Locally Decodable Codes From Nice Subsets of Finite Fields and Prime Factors of Mersenne Numbers

Kiran S. Kedlaya
MIT
kedlaya@mit.edu

Sergey Yekhanin
MIT
yekhanin@mit.edu

Abstract

A k-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 k bits of the codeword $C(x)$, even after some constant fraction of codeword bits has been corrupted. The major goal of LDC related research is to establish the optimal trade-off between length and query complexity of such codes.

Recently [35] introduced a novel technique for constructing locally decodable codes and vastly improved the upper bounds for code length. The technique is based on Mersenne primes. In this paper we extend the work of [35] and argue that further progress via these methods is tied to progress on an old number theory question regarding the size of the largest prime factors of Mersenne numbers.

Specifically, we show that every Mersenne number $m = 2^t - 1$ that has a prime factor $p > m^\gamma$ yields a family of $k(\gamma)$-query locally decodable codes of length $\exp\left(n^{1/t}\right)$. Conversely, if for some fixed k and all $\epsilon > 0$ one can use the technique of [35] to obtain a family of k-query LDCs of length $\exp\left(n^\epsilon\right)$; then infinitely many Mersenne numbers have prime factors larger than known currently.

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. It is well-known that 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 k randomly chosen coordinates of $C(x)$, where k can be as small as 2. Locally decodable codes have numerous applications in complexity theory [16, 30], cryptography [6, 12] and the theory of fault tolerant computation [25]. Below is a slightly informal definition of LDCs:

A (k, 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 $1 - \epsilon$, by a randomized decoding procedure that makes only k queries, even if the codeword $C(x)$ is corrupted in up to δN locations.

One should think of $\delta > 0$ and $\epsilon < 1/2$ as constants. The main parameters of interest in LDCs are the length N and the query complexity k. Ideally we would like to have both of them as small as possible. The concept of locally decodable codes was explicitly discussed in various papers in the early 1990s [2, 29, 22]. Katz and

Trevisan [16] were the first to provide a formal definition of LDCs. Further work on locally decodable codes includes [3, 8, 21, 4, 17, 31, 35, 34, 15, 24].

Below is a brief summary of what was known regarding the length of LDCs prior to [35]. The length of optimal 2-query LDCs was settled by Kerenidis and de Wolf in [17] and is $\exp(n)$.[1] The best upper bound for the length of 3-query LDCs was $\exp\left(n^{1/2}\right)$ due to Beimel et al. [3], and the best lower bound is $\tilde{\Omega}(n^2)$ [34]. For general (constant) k the best upper bound was $\exp\left(n^{O(\log\log k/(k\log k))}\right)$ due to Beimel et al. [4] and the best lower bound is $\tilde{\Omega}\left(n^{1+1/(\lceil k/2\rceil-1)}\right)$ [34].

The recent work [35] improved the upper bounds to the extent that it changed the common perception of what may be achievable [13, 12]. [35] introduced a novel technique to construct codes from so-called nice subsets of finite fields and showed that every Mersenne prime $p = 2^t - 1$ yields a family of 3-query LDCs of length $\exp\left(n^{1/t}\right)$. Based on the largest known Mersenne prime [9], this translates to a length of less than $\exp\left(n^{10^{-7}}\right)$. Combined with the recursive construction from [4], this result yields vast improvements for all values of $k > 2$. It has often been conjectured that the number of Mersenne primes is infinite. If indeed this conjecture holds, [35] gets three query locally decodable codes of length $N = \exp\left(n^{O\left(\frac{1}{\log\log n}\right)}\right)$ *for infinitely many* n. Finally, assuming that the conjecture of Lenstra, Pomerance and Wagstaff [32, 23, 33] regarding the density of Mersenne primes holds, [35] gets three query locally decodable codes of length $N = \exp\left(n^{O\left(\frac{1}{\log^{1-\epsilon}\log n}\right)}\right)$ *for all* n, for every $\epsilon > 0$.

1.1 Our results

In this paper we address two natural questions left open by [35]:

1. Are Mersenne primes necessary for the constructions of [35]?
2. Has the technique of [35] been pushed to its limits, or one can construct better codes through a more clever choice of nice subsets of finite fields?

We extend the work of [35] and answer both of the questions above. In what follows let $P(m)$ denote the largest prime factor of m. We show that one does not necessarily need to use Mersenne primes. It suffices to have Mersenne numbers with polynomially large prime factors. Specifically, every Mersenne number $m = 2^t - 1$ such that $P(m) \geq m^\gamma$ yields a family of $k(\gamma)$-query locally decodable codes of length $\exp\left(n^{1/t}\right)$. A partial converse also holds. Namely, if for some fixed $k \geq 3$ and all $\epsilon > 0$ one can use the technique of [35] to (unconditionally) obtain a family of k-query LDCs of length $\exp\left(n^\epsilon\right)$; then for infinitely many t we have

$$P(2^t - 1) \geq (t/2)^{1+1/(k-2)}. \tag{1}$$

The bound (1) may seem quite weak in light of the widely accepted conjecture saying that the number of Mersenne primes is infinite. However (for any $k \geq 3$) this bound is substantially stronger than what is currently known unconditionally. Lower bounds for $P(2^t - 1)$ have received a considerable amount of attention in the number theory literature [26, 27, 10, 28, 20, 19, 11]. The strongest result to date is due to Stewart [28]. It says that for all integers t ignoring a set of asymptotic density zero, and for all functions $\epsilon(t) > 0$ where $\epsilon(t)$ tends to zero monotonically and arbitrarily slowly:

$$P(2^t - 1) > \epsilon(t)t\left(\log t\right)^2 / \log\log t. \tag{2}$$

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

There are no better bounds known to hold for infinitely many values of t, unless one is willing to accept some number theoretic conjectures [20, 19]. We hope that our work will further stimulate the interest in proving lower bounds for $P(2^t - 1)$ in the number theory community.

In summary, we show that one may be able to improve the unconditional bounds of [35] (say, by discovering a new Mersenne number with a very large prime factor) using the same technique. However any attempts to reach the $\exp(n^\epsilon)$ length for some fixed query complexity and all $\epsilon > 0$ require either progress on an old number theory problem or some radically new ideas.

In this paper we deal only with binary codes for the sake of clarity of presentation. We remark however that our results as well as the results of [35] can be easily generalized to larger alphabets. Such generalization will be discussed in detail in [36].

1.2 Outline

In section 3 we introduce the key concepts of [35], namely that of combinatorial and algebraic niceness of subsets of finite fields. We also briefly review the construction of locally decodable codes from nice subsets. In section 4 we show how Mersenne numbers with large prime factors yield nice subsets of prime fields. In section 5 we prove a partial converse. Namely, we show that every finite field $\mathbb{F}_q$ containing a sufficiently nice subset, is an extension of a prime field $\mathbb{F}_p$, where p is a large prime factor of a large Mersenne number. Our main results are summarized in sections 4.3 and 5.4.

2 Notation

We use the following standard mathematical notation:

- $[s] = \{1, \ldots, s\}$;
- $\mathbb{Z}_n$ denotes integers modulo n;
- $\mathbb{F}_q$ is a finite field of q elements;
- $d_H(x, y)$ 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, $L^\perp = \{u \in \mathbb{F}_2^m \mid \forall v \in L, (u, v) = 0\}$;
- For an odd prime p, $\mathrm{ord}_p(2)$ denotes the smallest integer t such that $p \mid 2^t - 1$.

3 Nice subsets of finite fields and locally decodable codes

In this section we introduce the key technical concepts of [35], namely that of combinatorial and algebraic niceness of subsets of finite fields. We briefly review the construction of locally decodable codes from nice subsets. Our review is concise although self-contained. We refer the reader interested in a more detailed and intuitive treatment of the construction to the original paper [35]. We start by formally defining locally decodable codes.

Definition 1 *A binary code $C : \{0,1\}^n \to \{0,1\}^N$ is said to be (k, 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* $d_H(C(x), y) \leq \delta N$: $Pr[\mathcal{A}^y(i) = x_i] \geq 1 - \epsilon$, *where the probability is taken over the random coin tosses of the algorithm* $\mathcal{A}$.

2. $\mathcal{A}$ *makes at most* k *queries to* y.

We now introduce the concepts of combinatorial and algebraic niceness of subsets of finite fields. Our definitions are syntactically slightly different from the original definitions in [35]. We prefer these formulations since they are more appropriate for the purposes of the current paper. In what follows let $\mathbb{F}_q^*$ denote the multiplicative group of $\mathbb{F}_q$.

