \geq 104$. If r is even, we take $r/2$ distinct elements from $\mathbb{M}_{2,\text{Mersenne}}$ and if r is odd, we take $(r-3)/2$ distinct elements from $\mathbb{M}_{2,\text{Mersenne}}$ together with m, a product of three distinct odd primes such that $\gcd(m, m_i) = 1$ for all $i \in [(r-3)/2]$. In both cases, an application of Theorem 2.8 yields the required conclusion. □

4.2. Private Information Retrieval Schemes with Fewer Servers. An important application of LDCs is in the construction of information-theoretic PIR schemes. A PIR scheme allows a user $\mathcal{U}$ to retrieve a data item x_i from a database $x = (x_1, \ldots, x_n) \in \{0,1\}^n$ while keeping the identity i secret from the database operator. Since its introduction by Chor *et al.* (1998), many constructions have been proposed (Chor *et al.* 1998; Ambainis 1997; Itoh 1999; Beimel *et al.* 2005, 2002; Woodruff & Yekhanin 2007; Yekhanin 2008; Raghavendra 2007; Efremenko 2009; Itoh & Suzuki 2010). The efficiency of a PIR scheme is mainly measured by its communication complexity. In this section, we turn our new query-efficient LDCs into PIR schemes that are more efficient than those of Efremenko (2009) and Itoh & Suzuki (2010).

Definition 4.2 (PIR Scheme). *A one-round k-server PIR scheme is a triplet of algorithms $\mathcal{P} = (\mathcal{Q}, \mathcal{A}, \mathcal{C})$, where $\mathcal{Q}$ is a probabilistic query algorithm, $\mathcal{A}$ is an answer algorithm, and $\mathcal{C}$ is a reconstruction algorithm. At the beginning of the scheme, $\mathcal{U}$ picks a random string* $\mathtt{aux}$, *computes a k-tuple of queries* $\mathtt{que} = (\mathtt{que}_1, \ldots, \mathtt{que}_k) = \mathcal{Q}(k, n, i, \mathtt{aux})$ *and sends each query*

$\mathtt{que}_j$ *to server* S_j. *After receiving* $\mathtt{que}_j$, *the server* S_j *replies to* $\mathcal{U}$ *with* $\mathtt{ans}_j = \mathcal{A}(k, n, j, x, \mathtt{que}_j)$. *At last,* $\mathcal{U}$ *outputs* $\mathcal{C}(k, n, i, \mathtt{aux}, \mathtt{ans}_1, \ldots, \mathtt{ans}_k)$ *such that*

Correctness: *For every integer* $n, x \in \{0,1\}^n, i \in [n]$, *and* $\mathtt{aux}$,

$$\mathcal{C}(k, n, i, \mathtt{aux}, \mathtt{ans}_1, \ldots, \mathtt{ans}_k) = x_i.$$

Privacy: *For every* $i_1, i_2 \in [n], j \in [k]$, *and query* $\mathtt{que}$,

$$\begin{aligned}&\Pr[\mathcal{Q}_j(k, n, i_1, \mathtt{aux}) = \mathtt{que}]\\ &\quad = \Pr[\mathcal{Q}_j(k, n, i_2, \mathtt{aux}) = \mathtt{que}].\end{aligned}$$

The *communication complexity* of $\mathcal{P}$, denoted $C_{\mathcal{P}}(k, n)$, is the total number of bits exchanged between the user and all servers, maximized over $x \in \{0,1\}^n, i \in [n]$, and random string $\mathtt{aux}$. We denote by $(k, n; C_{\mathcal{P}}(k, n))$-PIR a k-server PIR scheme with communication complexity $C_{\mathcal{P}}(k, n)$.

Katz & Trevisan (2000) were the first to show generic transformations between information-theoretic PIR schemes and LDCs. Subsequently, Trevisan (2004) introduced the notion of perfectly smooth decoders:

Definition 4.3 (Trevisan 2004). *A* k*-query LDC* $\mathbf{C} : \Sigma^n \to \Gamma^N$ *is said to have a perfectly smooth decoder if it has a local decoding algorithm* $\mathcal{D}$ *satisfying*:

(i) *In every invocation, each query of* $\mathcal{D}$ *is uniformly distributed over* $[N]$.

(ii) *For every* $x \in \Sigma^n$ *and* $i \in [n], \Pr[\mathcal{D}^{\mathbf{C}(x)}(i) = x_i] = 1$.

LDCs with perfectly smooth decoders directly give information-theoretic PIR schemes.

Proposition 4.4 (Trevisan 2004). *If there is a* k*-query LDC* $C : \Sigma^n \to \Gamma^N$, *which has a perfectly smooth decoder, then there is a* $(k, n; k(\log N + \log |\Gamma|))$*-PIR scheme.*

The LDCs obtained by Efremenko (2009) and Itoh & Suzuki (2010) both have perfectly smooth decoders, and so do the LDCs we construct in Section 4.1. Applying Proposition 4.4 to the Itoh-Suzuki LDCs, one obtains a family of positive integers $\{k^{(r)}\}_{r\geq 4}$ for which $k^{(r)} \leq 3 \cdot 2^{r-2}$, such that for every $r \geq 4$, there is a $k^{(r)}$-server PIR scheme whose communication complexity is $\exp(O(\sqrt[s]{\log n (\log\log n)^{s-1}}))$, where $s = \log k^{(r)} + 2 - \log 3$. These PIR schemes are among the most efficient PIR schemes before this work. Here, we improve their results with the following theorem (an easy consequence of Theorem 4.1 and Proposition 4.4).

