A Deterministic Construction of a Large Distance Code from the Wozencraft Ensemble

Venkatesan Guruswami ✉ 🏠 ⓘ
Department of EECS, University of California, Berkeley, CA, USA

Shilun Li ✉ ⓘ
Department of Mathematics, University of California, Berkeley, CA, USA

—— **Abstract** ——

We present an explicit construction of a sequence of rate $1/2$ Wozencraft ensemble codes (over any fixed finite field $\mathbb{F}_q$) that achieve minimum distance $\Omega(\sqrt{k})$ where k is the message length. The coefficients of the Wozencraft ensemble codes are constructed using Sidon Sets and the cyclic structure of $\mathbb{F}_{q^k}$ where $k+1$ is prime with q a primitive root modulo $k+1$. Assuming Artin's conjecture, there are infinitely many such k for any prime power q.

1 Introduction

The explicit construction of binary error-correcting codes with a rate vs. distance trade-off approaching that of random constructions, i.e., the so-called Gilbert-Varshamov (GV) bound, remains an outstanding challenge in coding theory and combinatorics.

For large n, a random binary linear code of rate $R \in (0, 1)$, defined for example as the column span of a random matrix $G \in \mathbb{F}_2^{n \times Rn}$, has relative distance $h^{-1}(1 - R)$ with high probability, where $h^{-1}(\cdot)$ is the inverse of the binary entropy function. There is a similar GV bound $h_q^{-1}(1 - R)$, involving the q-ary entropy function, for codes over other finite fields $\mathbb{F}_q$.

While explicit constructions meeting the GV bound remain elusive[1], there are known derandomizations showing that codes drawn randomly from much smaller, structured ensembles can also achieve the GV bound. One of the most classical and famous such ensemble is the Wozencraft ensemble, which consists of codes $C_{\mathrm{WE}}^\alpha = \{(x, \alpha x) : x \in \mathbb{F}_{q^k}\}$ as α varies over nonzero elements of the field $\mathbb{F}_{q^k}$, and one uses some fixed basis to express elements of $\mathbb{F}_{q^k}$ as length k vectors over $\mathbb{F}_q$. Note that each code C_{WE}^α has rate $1/2$. The construction of Wozencraft ensemble codes C_{WE}^α first appeared in a paper by Massey [13], who attributed the discovery of these codes to John M. Wozencraft.

---

[1] Ta-Shma [19] recently constructed explicit binary codes near the GV bound for low rates. The codes have distance $\frac{1-\epsilon}{2}$ and rate $\Omega(\epsilon^{2+o(1)})$ which is asymptotically close to the rate $\Omega(\epsilon^2)$ guaranteed by the GV bound.

It is a standard exercise to show that for most choices of α, the code C_{WE}^{α} has distance close to $h_q^{-1}(1/2)$ and thus achieves the GV bound. Puncturing the Wozencraft ensemble gives codes of higher rates that also meet the GV bound. The property of the Wozencraft ensemble and its punctured variants behind this phenomenon is that every nonzero word appears as a codeword in an equal number of codes in the ensemble (if it appears in any code of the ensemble at all).

Using this property, Justesen [10] in 1972 gave the first strongly explicit asymptotically good binary linear codes, by concatenating an outer Reed–Solomon code with different inner codes drawn from the Wozencraft ensemble for different positions of the Reed-Solomon code. The Justesen construction achieves a trade-off between rate vs. distance called the Zyablov bound for rates at least 0.3. There are variants of Wozencraft codes which give Zyablov-bound-achieving codes for lower rates as well (see Section 5). In recent years, Wozencraft ensemble codes have also found other varied uses, for example in constructing covering codes of small block length [14] and write-once-memory codes [16, 17].

Given that for most α, the code C_{WE}^{α} meets the GV bound, it is a natural question whether one can find an explicit α for which the code has good distance (even if it doesn't quite meet the GV bound). Gaborit and Zemor [8] showed that it suffices to consider random α in a subset of size $\approx q^k/k$ and used it to show the existence of linear codes which are a factor k larger in terms of size than the Gilbert-Varshamov bound (such a result was shown earlier for general codes in [9]).

However, it remains an outstanding challenge to find some α in deterministic poly(k) time for which C_{WE}^{α} has distance $\Omega(k)$. This question is relatively well-known, eg. it received mention in a blog post by Dick Lipton [12], but has resisted progress. To the best of our knowledge, even an explicit α for which C_{WE}^{α} has distance $k^{\Omega(1)}$ was not known.

For certain structured fields $\mathbb{F}_{q^k}$ (of which there are an infinite family under Artin's conjecture), we give an explicit construction of $\alpha \in \mathbb{F}_{q^k}$ for which C_{WE}^{α} has distance $\Omega(\sqrt{k})$. We also give an explicit puncturing of these codes to achieve any desired rate $r < 1$, and $\Omega_r(\sqrt{k})$ distance (the constant in the $\Omega()$ depends on r). Our theorems are informally stated below.