Definition 2 *A set* $S \subseteq \mathbb{F}_q^*$ *is called* t *combinatorially nice if for some constant* $c > 0$ *and every positive integer* m *there exist two* $n = \lfloor cm^t \rfloor$*-sized collections of vectors* $\{u_1, \ldots, u_n\}$ *and* $\{v_1, \ldots, v_n\}$ *in* $\mathbb{F}_q^m$, *such that*

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

Definition 3 *A set* $S \subseteq \mathbb{F}_q^*$ *is called* k *algebraically nice if* k *is odd and there exists an odd* $k' \leq k$ *and two sets* $S_0, S_1 \subseteq \mathbb{F}_q$ *such that*

- S_0 *is not empty;*
- $|S_1| = k'$;
- *For all* $\alpha \in \mathbb{F}_q$ *and* $\beta \in S$: $|S_0 \cap (\alpha + \beta S_1)| \equiv 0 \mod (2)$.

The following lemma shows that for an algebraically nice set S, the set S_0 can always be chosen to be large. It is a straightforward generalization of [35, lemma 15].

Lemma 4 *Let* $S \subseteq \mathbb{F}_q^*$ *be a* k *algebraically nice set. Let* $S_0, S_1 \subseteq \mathbb{F}_q$ *be sets from the definition of algebraic niceness of* S. *One can always redefine the set* S_0 *to satisfy* $|S_0| \geq \lceil q/2 \rceil$.

Proof: Let L be the linear subspace of $\mathbb{F}_2^q$ spanned by the incidence vectors of the sets $\alpha + \beta S_1$, for $\alpha \in \mathbb{F}_q$ and $\beta \in S$. Observe that L is invariant under the actions of a 1-transitive permutation group (permuting the coordinates in accordance with addition in $\mathbb{F}_q$). This implies that the space $L^\perp$ is also invariant under the actions of the same group. Note that $L^\perp$ has positive dimension since it contains the incidence vector of the set S_0. The last two observations imply that $L^\perp$ has *full support*, i.e., for every $i \in [q]$ there exists a vector $v \in L^\perp$ such that $v_i \neq 0$. It is easy to verify that any linear subspace of $\mathbb{F}_2^q$ that has full support contains a vector of Hamming weight at least $\lceil q/2 \rceil$. Let $v \in L^\perp$ be such a vector. Redefining the set S_0 to be the set of nonzero coordinates of v we conclude the proof. ■

We now proceed to the core proposition of [35] that shows how sets exhibiting both combinatorial and algebraic niceness yield locally decodable codes.

Proposition 5 *Suppose* $S \subseteq \mathbb{F}_q^*$ *is* t *combinatorially nice and* k *algebraically nice; then for every message length* n *there exists a code of length* $\exp(n^{1/t})$ *that is* $(k, \delta, 2k\delta)$ *locally decodable for all* $\delta > 0$.

Proof: Our proof comes in three steps. We specify encoding and local decoding procedures for our codes and then argue the lower bound for the probability of correct decoding. We use the notation from definitions 2 and 3.

Encoding: We assume that our message has length $n = \lfloor cm^t \rfloor$ for some value of m. (Otherwise we pad the message with zeros. It is easy to see that such padding does not not affect the asymptotic length of the code.) Our

code will be linear. Therefore it suffices to specify the encoding of unit vectors $e_1, \ldots, e_n$, where e_j has length n and a unique non-zero coordinate j. We define the encoding of e_j to be a q^m long vector, whose coordinates are labelled by elements of $\mathbb{F}_q^m$. For all $w \in \mathbb{F}_q^m$ we set:

$$\text{Enc}(e_j)_w = \begin{cases} 1, & \text{if } (u_j, w) \in S_0; \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

It is straightforward to verify that we defined a code encoding n bits to $\exp(n^{1/t})$ bits.

Local decoding: Given a (possibly corrupted) codeword y and an index $i \in [n]$, the decoding algorithm $\mathcal{A}$ picks $w \in \mathbb{F}_q^m$, such that $(u_i, w) \in S_0$ uniformly at random, reads $k' \leq k$ coordinates of y, and outputs the sum:

$$\sum_{\lambda \in S_1} y_{w+\lambda v_i}. \tag{4}$$

Probability of correct decoding: First we argue that decoding is always correct if $\mathcal{A}$ picks $w \in \mathbb{F}_q^m$ such that all bits of y in locations $\{w + \lambda v_i\}_{\lambda \in S_1}$ are not corrupted. We need to show that for all $i \in [n]$, $x \in \{0,1\}^n$ and $w \in \mathbb{F}_q^m$, such that $(u_i, w) \in S_0$:

$$\sum_{\lambda \in S_1} \left(\sum_{j=1}^{n} x_j \, \text{Enc}(e_j) \right)_{w+\lambda v_i} = x_i. \tag{5}$$

Note that

$$\sum_{\lambda \in S_1} \left(\sum_{j=1}^{n} x_j \, \text{Enc}(e_j) \right)_{w+\lambda v_i} = \sum_{j=1}^{n} x_j \sum_{\lambda \in S_1} \text{Enc}(e_j)_{w+\lambda v_i} = \sum_{j=1}^{n} x_j \sum_{\lambda \in S_1} I\left[(u_j, w + \lambda v_i) \in S_0\right], \tag{6}$$

where $I[\gamma \in S_0] = 1$ if $\gamma \in S_0$ and zero otherwise. Now note that

$$\sum_{\lambda \in S_1} I\left[(u_j, w + \lambda v_i) \in S_0\right] = \sum_{\lambda \in S_1} I\left[(u_j, w) + \lambda (u_j, v_i) \in S_0\right] = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

The last identity in (7) for $i = j$ follows from: $(u_i, v_i) = 0$, $(u_i, w) \in S_0$ and $k' = |S_1|$ is odd. The last identity for $i \neq j$ follows from $(u_j, v_i) \in S$ and the algebraic niceness of S. Combining identities (6) and (7) we get (5).

Now assume that up to δ fraction of bits of y are corrupted. Let T_i denote the set of coordinates whose labels belong to $\left\{w \in \mathbb{F}_q^m \mid (u_i, w) \in S_0\right\}$. Recall that by lemma 4, $|T_i| \geq q^m/2$. Thus at most 2δ fraction of coordinates in T_i contain corrupted bits. Let $Q_i = \left\{\{w + \lambda v_i\}_{\lambda \in S_1} \mid w : (u_i, w) \in S_0\right\}$ be the family of k'-tuples of coordinates that may be queried by $\mathcal{A}$. $(u_i, v_i) = 0$ implies that elements of Q_i uniformly cover the set T_i. Combining the last two observations we conclude that with probability at least $1 - 2k\delta$ $\mathcal{A}$ picks an uncorrupted k'-tuple and outputs the correct value of x_i. ■

All locally decodable codes constructed in this paper are obtained by applying proposition 5 to certain nice sets. Thus all our codes have the same dependence of ϵ (the probability of the decoding error) on δ (the fraction of corrupted bits). In what follows we often ignore these parameters and consider only the length and query complexity of codes.

4 Mersenne numbers with large prime factors yield nice subsets of prime fields

In what follows let $\langle 2\rangle \subseteq \mathbb{F}_p^*$ denote the multiplicative subgroup of $\mathbb{F}_p^*$ generated by 2. In [35] it is shown that for every Mersenne prime $p = 2^t - 1$ the set $\langle 2\rangle \subseteq \mathbb{F}_p^*$ is simultaneously 3 algebraically nice and $\mathrm{ord}_p(2)$ combinatorially nice. In this section we prove the same conclusion for a substantially broader class of primes.

Lemma 6 *Suppose p is an odd prime; then $\langle 2\rangle \subseteq \mathbb{F}_p^*$ is $\mathrm{ord}_p(2)$ combinatorially nice.*

Proof: Let $t = \mathrm{ord}_p(2)$. Clearly, t divides $p-1$. We need to specify a constant $c > 0$ such that for every positive integer m there exist two $n = \lfloor cm^t \rfloor$-sized collections of m long vectors over $\mathbb{F}_p$ satisfying:

- For all $i \in [n]$, $(u_i, v_i) = 0$;
- For all $i, j \in [n]$ such that $i \neq j$, $(u_j, v_i) \in \langle 2\rangle$.