THEOREM 4.5. *The following statements hold*:

(a) *There is a family of positive integers* $\{k^{\langle r\rangle}\}_{1\leq r\leq 103}$ *for which* $k^{\langle r\rangle} \leq (\sqrt{3})^r$ *if* r *is even, and* $k^{\langle r\rangle} \leq 8 \cdot (\sqrt{3})^{r-3}$ *if* r *is odd, such that for every* $r \in [103]$, *there is a* $k^{\langle r\rangle}$*-server PIR scheme with communication complexity* $\exp(O(\sqrt[s]{\log n (\log\log n)^{s-1}}))$, *where* $s = 2\log k^{\langle r\rangle}/\log 3$ *if* r *is even, and* $s = (2\log k^{\langle r\rangle} - 6 + 3\log 3)/\log 3$ *if* r *is odd.*

(b) *There is a family of positive integers* $\{k^{\langle r\rangle}\}_{r\geq 104}$ *for which* $k^{\langle r\rangle} \leq (3/4)^{51} \cdot 2^r$, *such that for every* $r \geq 104$, *there is a* $k^{\langle r\rangle}$*-server PIR scheme with communication complexity* $\exp(O(\sqrt[s]{\log n (\log\log n)^{s-1}}))$, *where* $s = \log k^{\langle r\rangle} + 102 - 51\log 3$.

(c) *If* $|\mathbb{M}_{2,\mathrm{Mersenne}}| = \infty$, *then there is a family of positive integers* $\{k^{\langle r\rangle}\}_{r\geq 1}$ *for which* $k^{\langle r\rangle} \leq (\sqrt{3})^r$ *if* r *is even, and* $k^{\langle r\rangle} \leq 8 \cdot (\sqrt{3})^{r-3}$ *if* r *is odd, such that for every* $r \geq 1$, *there is a* $k^{\langle r\rangle}$*-server PIR scheme with communication complexity* $\exp(O(\sqrt[s]{\log n (\log\log n)^{s-1}}))$, *where* $s = 2\log k^{\langle r\rangle}/\log 3$ *if* r *is even, and* $s = (2\log k^{\langle r\rangle} - 6 + 3\log 3)/\log 3$ *if* r *is odd.*

5. Conclusion

In this paper, we showed that every Mersenne number, which is the product of two primes, can be used to improve the query complexity by a factor of 3/4 in Efremenko's framework for constructing LDCs. Based on the 50 elements in $\mathbb{M}_{2,\mathrm{Mersenne}}$ we discovered, a new family of query-efficient LDCs of subexponential length with better performance than those of Efremenko (2009) and Itoh & Suzuki (2010) was obtained. Applying our new LDCs to the construction of PIR schemes, we obtained a new family of PIR schemes, which are also more efficient than those of Efremenko (2009) and Itoh & Suzuki (2010). It is an interesting open problem to determine whether $|\mathbb{M}_{2,\mathrm{Mersenne}}| = \infty$. Furthermore, identifying new elements in $\mathbb{M}_{2,\mathrm{Mersenne}}$ can improve our results and is also of interest on its own right.

Acknowledgements

The authors are grateful to Oded Goldreich for valuable suggestions that helped improve the presentation of the paper. The authors also thank Joachim von zur Gathen and the anonymous referee for helpful comments.

Research of Y. M. Chee, S. Ling, and H. Wang is supported in part by the National Research Foundation of Singapore under Research Grant NRF-CRP2-2007-03.

References

A. Ambainis (1997). Upper bound on the communication complexity of private information retrieval. In *ICALP '97: Proccedings of the 24th International Colloquium on Automata, Languages and Programming*, volume 1256 of *Lecture Notes in Comput. Sci.*, 401–407. Springer, Berlin.

A. Beimel, Y. Ishai & E. Kushilevitz (2005). General constructions for information-theoretic private information retrieval. *J. Comput. System Sci.* **71**(2), 213–247.

A. Beimel, Y. Ishai, E. Kushilevitz & J.-F. Raymond (2002) Breaking the $O(n^{\frac{1}{2k-1}})$ barrier for information-theoretic private information retrieval. In *FOCS '02: Proceedings of the 43rd Symposium on Foundations of Computer Science*, 261–270. IEEE Computer Society, Washington, DC, USA.

B. Chor, O. Goldreich, E. Kushilevitz & M. Sudan (1998). Private information retrieval. *J. ACM* **45**(6), 965–982.

C. W. Curtis & I. Reiner (2006). *Representation Theory of Finite Groups and Associative Algebras.* AMS Chelsea Publishing, Providence, RI, xiv+689.