▶ **Theorem 1** (Informal). *Fix a field $\mathbb{F}_q$ and consider an integer k such that $k + 1$ is prime and q is a primitive root modulo $k + 1$. There exist $\alpha^* \in \mathbb{F}_{q^k}$ which can be constructed in deterministic poly(k) time such that:*
- *$C_{WE}^{\alpha^*}$ has distance $\Omega(\sqrt{k})$.*
- *For any $r > \frac{1}{2}$, there is an explicit puncturing of $C_{WE}^{\alpha^*}$ with rate at least r and distance $\Omega_r(\sqrt{k})$.*

Please refer to Theorem 7 and Theorem 17 for construction of α^* and choice of puncturing.

2 Preliminaries

Throughout this paper, we will assume the alphabet has size q, where q is a prime power. Furthermore, we will assume k' is a prime such that q is a primitive root modulo k'. Assuming Artin's conjecture, such k' exists infinitely often at sufficiently high density and can be efficiently found in deterministic poly(k') time.

Denote $k = k' - 1$ for ease of notation. Let $p(x) = 1 + x + x^2 + \ldots + x^{k'-1}$ be the cyclotomic polynomial which is irreducible over $\mathbb{F}_q$ (see Proposition 3). Note that $\mathbb{F}_q[x]/(p(x)) \cong \mathbb{F}_{q^k}$ for the extension field $\mathbb{F}_q[x]/(p(x))$ and we will fix the representation of $\mathbb{F}_{q^k}$ as polynomials in x of degree less than k, with operations performed modulo $p(x)$. We will fix the $\mathbb{F}_q$-linear isomorphism $\varphi : \mathbb{F}_{q^k} \to \mathbb{F}_q^k$ that maps a polynomial of degree less than k to its coefficient vector. That is, $\varphi(\sum_{i=0}^{k-1} a_i x^i) = (a_0, a_1, \ldots, a_{k-1})$. For $\alpha \in \mathbb{F}_{q^k}$, define wt($\alpha$) to be the Hamming weight of $\varphi(\alpha)$.

The Wozencraft ensemble is a classic family of codes defined as follows.

▶ **Definition 2.** *For $\alpha \in \mathbb{F}_{q^k}$, the **Wozencraft ensemble code** C^α_{WE} parameterized by α is given by*

$$C^\alpha_{WE} = \{(\varphi(x), \varphi(\alpha x)) : x \in \mathbb{F}_{q^k}\} \ .$$

Note that C^α_{WE} is an $\mathbb{F}_q$-linear code of rate $1/2$.

▶ **Proposition 3.** *If k' is a prime such that q is a primitive root modulo k', then $p(x) = \sum_{i=0}^{k'-1} x^i$ is an irreducible polynomial of degree $k' - 1$ in $\mathbb{F}_q[x]$.*

Proof. Since q is a primitive root modulo k', $d = k' - 1$ is the smallest integer satisfying $q^d \equiv 1 \mod k'$. Note that a field extension $\mathbb{F}_{q^d}$ of $\mathbb{F}_q$ contains a primitive k'-th root of unity ζ if and only if $k'|q^d - 1$, i.e., $q^d \equiv 1 \mod k'$. So $\mathbb{F}_{q^{k'-1}}$ is the smallest field extension of $\mathbb{F}_q$ containing ζ and the mimimal polynomial of ζ has degree $k' - 1$. Since $p = \sum_{i=0}^{k'-1} x^i$ is a degree $k' - 1$ polynomial such that $p(\zeta) = 0$, p is the minimal polynomial of ζ and thus irreducible. ◀

We will use Sidon sets [18, 1] to construct the parameter α for C^α_{WE}.

▶ **Definition 4.** *A **Sidon set** is a set of integers $A = \{a_1, \ldots, a_d\}$ where $a_1 < a_2 < \ldots < a_d$ such that for all $i, j, k, l \in [d]$ with $i \neq j$ and $k \neq l$,*

$$a_i - a_j = a_k - a_l \Longleftrightarrow i = k \text{ and } j = l.$$

*A **Sidon set modulo n** is a **Sidon set** such that for all $i, j, k, l \in [d]$ with $i \neq j$ and $k \neq l$,*

$$a_i - a_j \equiv a_k - a_l \pmod{n} \Longleftrightarrow i = k \text{ and } j = l.$$

*Size d of the Sidon set A is referred to as its **order** and $a_d - a_1$ as its **length**.*

▶ **Remark 5.** For any Sidon set with order d and length m, the $\binom{d}{2}$ distances between each pair of points need to be distinct. So $m \geq \binom{d}{2}$ and this gives a trivial upper bound $d \leq \sqrt{2m}$. This upper bound on d can be improved to $d \leq \sqrt{m} + O(m^{1/4})$ [7] and further to $d \leq \sqrt{m} + m^{1/4} + 1$ [11]. On the other hand, the maximal d given m satisfies $d \geq \sqrt{m} - O(m^{5/16})$ [7] but it is believed that we can have $d > \sqrt{m}$ [6].

We will introduce the Bose-Chowla construction of Sidon sets [4, 3].