First assume that m has the shape $m = \binom{m'-1+(p-1)/t}{(p-1)/t}$, for some integer $m' \geq p-1$. In this case [35, lemma 13] gives us a collection of $n = \binom{m'}{p-1}$ vectors with the right properties. Observe that $n \geq cm^t$ for a constant c that depends only on p and t. Now assume m does not have the right shape, and let m_1 be the largest integer smaller than m that does have it. In order to get vectors of length m we use vectors of length m_1 coming from [35, lemma 13] padded with zeros. It is not hard to verify such a construction still gives us $n \geq cm^t$ large families of vectors for a suitably chosen constant c. ■

We use the standard notation $\overline{\mathbb{F}}$ to denote the algebraic closure of the field $\mathbb{F}$. Also let $C_p \subseteq \overline{\mathbb{F}}_2^*$ denote the multiplicative subgroup of p-th roots of unity in $\overline{\mathbb{F}}_2$. The next lemma generalizes [35, lemma 14].

Lemma 7 *Let p be a prime and k be odd. Suppose there exist $\zeta_1, \ldots, \zeta_k \in C_p$ such that*

$$\zeta_1 + \ldots + \zeta_k = 0; \tag{8}$$

then $\langle 2\rangle \subseteq \mathbb{F}_p^*$ is k algebraically nice.

Proof: In what follows we define the set $S_1 \subseteq \mathbb{F}_p$ and prove the existence of a set S_0 such that together S_0 and S_1 yield k algebraic niceness of $\langle 2\rangle$. Identity 8 implies that there exists an odd integer $k' \leq k$ and k' *distinct* p-th roots of unity $\zeta_1', \ldots, \zeta_k' \in C_p$ such that

$$\zeta_1' + \ldots + \zeta_{k'}' = 0. \tag{9}$$

Let $t = \mathrm{ord}_p(2)$. Observe that $C_p \subseteq \mathbb{F}_{2^t}$. Let g be a generator of C_p. Identity (9) yields $g^{\gamma_1} + \ldots + g^{\gamma_{k'}} = 0$, for some distinct values $\{\gamma_i\}_{i\in[k']}$ in $\mathbb{Z}_p$. Set $S_1 = \{\gamma_1, \ldots, \gamma_{k'}\}$.

Consider a natural one to one correspondence between subsets S' of $\mathbb{F}_p$ and polynomials $\phi_{S'}(x)$ in the ring $\mathbb{F}_2[x]/(x^p-1) : \phi_{S'}(x) = \sum_{s\in S'} x^s$. It is easy to see that for all sets $S' \subseteq \mathbb{F}_p$ and all $\alpha, \beta \in \mathbb{F}_p$, such that $\beta \neq 0$:

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

Let α be a variable ranging over $\mathbb{F}_p$ and β be a variable ranging over $\langle 2\rangle$. 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 $\phi_{S_1}(x)$. 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 $\beta \in \langle 2\rangle$. 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 $f \in \mathbb{F}_2[x]$ and any integer $i : f(x^{2^i}) = (f(x))^{2^i}$. ■

In what follows we present sufficient conditions for the existence of k-tuples of p-th roots of unity in $\overline{\mathbb{F}}_2$ that sum to zero. We treat the $k = 3$ case separately since in that case we can use a specialized argument to derive a more explicit conclusion.

4.1 A sufficient condition for the existence of three p-th roots of unity summing to zero

Lemma 8 *Let p be an odd prime. Suppose $\mathrm{ord}_p(2) < (4/3)\log_2 p$; then there exist three p-th roots of unity in $\overline{\mathbb{F}}_2$ that sum to zero.*

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

Let $t = \mathrm{ord}_p(2)$. Note that $C_p \subseteq \mathbb{F}_{2^t}$. Consider a projective plane Fermat curve χ defined by

$$x^{(2^t-1)/p} + y^{(2^t-1)/p} + z^{(2^t-1)/p} = 0. \tag{10}$$

Let us call a point a on χ trivial if one of the coordinates of a is zero. Cyclicity of $\mathbb{F}^*_{2^t}$ implies that χ contains exactly $3(2^t - 1)/p$ trivial $\mathbb{F}_{2^t}$-rational points. Note that every nontrivial point of χ yields a triple of elements of C_p that sum to zero. The classical Weil bound [18, p. 330] provides an estimate

$$|N_q - (q + 1)| \le (d - 1)(d - 2)\sqrt{q} \tag{11}$$

for the number N_q of $\mathbb{F}_q$-rational points on an arbitrary smooth projective plane curve of degree d. (11) implies that in case

$$2^t + 1 > \left(\frac{2^t - 1}{p} - 1\right)\left(\frac{2^t - 1}{p} - 2\right)2^{t/2} + 3\frac{2^t - 1}{p} \tag{12}$$

there exists a nontrivial point on the curve (10). Note that (12) follows from

$$2^t + 1 > \left(\frac{2^t}{p}\right)\left(\frac{2^t}{p}\right)2^{t/2} - \frac{2^{3t/2+1}}{p} + \frac{3 * 2^t}{p}, \tag{13}$$

and (13) follows from

$$2^t > 2^{2t+t/2}/p^2 \quad \text{and} \quad 2^{t/2+1} > 3.$$

Now note that the first inequality above follows from $t < (4/3)\log_2 p$ and the second follows from $t > 1$. ■

Note that the constant $4/3$ in lemma 8 cannot be improved to 2: there are no three elements of $C_{13264529}$ that sum to zero, even though $\mathrm{ord}_2(13264529) = 47 < 2 * \log_2 13264529 \approx 47.3$.

4.2 A sufficient condition for the existence of k p-th roots of unity summing to zero

Our argument in this section comes in three steps. First we briefly review the notion of (additive) Fourier coefficients of subsets of $\mathbb{F}_{2^t}$. Next, we invoke a folklore argument to show that subsets of $\mathbb{F}_{2^t}$ with appropriately small nontrivial Fourier coefficients contain k-tuples of elements that sum to zero. Finally, we use a recent result of Bourgain and Chang [5] (generalizing the classical estimate for Gauss sums) to argue that (under certain constraints on p) all nontrivial Fourier coefficients of C_p are small.

For $x \in \mathbb{F}_{2^t}$ let $\mathrm{Tr}(x) = x + x^2 + \ldots + x^{2^{t-1}}$ denote the trace of x. It is not hard to verify that for all x, $\mathrm{Tr}(x) \in \mathbb{F}_2$. Characters of $\mathbb{F}_{2^t}$ are homomorphisms from the additive group of $\mathbb{F}_{2^t}$ into the multiplicative group

$\{\pm 1\}$. There exist 2^t characters. We denote characters by χ_a, where a ranges in $\mathbb{F}_{2^t}$, and set $\chi_a(x) = (-1)^{\mathrm{Tr}(ax)}$. Let $C(x)$ denote the incidence function of a set $C \subseteq \mathbb{F}_{2^t}$. For arbitrary $a \in \mathbb{F}_2^t$ the Fourier coefficient $\chi_a(C)$ is defined by $\chi_a(C) = \sum \chi_a(x)C(x)$, where the sum is over all $x \in \mathbb{F}_{2^t}$. Fourier coefficient $\chi_0(C) = |C|$ is called trivial, and other Fourier coefficients are called nontrivial. In what follows $\sum_\chi$ stands for summation over all 2^t characters of $\mathbb{F}_{2^t}$. We need the following two standard properties of characters and Fourier coefficients.

$$\sum_\chi \chi(x) = \begin{cases} 2^t, & \text{if } x = 0; \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

$$\sum_\chi \chi^2(C) = 2^t|C|. \tag{15}$$

The following lemma is a folklore.

Lemma 9 *Let $C \subseteq \mathbb{F}_{2^t}$ and $k \geq 3$ be a positive integer. Let F be the largest absolute value of a nontrivial Fourier coefficient of C. Suppose*

$$\frac{F}{|C|} < \left(\frac{|C|}{2^t}\right)^{1/(k-2)} \tag{16}$$

then there exist k elements of C that sum to zero.