A. Deshpande, R. Jain, T. Kavitha, J. Radhakrishnan & S. V. Lokam (2002). Better Lower Bounds for Locally Decodable Codes. In *CCC '02: Proceedings of the 17th IEEE Annual Conference on Computational Complexity*, 184. IEEE Computer Society, Washington, DC, USA.

Z. Dvir & A. Shpilka (2005). Locally decodable codes with 2 queries and polynomial identity testing for depth 3 circuits. In *STOC '05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, 592–601. ACM, New York.

K. Efremenko (2009). 3-query locally decodable codes of subexponential length. In *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing*, 39–44. ACM, New York.

W. Gasarch (2004). A survey on private information retrieval. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* **82**, 72–107.

O. Goldreich, H. Karloff, L. J. Schulman & L. Trevisan (2006). Lower bounds for linear locally decodable codes and private information retrieval. *Comput. Complexity* **15**(3), 263–296.

P. Gopalan (2009). A note on Efremenko's locally decodable codes. *Electronic Colloquium on Computational Complexity (ECCC)* TR09-069.

V. Grolmusz (2000). Superpolynomial size set-systems with restricted intersections mod 6 and explicit Ramsey graphs. *Combinatorica* **20**(1), 71–85.

T. Itoh (1999). Efficient private information retrieval. *IEICE Trans. Fund. Electronics Comm. E82-A* **1**, 11–20.

T. Itoh & Y. Suzuki (2010). New constructions for query-efficient locally decodable codes of subexponential length. *IEICE Trans. Inform. Syst. E93-D* **2**, 263–270.

J. Katz & L. Trevisan (2000). On the efficiency of local decoding procedures for error-correcting codes. In *STOC '00: Proceedings of* *the Thirty-Second Annual ACM Symposium on Theory of Computing*, 80–86 (electronic). ACM, New York.

K. S. Kedlaya & S. Yekhanin (2008). Locally decodable codes from nice subsets of finite fields and prime factors of Mersenne numbers. *SIAM J. Comput.* **38**(5), 1952–1969.

I. Kerenidis & R. de Wolf (2004). Exponential lower bound for 2-query locally decodable codes via a quantum argument. *J. Comput. System Sci.* **69**(3), 395–420.

F. J. MacWilliams & N. J. A. Sloane (1977). *The Theory of Error-Correcting Codes.* North-Holland Publishing Co., Amsterdam.

B. R. McDonald (1974). *Finite Rings with Identity.* Marcel Dekker Inc., New York, ix+429. Pure and Applied Mathematics, Vol. 28.

K. Obata (2002). Optimal lower bounds for 2-query locally decodable linear codes. In *Randomization and Approximation Techniques in Computer Science*, volume 2483 of *Lecture Notes in Comput. Sci.*, 39–50. Springer, Berlin.

P. Raghavendra (2007). A note on Yekhanin's locally decodable codes. *Electronic Colloquium on Computational Complexity (ECCC)* TR07-016.

D. Shiowattana & S. V. Lokam (2006). An optimal lower bound for 2-query locally decodable linear codes. *Inform. Process. Lett.* **97**(6), 244–250.

L. Trevisan (2004). Some applications of coding theory in computational complexity. In *Complexity of Computations and Proofs*, volume 13 of *Quad. Mat.*, 347–424. Dept. Math., Seconda Univ. Napoli, Caserta.

L. C. Washington (1997). *Introduction to cyclotomic fields*, volume 83 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 2nd edition, xiv+487.

S. Wehner & R. de Wolf (2005). Improved lower bounds for locally decodable codes and private information retrieval. In *ICALP '05: Proceedings of the 32nd International Colloquium on Automata, Languages and Programming*, volume 3580 of *Lecture Notes in Comput. Sci.*, 1424–1436. Springer, Berlin.

D. Woodruff & S. Yekhanin (2007). A geometric approach to information-theoretic private information retrieval. *SIAM J. Comput.* **37**(4), 1046–1056.

D. P. Woodruff (2007). New lower bounds for general locally decodable codes. *Electronic Colloquium on Computational Complexity (ECCC)* TR07-006.

S. Yekhanin (2008). Towards 3-query locally decodable codes of subexponential length. *J. ACM* **55**(1), 1–16.

Manuscript received 8 February 2010

Yeow Meng Chee
Division of Mathematical Sciences,
School of Physical and
Mathematical Sciences,
Nanyang Technological University,
Singapore 637371, Singapore.
ymchee@ntu.edu.sg

Tao Feng
Department of Mathematical
Sciences,
University of Delaware,
Newark, DE 19716, USA.
feng@math.udel.edu

San Ling
Division of Mathematical Sciences,
School of Physical and
Mathematical Sciences,
Nanyang Technological University,
Singapore 637371, Singapore.
lingsan@ntu.edu.sg

Huaxiong Wang
Division of Mathematical Sciences,
School of Physical and
Mathematical Sciences,
Nanyang Technological University,
Singapore 637371, Singapore.
hxwang@ntu.edu.sg

Liang Feng Zhang
Division of Mathematical Sciences,
School of Physical and
Mathematical Sciences,
Nanyang Technological University,
Singapore 637371, Singapore.
liangf.zhang@gmail.com