Proof: Let $M(C) = \#\{\zeta_1, \ldots, \zeta_k \in C \mid \zeta_1 + \ldots + \zeta_k = 0\}$. (14) yields

$$M(C) = \frac{1}{2^t} \sum_{x_1,\ldots,x_k \in \mathbb{F}_{2^t}} C(x_1)\ldots C(x_k) \sum_\chi \chi(x_1 + \ldots + x_k). \tag{17}$$

Note that $\chi(x_1 + \ldots + x_k) = \chi(x_1)\ldots\chi(x_k)$. Changing the order of summation in (17) we get

$$M(C) = \frac{1}{2^t} \sum_\chi \sum_{x_1,\ldots,x_k \in \mathbb{F}_{2^t}} C(x_1)\ldots C(x_k)\chi(x_1)\ldots\chi(x_k) = \frac{1}{2^t}\sum_\chi \chi^k(C). \tag{18}$$

Note that

$$\frac{1}{2^t}\sum_\chi \chi^k(C) = \frac{|C|^k}{2^t} + \frac{1}{2^t}\sum_{\chi \neq \chi_0} \chi^k(C) \geq \frac{|C|^k}{2^t} - F^{k-2}\frac{1}{2^t}\sum_\chi \chi^2(C) = \frac{|C|^k}{2^t} - F^{k-2}|C|, \tag{19}$$

where the last identity follows from (15). Combining (18) and (19) we conclude that (16) implies $M(C) > 0$. ■

The following lemma is a special case of [5, theorem 1].

Lemma 10 *Assume that $n \mid 2^t - 1$ and satisfies the condition*

$$\gcd\left(n, \frac{2^t - 1}{2^{t'} - 1}\right) < 2^{t(1-\epsilon)-t'}, \quad \textit{for all} \quad 1 \leq t' < t,\ t' \mid t,$$

where $\epsilon > 0$ is arbitrary and fixed. Then for all $a \in \mathbb{F}_{2^t}^*$

$$\left|\sum_{x \in \mathbb{F}_{2^t}} (-1)^{\mathrm{Tr}(ax^n)}\right| < c_1 2^{t(1-\delta)}, \tag{20}$$

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

Below is the main result of this section. Recall that C_p denotes the set of p-th roots of unity in $\overline{\mathbb{F}}_2$.

Lemma 11 *For every $c > 0$ there exists an odd integer $k = k(c)$ such that the following implication holds. If p is an odd prime and $\mathrm{ord}_p(2) < c \log_2 p$ then some k elements of C_p sum to zero.*

Proof: Note that if there exist k' elements of a set $C \subseteq \overline{\mathbb{F}}_2$ that sum to zero, where k' is odd; then there exist k elements of C that sum to zero for every odd $k \geq k'$. Also note that the sum of all p-th roots of unity is zero. Therefore given c it suffices to prove the existence of an odd $k = k(c)$ that works for all *sufficiently large* p. Let $t = \mathrm{ord}_p(2)$. Observe that $p > 2^{t/c}$. Assume p is sufficiently large so that $t > 2c$. Next we show that the precondition of lemma 10 holds for $n = (2^t - 1)/p$ and $\epsilon = 1/(2c)$. Let $t' \mid t$ and $1 \leq t' < t$. Clearly $\gcd(2^{t'} - 1, p) = 1$. Therefore

$$\gcd\left(\frac{2^t - 1}{p}, \frac{2^t - 1}{2^{t'} - 1}\right) = \frac{2^t - 1}{p(2^{t'} - 1)} < \frac{2^{t(1-1/c)})}{2^{t'} - 1}, \tag{21}$$

where the inequality follows from $p > 2^{t/c}$. Clearly, $t > 2c$ yields $2^{t/(2c)}/2 > 1$. Multiplying the right hand side of (21) by $2^{t/(2c)}/2$ and using $2(2^{t'} - 1) \geq 2^{t'}$ we get

$$\gcd\left(\frac{2^t - 1}{p}, \frac{2^t - 1}{2^{t'} - 1}\right) < 2^{t(1-1/(2c))-t'}. \tag{22}$$

Combining (22) with lemma 10 we conclude that there exist $\delta > 0$ and c_1 such that for all $a \in \mathbb{F}_{2^t}^*$

$$\left| \sum_{x \in \mathbb{F}_{2^t}} (-1)^{\mathrm{Tr}\left(ax^{(2^t-1)/p}\right)} \right| < c_1 2^{t(1-\delta)}. \tag{23}$$

Observe that $x^{(2^t-1)/p}$ takes every value in C_p exactly $(2^t - 1)/p$ times when x ranges over $\mathbb{F}_{2^t}^*$. Thus (23) implies

$$(2^t - 1)(F/p) < c_1 2^{t(1-\delta)} + 1, \tag{24}$$

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

$$(2^t - 1)(F/p) < c_2 2^{t(1-\delta)}, \tag{25}$$

for a suitably chosen constant c_2. (25) yields $F/p < (2c_2)2^{-\delta t}$. Pick $k \geq 3$ to be the smallest odd integer such that $(1 - 1/c)/(k - 2) < \delta$. We now have

$$\frac{F}{p} < 2^{-\frac{(1-1/c)t}{(k-2)}} \tag{26}$$

for all sufficiently large values of p. Combining $p > 2^{t/c}$ with (26) we get

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

and the application of lemma 9 concludes the proof. ■

4.3 Summary

In this section we summarize our positive results and show that one does not necessarily need to use Mersenne primes to construct locally decodable codes via the methods of [35]. It suffices to have Mersenne numbers with polynomially large prime factors. Recall that $P(m)$ denotes the largest prime factor of an integer m. Our first theorem gets 3-query LDCs from Mersenne numbers m with prime factors larger than $m^{3/4}$.

Theorem 12 *Suppose $P(2^t - 1) > 2^{0.75t}$; then for every message length n there exists a three query locally decodable code of length $\exp(n^{1/t})$.*

Proof: Let $P(2^t - 1) = p$. Observe that $p \mid 2^t - 1$ and $p > 2^{0.75t}$ yield $\mathrm{ord}_p(2) < (4/3)\log_2 p$. Combining lemmas 8,7 and 6 with proposition 5 we obtain the statement of the theorem. ■

As an example application of theorem 12 one can observe that $P(2^{23} - 1) = 178481 > 2^{(3/4)*23} \approx 155872$ yields a family of three query locally decodable codes of length $\exp(n^{1/23})$. Theorem 12 immediately yields:

Theorem 13 *Suppose for infinitely many t we have $P(2^t - 1) > 2^{0.75t}$; then for every $\epsilon > 0$ there exists a family of three query locally decodable codes of length $\exp(n^{\epsilon})$.*

The next theorem gets constant query LDCs from Mersenne numbers m with prime factors larger than m^{γ} for every value of γ.

Theorem 14 *For every $\gamma > 0$ there exists an odd integer $k = k(\gamma)$ such that the following implication holds. Suppose $P(2^t - 1) > 2^{\gamma t}$; then for every message length n there exists a k-query locally decodable code of length $\exp(n^{1/t})$.*

Proof: Let $P(2^t - 1) = p$. Observe that $p \mid 2^t - 1$ and $p > 2^{\gamma t}$ yield $\mathrm{ord}_p(2) < (1/\gamma)\log_2 p$. Combining lemmas 22,7 and 6 with proposition 5 we obtain the statement of the theorem. ■

As an immediate corollary we get:

Theorem 15 *Suppose for some $\gamma > 0$ and infinitely many t we have $P(2^t - 1) > 2^{\gamma t}$; then there is a fixed k such that for every $\epsilon > 0$ there exists a family of k-query locally decodable codes of length $\exp(n^{\epsilon})$.*

5 Nice subsets of finite fields yield Mersenne numbers with large prime factors

Definition 16 *We say that a sequence $\left\{S_i \subseteq \mathbb{F}_{q_i}^*\right\}_{i \geq 1}$ of subsets of finite fields is k-nice if every S_i is k algebraically nice and $t(i)$ combinatorially nice, for some integer valued monotonically increasing function t.*

The core proposition 5 asserts that a subset $S \subseteq \mathbb{F}_q^*$ that is k algebraically nice and t combinatorially nice yields a family of k-query locally decodable codes of length $\exp(n^{1/t})$. Clearly, to get k-query LDCs of length $\exp(n^{\epsilon})$ for some fixed k and every $\epsilon > 0$ via this proposition, one needs to exhibit a k-nice sequence. In this section we show how the existence of a k-nice sequence implies that infinitely many Mersenne numbers have large prime factors. Our argument proceeds in two steps. First we show that a k-nice sequence yields an infinite sequence of primes $\{p_i\}_{i \geq 1}$, where every C_{p_i} contains a k-tuple of elements summing to zero. Next we show that C_p contains a (nontrivial) short additive dependence only if p is a large factor of a Mersenne number.

5.1 A nice sequence yields infinitely many primes p with short dependencies between p-th roots of unity

We start with some notation. Consider a a finite field $\mathbb{F}_q = \mathbb{F}_{p^l}$, where p is prime. Fix a basis $e_1, \ldots, e_l$ of $\mathbb{F}_q$ over $\mathbb{F}_p$. In what follows we often write $(\alpha_1, \ldots, \alpha_l) \in \mathbb{F}_p^l$ to denote $\alpha = \sum_{i=1}^{l} \alpha_i e_i \in \mathbb{F}_q$. Let R denote the ring $\mathbb{F}_2[x_1, \ldots, x_l]/(x_1^p - 1, \ldots, x_l^p - 1)$. Consider a natural one to one correspondence between subsets S_1 of $\mathbb{F}_q$ and polynomials $\phi_{S_1}(x_1, \ldots, x_l) \in R$.

$$\phi_{S_1}(x_1, \ldots, x_l) = \sum_{(\alpha_1, \ldots, \alpha_l) \in S_1} x_1^{\alpha_1} \ldots x_l^{\alpha_l}.$$

It is easy to see that for all sets $S_1 \subseteq \mathbb{F}_q$ and all $\alpha, \beta \in \mathbb{F}_q$:

$$\phi_{(\alpha_1, \ldots, \alpha_l) + \beta S_1}(x_1, \ldots, x_l) = x_1^{\alpha_1} \ldots x_l^{\alpha_l} \phi_{\beta S_1}(x_1, \ldots, x_l). \tag{27}$$

Let Γ be a family of subsets of $\mathbb{F}_q$. It is straightforward to verify that a set $S_0 \subseteq \mathbb{F}_q$ has even intersections with every element of Γ if and only if ϕ_{S_0} belongs to $L^{\perp}$, where L is the linear subspace of R spanned by $\{\phi_{S_1}\}_{S_1 \in \Gamma}$. Combining the last observation with formula (27) we conclude that a set $S \subseteq \mathbb{F}_q^*$ is k algebraically nice if and only if there exists a set $S_1 \subseteq \mathbb{F}_q$ of odd size $k' \leq k$ such that the ideal generated by polynomials $\{\phi_{\beta S_1}\}_{\{\beta \in S\}}$ is a proper ideal of R. Note that polynomials $\{f_1, \ldots, f_h\} \in R$ generate a proper ideal if an only if polynomials $\{f_1, \ldots, f_h, x_1^p - 1, \ldots, x_l^p - 1\}$ generate a proper ideal in $\mathbb{F}_2[x_1, \ldots, x_l]$. Also note that a family of polynomials generates a proper ideal in $\mathbb{F}_2[x_1, \ldots, x_l]$ if and only if it generates a proper ideal in $\overline{\mathbb{F}}_2[x_1, \ldots, x_l]$. Now an application of Hilbert's Nullstellensatz [7, p. 168] implies that a set $S \subseteq \mathbb{F}_q^*$ is k algebraically nice if and only if there is a set $S_1 \subseteq \mathbb{F}_q$ of odd size $k' \leq k$ such that the polynomials $\{\phi_{\beta S_1}\}_{\{\beta \in S\}}$ and $\{x_i^p - 1\}_{1 \leq i \leq l}$ have a common root in $\overline{\mathbb{F}}_2$.

Lemma 17 *Let $\mathbb{F}_q = \mathbb{F}_{p^l}$, where p is prime. Suppose $\mathbb{F}_q$ contains a nonempty k algebraically nice subset; then there exist $\zeta_1, \ldots, \zeta_k \in C_p$ such that $\zeta_1 + \ldots + \zeta_k = 0$.*

Proof: Assume $S \subseteq \mathbb{F}_q^*$ is nonempty and k algebraically nice. The discussion above implies that there exists $S_1 \subseteq \mathbb{F}_q$ of odd size $k' \leq k$ such that all polynomials $\{\phi_{\beta S_1}\}_{\{\beta \in S\}}$ vanish at some $(\zeta_1, \ldots, \zeta_l) \in C_p^l$. Fix an arbitrary $\beta_0 \in S$, and note that C_p is closed under multiplication. Thus,

$$\phi_{\beta_0 S_1}(\zeta_1, \ldots, \zeta_l) = 0 \tag{28}$$

yields k' p-th roots of unity that add up to zero. It is readily seen that one can extend (28) (by adding an appropriate number of pairs of identical roots) to obtain k p-th roots of unity that add up to zero for any odd $k \geq k'$. ■

Note that lemma 17 does not suffice to prove that a k-nice sequence $\left\{S_i \subseteq \mathbb{F}_{q_i}^*\right\}_{i \geq 1}$ yields infinitely many primes p with short (nontrivial) additive dependencies in C_p. We need to argue that the set $\left\{\mathrm{char}\mathbb{F}_{q_i}\right\}_{i \geq 1}$ can not be finite. To proceed, we need some more notation. Recall that $q = p^l$ and p is prime. For $x \in \mathbb{F}_q$ let $\mathrm{Tr}(x) = x + \ldots + x^{p^{l-1}} \in \mathbb{F}_p$ denote the (absolute) trace of x. For $\gamma \in \mathbb{F}_q, c \in \mathbb{F}_p^*$ we call the set $\pi_{\gamma,c} = \{x \in \mathbb{F}_q \mid \mathrm{Tr}(\gamma x) = c\}$ a *proper affine hyperplane* of $\mathbb{F}_q$.

Lemma 18 *Let $\mathbb{F}_q = \mathbb{F}_{p^l}$, where p is prime. Suppose $S \subseteq \mathbb{F}_q^*$ is k algebraically nice; then there exist $h \leq p^k$ proper affine hyperplanes $\{\pi_{\gamma_r, c_r}\}_{1 \leq r \leq h}$ of $\mathbb{F}_q$ such that $S \subseteq \bigcup_{r=1}^{h} \pi_{\gamma_r, c_r}$.*

Proof: Discussion preceding lemma 17 implies that there exists a set $S_1 = \{\sigma_1, \ldots, \sigma_{k'}\} \subseteq \mathbb{F}_q$ of odd size $k' \leq k$ such that all polynomials $\{\phi_{\beta S_1}\}_{\{\beta \in S\}}$ vanish at some $(\zeta_1, \ldots, \zeta_l) \in C_p^l$. Let ζ be a generator of C_p. For every $1 \leq i \leq l$ pick $\omega_i \in \mathbb{Z}_p$ such that $\zeta_i = \zeta^{\omega_i}$. For every $\beta \in S$, $\phi_{\beta S_1}(\zeta_1, \ldots, \zeta_l) = 0$ yields

$$\sum_{\mu=(\mu_1,\ldots,\mu_l)\in\beta S_1} \zeta^{\sum_{i=1}^{l}\mu_i\omega_i} = 0. \tag{29}$$

Observe that for fixed values $\{\omega_i\}_{1\leq i\leq l} \in \mathbb{Z}_p$ the map $D(\mu) = \sum_{i=1}^{l}\mu_i\omega_i$ is a linear map from $\mathbb{F}_q$ to $\mathbb{F}_p$. It is not hard to prove that every such map can be expressed as $D(\mu) = \mathrm{Tr}(\delta\mu)$ for an appropriate choice of $\delta \in \mathbb{F}_q$. Therefore we can rewrite (29) as

$$\sum_{\mu\in\beta S_1} \zeta^{\mathrm{Tr}(\delta\mu)} = \sum_{\sigma\in S_1} \zeta^{\mathrm{Tr}(\delta\beta\sigma)} = 0. \tag{30}$$

Let $W = \left\{(w_1, \ldots, w_{k'}) \in \mathbb{Z}_p^{k'} \mid \zeta^{w_1} + \ldots + \zeta^{w_{k'}} = 0\right\}$ denote the set of exponents of k'-dependencies between powers of ζ. Clearly, $|W| \leq p^k$. Identity (30) implies that every $\beta \in S$ satisfies

$$\begin{cases} \mathrm{Tr}((\delta\sigma_1)\beta) &= w_1, \\ \vdots & \\ \mathrm{Tr}((\delta\sigma_{k'})\beta) &= w_{k'}; \end{cases} \tag{31}$$

for an appropriate choice of $(w_1, \ldots, w_{k'}) \in W$. Note that the all-zeros vector does not lie in W since k' is odd. Therefore at least one of the identities in (31) has a non-zero right-hand side, and defines a proper affine hyperplane of $\mathbb{F}_q$. Collecting one such hyperplane for every element of W we get a family of $|W|$ proper affine hyperplanes containing every element of S. ■

Lemma 18 gives us some insight into the structure of algebraically nice subsets of $\mathbb{F}_q$. Our next goal is to develop an insight into the structure of combinatorially nice subsets. We start by reviewing some relations between tensor and dot products of vectors. For vectors $u \in \mathbb{F}_q^m$ and $v \in \mathbb{F}_q^n$ let $u \otimes v \in \mathbb{F}_q^{mn}$ denote the tensor product of u and v. Coordinates of $u \otimes v$ are labelled by all possible elements of $[m] \times [n]$ and $(u \otimes v)_{i,j} = u_i v_j$. Also, let $u^{\otimes l}$ denote the l-the tensor power of u and $u \circ v$ denote the concatenation of u and v. The following identity is standard. For any $u, x \in \mathbb{F}_q^m$ and $v, y \in \mathbb{F}_q^n$:

$$(u \otimes v, x \otimes y) = \sum_{i\in[m],j\in[n]} u_i v_j x_i y_j = \left(\sum_{i\in[m]} u_i x_i\right)\left(\sum_{j\in[n]} v_j y_j\right) = (u, x)(v, y). \tag{32}$$

In what follows we need a generalization of identity (32). Let $f(x_1, \ldots, x_h) = \sum_i c_i x_1^{\alpha_1^i} \ldots x_h^{\alpha_h^i}$ be a polynomial in $\mathbb{F}_q[x_1, \ldots, x_h]$. Given f we define $\bar{f} \in \mathbb{F}_q[x_1, \ldots, x_h]$ by $\bar{f} = \sum_i x_1^{\alpha_1^i} \ldots x_h^{\alpha_h^i}$, i.e., we simply set all nonzero coefficients of f to 1. For vectors $u_1, \ldots, u_h$ in $\mathbb{F}_q^m$ define

$$f(u_1, \ldots, u_h) = \circ_i\, c_i u_1^{\otimes\alpha_1^i} \otimes \ldots \otimes u_h^{\otimes\alpha_h^i}. \tag{33}$$

Note that to obtain $f(u_1, \ldots, u_h)$ we replaced products in f by tensor products and addition by concatenation. Clearly, $f(u_1, \ldots, u_h)$ is a vector whose length may be larger than m.

Claim 19 *For every $f \in \mathbb{F}_q[x_1, \ldots, x_h]$ and $u_1, \ldots, u_h, v_1, \ldots, v_h \in \mathbb{F}_q^m$:*

$$\left(f(u_1, \ldots, u_h), \bar{f}(v_1, \ldots, v_h)\right) = f((u_1, v_1), \ldots, (u_h, v_h)). \tag{34}$$

Proof: Let $\mathbf{u} = (u_1, \ldots, u_h)$ and $\mathbf{v} = (v_1, \ldots, v_h)$. Observe that if (34) holds for polynomials f_1 and f_2 defined over disjoint sets of monomials then it also holds for $f = f_1 + f_2$:

$$\left(f(\mathbf{u}), \bar{f}(\mathbf{v})\right) = \left((f_1 + f_2)(\mathbf{u}), (\bar{f}_1 + \bar{f}_2)(\mathbf{v})\right) = \left(f_1(\mathbf{u}) \circ f_2(\mathbf{u}), \bar{f}_1(\mathbf{v}) \circ \bar{f}_2(\mathbf{v})\right) =$$
$$f_1\left((u_1, v_1), \ldots, (u_h, v_h)\right) + f_2\left((u_1, v_1), \ldots, (u_h, v_h)\right) = f\left((u_1, v_1), \ldots, (u_h, v_h)\right).$$

Therefore it suffices to prove (34) for monomials $f = cx_1^{\alpha_1} \ldots x_h^{\alpha_h}$. It remains to notice identity (34) for monomials $f = cx_1^{\alpha_1} \ldots x_h^{\alpha_h}$ follows immediately from formula (32) using induction on $\sum_{i=1}^{h} \alpha_i$. ■

The next lemma bounds combinatorial niceness of certain subsets of $\mathbb{F}_q^*$.

Lemma 20 *Let $\mathbb{F}_q = \mathbb{F}_{p^l}$, where p is prime. Let $S \subseteq \mathbb{F}_q^*$. Suppose there exist h proper affine hyperplanes $\{\pi_{\gamma_r, c_r}\}_{1 \le r \le h}$ of $\mathbb{F}_q$ such that $S \subseteq \bigcup_{r=1}^{h} \pi_{\gamma_r, c_r}$; then S is at most $h(p-1)$ combinatorially nice.*

Proof: Assume S is t combinatorially nice. This implies that for some $c > 0$ and every m there exist two $n = \lfloor cm^t \rfloor$-sized collections of vectors $\{u_i\}_{i \in [n]}$ and $\{v_i\}_{i \in [n]}$ in $\mathbb{F}_q^m$, such that:

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

For a vector $u \in \mathbb{F}_q^m$ and integer e let u^e denote a vector resulting from raising every coordinate of u to the power e. For every $i \in [n]$ and $r \in [h]$ define vectors $u_i^{(r)}$ and $v_i^{(r)}$ in $\mathbb{F}_q^{ml}$ by

$$u_i^{(r)} = (\gamma_r u_i) \circ (\gamma_r u_i)^p \circ \ldots \circ (\gamma_r u_i)^{p^{l-1}} \quad \text{and} \quad v_i^{(r)} = v_i \circ v_i^p \circ \ldots \circ v_i^{p^{l-1}}. \tag{35}$$

Note that for every $r_1, r_2 \in [h]$, $v_i^{(r_1)} = v_i^{(r_2)}$. It is straightforward to verify that for every $i, j \in [n]$ and $r \in [h]$:

$$\left(u_j^{(r)}, v_i^{(r)}\right) = \mathrm{Tr}(\gamma_r(u_j, v_i)). \tag{36}$$

Combining (36) with the fact that S is covered by proper affine hyperplanes $\{\pi_{\gamma_r, c_r}\}_{r \in [h]}$ we conclude that

- For all $i \in [n]$ and $r \in [h]$, $\left(u_i^{(r)}, v_i^{(r)}\right) = 0$;
- For all $i, j \in [n]$ such that $i \neq j$, there exists $r \in [h]$ such that $\left(u_j^{(r)}, v_i^{(r)}\right) \in \mathbb{F}_p^*$.

Pick $g(x_1, \ldots, x_h) \in \mathbb{F}_p[x_1, \ldots, x_h]$ to be a homogeneous degree h polynomial such that for $\mathbf{a} = (a_1, \ldots, a_h) \in \mathbb{F}_p^h$: $g(\mathbf{a}) = 0$ if and only if $\mathbf{a}$ is the all-zeros vector. The existence of such a polynomial g follows from [18, Example 6.7]. Set $f = g^{p-1}$. Note that for $\mathbf{a} \in \mathbb{F}_p^h$: $f(a) = 0$ if $\mathbf{a}$ is the all-zeros vector, and $f(a) = 1$ otherwise. For all $i \in [n]$ define

$$u_i' = f\left(u_i^{(1)}, \ldots, u_i^{(h)}\right) \circ (1) \quad \text{and} \quad v_i' = \bar{f}\left(v_i^{(1)}, \ldots, v_i^{(h)}\right) \circ (-1). \tag{37}$$

Note that f and $\bar{f}$ are homogeneous degree $(p-1)h$ polynomials in h variables. Therefore (33) implies that for all i vectors u_i' and v_i' have length $m' \le h^{(p-1)h}(ml)^{(p-1)h} + 1$. Combining identities (37) and (34) and using the properties of dot products between vectors $\left\{u_i^{(r)}\right\}$ and $\left\{v_i^{(r)}\right\}$ discussed above we conclude that for every m there exist two $n = \lfloor cm^t \rfloor$-sized collections of vectors $\{u_i'\}_{i \in [n]}$ and $\{v_i'\}_{i \in [n]}$ in $\mathbb{F}_q^{m'}$, such that:

- For all $i \in [n]$, $(u'_i, v'_i) = -1$;
- For all $i, j \in [n]$ such that $i \neq j$, $(u_j, v_i) = 0$.

It remains to notice that a family of vectors with such properties exists only if $n \leq m'$, i.e.,

$$\lfloor cm^t \rfloor \leq h^{(p-1)h}(ml)^{(p-1)h} + 1.$$

Given that we can pick m to be arbitrarily large, this implies that $t \leq (p-1)h$. ■

The next lemma presents the main result of this section.

Lemma 21 *Let k be an odd integer. Suppose there exists a k-nice sequence; then for infinitely many primes p some k of elements of C_p add up to zero.*

Proof: Assume $\left\{S_i \subseteq \mathbb{F}_{q_i}^*\right\}_{i\geq 1}$ is k-nice. Let p be a fixed prime. Combining lemmas 18 and 20 we conclude that every k algebraically nice subset $S \subseteq \mathbb{F}_{p^l}^*$ is at most $(p-1)p^k$ combinatorially nice. Note that our bound on combinatorial niceness is independent of l. Therefore there are only finitely many extensions of the field $\mathbb{F}_p$ in the sequence $\{\mathbb{F}_{q_i}\}_{i\geq 1}$, and the set $\mathbb{P} = \{\mathrm{char}\mathbb{F}_{q_i}\}_{i\geq 1}$ is infinite. It remains to notice that according to lemma 17 for every $p \in \mathbb{P}$ there exist k elements of C_p that add up to zero. ■

In what follows we present necessary conditions for the existence of k-tuples of p-th roots of unity in $\overline{\mathbb{F}}_2$ that sum to zero. We treat the $k = 3$ case separately since in that case we can use a specialized argument to derive a slightly stronger conclusion.

5.2 A necessary condition for the existence of k p-th roots of unity summing to zero

Lemma 22 *Let $k \geq 3$ be odd and p be a prime. Suppose there exist $\zeta_1, \ldots, \zeta_k \in C_p$ such that $\sum_{i=1}^k \zeta_i = 0$; then*

$$\mathrm{ord}_p(2) \leq 2p^{1-1/(k-1)}. \tag{38}$$

Proof: Let $t = \mathrm{ord}_p(2)$. Note that $C_p \subseteq \mathbb{F}_{2^t}$. Note also that all elements of C_p other than the multiplicative identity are proper elements of $\mathbb{F}_{2^t}$. Therefore for every $\zeta \in C_p$ where $\zeta \neq 1$ and every nonzero $f(x) \in \mathbb{F}_2[x]$ such that $\deg f \leq t-1$ we have: $f(\zeta) \neq 0$.

By multiplying $\sum_{i=1}^k \zeta_i = 0$ through by ζ_k^{-1}, we may reduce to the case $\zeta_k = 1$. Let ζ be the generator of C_p. For every $i \in [k-1]$ pick $w_i \in \mathbb{Z}_p$ such that $\zeta_i = \zeta^{w_i}$. We now have $\sum_{i=1}^{k-1} \zeta^{w_i} + 1 = 0$. Set $h = \lfloor (t-1)/2 \rfloor$. Consider the $(k-1)$-tuples:

$$(mw_1 + i_1, \ldots, mw_{k-1} + i_{k-1}) \in \mathbb{Z}_p^{k-1}, \text{ for } m \in \mathbb{Z}_p \text{ and } i_1, \ldots, i_{k-1} \in [0, h]. \tag{39}$$

Suppose two of these coincide, say

$$(mw_1 + i_1, \ldots, mw_{k-1} + i_{k-1}) = (m'w_1 + i'_1, \ldots, m'w_{k-1} + i'_{k-1}),$$

with $(m, i_1, \ldots, i_{k-1}) \neq (m', i'_1, \ldots, i'_{k-1})$. Set $n = m - m'$ and $j_l = i'_l - i_l$ for $l \in [k-1]$. We now have

$$(nw_1, \ldots, nw_{k-1}) = (j_1, \ldots, j_{k-1})$$

with $-h \leq j_1, \ldots, j_{k-1} \leq h$. Observe that $n \neq 0$, and thus it has a multiplicative inverse $g \in \mathbb{Z}_p$. Consider a polynomial

$$P(z) = z^{j_1+h} + \ldots + z^{j_{k-1}+h} + z^h \in \mathbb{F}_2[z].$$

Note that $\deg P \leq 2h \leq t-1$. Note also that $P(1) = 1$ and $P(\zeta^g) = 0$. The latter identity contradicts the fact that ζ^g is a proper element of $\mathbb{F}_{2^t}$. This contradiction implies that all $(k-1)$-tuples in (39) are distinct. This yields

$$p^{k-1} \geq p \left(\frac{t}{2}\right)^{k-1},$$

which is equivalent to (38). ■

5.3 A necessary condition for the existence of three p-th roots of unity summing to zero

In this section we slightly strengthen lemma 22 in the special case when $k = 3$. Our argument is loosely inspired by the Agrawal-Kayal-Saxena deterministic primality test [1].

Lemma 23 *Let p be a prime. Suppose there exist $\zeta_1, \zeta_2, \zeta_3 \in C_p$ that sum up to zero; then*

$$\mathrm{ord}_p(2) \leq ((4/3)p)^{1/2}. \tag{40}$$

Proof: Let $t = \mathrm{ord}_p(2)$. Note that $C_p \subseteq \mathbb{F}_{2^t}$. Note also that all elements of C_p other than the multiplicative identity are proper elements of $\mathbb{F}_{2^t}$. Therefore for every $\zeta \in C_p$ where $\zeta \neq 1$ and every nonzero $f(x) \in \mathbb{F}_2[x]$ such that $\deg f \leq t-1$ we have: $f(\zeta) \neq 0$.

Observe that $\zeta_1 + \zeta_2 + \zeta_3 = 0$ implies $\zeta_1\zeta_2^{-1} + 1 = \zeta_3\zeta_2^{-1}$. This yields $\left(\zeta_1\zeta_2^{-1} + 1\right)^p = 1$. Put $\zeta = \zeta_1\zeta_2^{-1}$. Note that $\zeta \neq 1$ and $\zeta, 1+\zeta \in C_p$. Consider the products $\pi_{i,j} = \zeta^i(1+\zeta)^j \in C_p$ for $0 \leq i, j \leq t-1$. Note that $\pi_{i,j}, \pi_{k,l}$ cannot be the same if $i \geq k$ and $l \geq j$, as then

$$\zeta^{i-k} - (1+\zeta)^{l-j} = 0,$$

but the left side has degree less than t. In other words, if $\pi_{i,j} = \pi_{k,l}$ and $(i,j) \neq (k,l)$, then the pairs (i,j) and (k,l) are comparable under termwise comparison. In particular, either $(k,l) = (i+a, j+b)$ or $(i,j) = (k+a, l+b)$ for some pair (a,b) with $\pi_{a,b} = 1$.

We next check that there cannot be two distinct nonzero pairs $(a,b), (a',b')$ with $\pi_{a,b} = \pi_{a',b'} = 1$. As above, these pairs must be comparable; we may assume without loss of generality that $a \leq a', b \leq b'$. The equations $\pi_{a,b} = 1$ and $\pi_{a'-a,b'-b} = 1$ force $a + b \geq t$ and $(a'-a) + (b'-b) \geq t$, so $a' + b' \geq 2t$. But $a', b' \leq t-1$, contradiction.

If there is no nonzero pair (a,b) with $0 \leq a, b \leq t-1$ and $\pi_{a,b} = 1$, then all $\pi_{i,j}$ are distinct, so $p \geq t^2$. Otherwise, as above, the pair (a,b) is unique, and the pairs (i,j) with $0 \leq i, j \leq t-1$ and $(i,j) \not\geq (a,b)$ are pairwise distinct. The number of pairs excluded by the condition $(i,j) \not\geq (a,b)$ is $(t-a)(t-b)$; since $a + b \geq t$, $(t-a)(t-b) \leq t^2/4$. Hence $p \geq t^2 - t^2/4 = 3t^2/4$ as desired. ■

While the necessary condition given by lemma 23 is quite far away from the sufficient condition given by lemma 8, it nonetheless suffices for checking that for most primes p, there do not exist three p-th roots of unity summing to zero. For instance, among the 664578 odd primes $p \leq 10^8$, all but 550 are ruled out by Lemma 23. (There is an easy argument that t must be odd if $p > 3$; this cuts the list down to 273 primes.) Each remaining p can be tested by computing $\gcd(x^p + 1, (x+1)^p + 1)$; the only examples we found that did not satisfy the condition of lemma 8 were $(p,t) = (73,9), (262657,27), (599479,33), (121369,39)$.

5.4 Summary

In the beginning of this section 5 we argued that in order to use the method of [35], (i.e., proposition 5) to obtain k-query locally decodable codes of length $\exp(n^\epsilon)$ for some fixed k and all $\epsilon > 0$, one needs to exhibit a k-nice sequence of subsets of finite fields. In what follows we use technical results of the previous subsections to show that the existence of a k-nice sequence implies that infinitely many Mersenne numbers have large prime factors.

Theorem 24 *Let k be odd. Suppose there exists a k-nice sequence of subsets of finite fields; then for infinitely many values of t we have*

$$P(2^t - 1) \geq (t/2)^{1+1/(k-2)}. \tag{41}$$

Proof: Using lemmas 21 and 22 we conclude that a k-nice sequence yields infinitely many primes p such that $\mathrm{ord}_p(2) \leq 2p^{1-1/(k-1)}$. Let p be such a prime and $t = \mathrm{ord}_p(2)$. Then $P(2^t - 1) \geq (t/2)^{1+1/(k-2)}$. ■

A combination of lemmas 21 and 23 yields a slightly stronger bound for the special case of 3-nice sequences.

Theorem 25 *Suppose there exists a 3-nice sequence of subsets; then for infinitely many values of t we have*

$$P(2^t - 1) \geq (3/4)t^2. \tag{42}$$

We would like to remind the reader that although the lower bounds for $P(2^t - 1)$ given by (41) and (42) are extremely weak light of the widely accepted conjecture saying that the number of Mersenne primes is infinite, they are substantially stronger than what is currently known unconditionally (2).

Acknowledgements

Kiran Kedlaya's research is supported by NSF CAREER grant DMS-0545904 and by the Sloan Research Fellowship. Sergey Yekhanin would like to thank Swastik Kopparty for providing the reference [5] and outlining the proof of lemma 9. He would also like to thank Henryk Iwaniec, Carl Pomerance and Peter Sarnak for their feedback regarding the number theory problems discussed in this paper.

References

[1] M. Agrawal, N. Kayal, and N. Saxena, "PRIMES is in P," *Annals of Mathematics*, vol. 160, pp. 781-793, 2004.

[2] L. Babai, L. Fortnow, L. Levin, and M. Szegedy, "Checking computations in polylogarithmic time," In *Proc. of the 23th ACM Symposium on Theory of Computing (STOC)*, pp. 21-31, 1991.

[3] A. Beimel, Y. Ishai and E. Kushilevitz,"General constructions for information-theoretic private information retrieval," *Journal of Computer and System Sciences*, vol. 71, pp. 213-247, 2005. Preliminary versions in STOC 1999 and ICALP 2001.

[4] A. Beimel, Y. Ishai, E. Kushilevitz, and J. F. Raymond, "Breaking the $O\left(n^{1/(2k-1)}\right)$ barrier for information-theoretic private information retrieval," In *Proc. of the 43rd IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 261-270, 2002.

[5] J. Bourgain and M. Chang, "A Gauss sum estimate in arbitrary finite fields," *Comptes Rendus Mathematique*, vol. 342, pp. 643-646, 2006.

[6] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," In *Proc. of the 36rd IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 41-50, 1995. Also, in *Journal of the ACM*, vol. 45, 1998.

[7] D. Cox, J. Little, D. O'Shea, Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra. Springer, 1996.

[8] A. Deshpande, R. Jain, T. Kavitha, S. Lokam and J. Radhakrishnan, "Better lower bounds for locally decodable codes," In *Proc. of the 20th IEEE Computational Complexity Conference (CCC)*, pp. 184-193, 2002.

[9] Curtis Cooper, Steven Boone, http://www.mersenne.org/32582657.htm

[10] P. Erdos and T. Shorey, "On the greatest prime factor of $2^p - 1$ for a prime p and other expressions," *Acta. Arith.* vol. 30, pp. 257-265, 1976.

[11] K. Ford, F. Luca and I. Shparlinski, "On the largest prime factor of the Mersenne numbers," *a preprint,* arXiv:0704.1327, 2007.

[12] W. Gasarch, "A survey on private information retrieval," *The Bulletin of the EATCS*, vol. 82, pp. 72-107, 2004.

[13] O. Goldreich, "Short locally testable codes and proofs," Technical Report TR05-014, *Electronic Colloquim on Computational Complexity* (ECCC), 2005.

[14] O. Goldreich, H. Karloff, L. Schulman, and L. Trevisan, "Lower bounds for locally decodable codes and private information retrieval," In *Proc. of the 17th IEEE Computational Complexity Conference (CCC)*, pp. 175-183, 2002.

[15] B. Hemenway and R. Ostrovsky, "Public key encryption which is simultaneously a locally-decodable error-correcting code," In *Cryptology ePrint Archive*, Report 2007/083.

[16] J. Katz and L. Trevisan, "On the efficiency of local decoding procedures for error-correcting codes," In *Proc. of the 32th ACM Symposium on Theory of Computing (STOC)*, pp. 80-86, 2000.

[17] I. Kerenidis and R. de Wolf, "Exponential lower bound for 2-query locally decodable codes via a quantum argument," *Journal of Computer and System Sciences,* 69(3), pp. 395-420. Earlier version in STOC'03. quant-ph/0208062.

[18] R. Lidl and H. Niederreiter, Finite Fields. Cambridge: Cambridge University Press, 1983.

[19] L. Murata and C. Pomerance, "On the largest prime factor of a Mersenne number," *Number theory, CRM Proc. Lecture Notes of American Mathematical Society* vol. 36, pp. 209-218, 2004.

[20] M. Murty and S. Wong, "The ABC conjecture and prime divisors of the Lucas and Lehmer sequences," In *Proc. of Milennial Conference on Number Theory III,* (Urbana, IL, 2000) (A. K. Peters, Natick, MA, 2002) pp. 43-54.

[21] K. Obata, "Optimal lower bounds for 2-query locally decodable linear codes," In *Proc. of the 6th RANDOM*, vol. 2483 of Lecture Notes in Computer Science, pp. 39-50, 2002.

[22] A. Polishchuk and D. Spielman, "Nearly-linear size holographic proofs," In *Proc. of the 26th ACM Symposium on Theory of Computing (STOC)*, pp. 194-203, 1994.

[23] C. Pomerance, "Recent developments in primality testing," *Math. Intelligencer,* 3:3, pp. 97-105, (1980/81).

[24] P. Raghavendra, "A Note on Yekhanin's locally decodable codes," In *Electronic Colloquium on Computational Complexity* Report TR07-016, 2007.

[25] A. Romashchenko, "Reliable computations based on locally decodable codes," In *Proc. of the 23rd International Symposium on Theoretical Aspects of Computer Science (STACS)*, vol. 3884 of Lecture Notes in Computer Science, pp. 537-548, 2006.

[26] A. Schinzel, "On primitive factors of $a^n - b^n$," In *Proc. of Cambridge Philos. Soc.* vol. 58, pp. 555-562, 1962.

[27] C. Stewart, "The greatest prime factor of $a^n - b^n$," *Acta Arith.* vol. 26, pp. 427-433, 1974/75.

[28] C. Stewart, "On divisors of Fermat, Fibonacci, Lucas, and Lehmer numbers," In *Proc. of London Math. Soc.* vol. 35 (3), pp. 425-447, 1977.

[29] M. Sudan, Efficient checking of polynomials and proofs and the hardness of approximation problems. PhD thesis, University of California at Berkeley, 1992.

[30] L. Trevisan, "Some applications of coding theory in computational complexity," *Quaderni di Matematica,* vol. 13, pp. 347-424, 2004.

[31] S. Wehner and R. de Wolf, "Improved lower bounds for locally decodable codes and private information retrieval," In *Proc. of 32nd International Colloquium on Automata, Languages and Programming (ICALP'05)*, LNCS 3580, pp. 1424-1436.

[32] Lenstra-Pomerance-Wagstaff conjecture. (2006, May 22). In Wikipedia, The Free Encyclopedia. Retrieved 00:18, October 3, 2006, from http://en.wikipedia.org/w/index.php?title=Lenstra-Pomerance-Wagstaff_conjecture&oldid=54506577

[33] S. Wagstaff, "Divisors of Mersenne numbers," *Math. Comp.*, 40:161, pp. 385-397, 1983.

[34] D. Woodruff, "New lower bounds for general locally decodable codes," *Electronic Colloquium on Computational Complexity,* TR07-006, 2007.

[35] S. Yekhanin, "Towards 3-query locally decodable codes of subexponential length," In *Proc. of the 39th ACM Symposium on Theory of Computing (STOC)*, 2007.

[36] S. Yekhanin, Locally decodable codes and private information retrieval schemes. PhD thesis, MIT, to appear.