▶ **Theorem 6** (Bose-Chowla, [4]). *Let p be a power of a prime, g be a primitive root in $\mathbb{F}_{p^2}$. Then the sequence of p integers*

$$A = \{i \in [p^2 - 2] : g^i + g^{pi} = 1\}$$

forms a Sidon set modulo $p^2 - 1$.

This construction of Sidon set has order $d = p$ and length at most $m(d) = p^2 - 2$. They are asymptotically optimal in the sense that $\lim_{d \to \infty} \sqrt{m(d)}/d = 1$. Given p, such construction can be done in $O(p^2)$ time by fixing $s = 1$ and finding a primitive root g via naive search. We will later define our parameter α to be $\sum_{a \in A} x^a$ where A is a Sidon set and show that C^α_{WE} has good distance.

3   **Rate 1/2 Construction**

In this section, we will give explicit construction of rate $1/2$ Wozencraft ensemble codes with minimum distance $\Omega(\sqrt{k})$ using Sidon sets. To begin, we provide an intuitive explanation for the natural occurrence of Sidon sets in this specific context. Subsequently, we proceed with the analysis of the minimum distance of our construction.

3.1   Motivation

Fix a set of indices $A \subseteq [k]$ and an element $\alpha = \sum_{a \in A} x^a \in \mathbb{F}_{q^k}$ with coefficients either 0 or 1. Take any $y = \sum_{s \in S} b_s x^s$ where S is the set of non-zero indices of the coefficients of y, so $b_s \neq 0$ for all $s \in S$. To establish a lower bound on $\Delta(C_{\mathrm{WE}}^{\alpha})$, we would like show a lower bound on the weight of the product $\alpha y \in \mathbb{F}_{q^k}$ for any y. To simplify the analysis, we consider a ring extension of $\mathbb{F}_{q^k}$ (described in Section 3.2), in which the coefficient c_j in front of x^j of the product αy can be expressed as $c_j = \sum_{a \in A, s \in S} \mathbf{1}\{a + s \equiv j \pmod{k'}\} b_s$. We would like to establish a lower bound on the number of non-zero coefficients $c_j \neq 0$, which would transform into a lower bound on the weight of αy in $\mathbb{F}_{q^k}$. It is sufficient to demonstrate the existence of numerous choices of j satisfying $j \equiv a + s \pmod{k'}$ for a unique combination of a and s. In this case, c_j corresponds to the sum of one non-zero element and is therefore non-zero. To ensure there are an abundance of such choices for j with this uniqueness property, it is desirable to minimize collisions of the form $a + s \equiv a' + s' \pmod{k'}$ where $a, a' \in A$ and $s, s' \in S$. Since S can be selected adversarially with respect to A, it is advantageous to have $(a - a') \bmod k'$ be unique, which is exactly the property of Sidon sets modulo k'. The result of this construction is stated formally as follows, where the proof is propounded to Section 3.2:

▶ **Theorem 7.** *Let d be the largest prime smaller than $\sqrt{k}$ and let $A = \{a_1, \ldots, a_d\}$ be a Bose-Chowla Sidon set with order d. Define $\alpha^* = \sum_{a \in A} x^a \in \mathbb{F}_{q^k}$. Then $\Delta(C_{WE}^{\alpha^*}) \geq d = \Omega(\sqrt{k})$.*

▶ Remark 8. Since there exists a prime between $[\frac{1}{2}\sqrt{k}, \sqrt{k}]$ by Bertrand–Chebyshev theorem [5], d can be found efficiently via naive search and $d \geq \frac{1}{2}\sqrt{k}$. Moreover, Baker, Harman and Pintz [2] showed that there exists a prime in the interval $[\sqrt{k} - k^{0.27}, \sqrt{k},]$ for k sufficiently large. So $d = (1 - o(1))\sqrt{k}$ and the constructed code $C_{\mathrm{WE}}^{\alpha^*}$ has distance asymptotically $\Delta(C_{\mathrm{WE}}^{\alpha^*}) \geq (1 - o(1))\sqrt{k}$ as $k \to \infty$.

It is also worthwhile to note that when $a + s \equiv j \pmod{k'}$ holds for more than one pair of (a, s), c_j may still be non-zero as it is a sum of multiple non-zero elements. In fact, for α with large weight, it is common that αy has substantial weight yet few choices of j satisfy the uniqueness property. One possible approach to improving the construction of A involves analyzing scenarios where $j \equiv a + s \pmod{k'}$ for multiple pairs of (a, s).

3.2   Proof of Theorem 7

To analyze the minimum distance of C_{WE}^{α}, it is helpful to define the ring $R = \mathbb{F}_q[x]/(x^{k'} - 1)$, which consists of polynomials of degree less than $k' = k + 1$. We can identify $\mathbb{F}_{q^k} \cong R/(p)$ by the map sending $f \in R$ to $(f \bmod p) \in \mathbb{F}_{q^k}$. In addition, we can consider $\mathbb{F}_{q^k} \subseteq R$ and extend φ to the $\mathbb{F}_q$-linear map $\widetilde{\varphi} : R \to \mathbb{F}_q^{k'}$ mapping polynomials of degree less than k' to its coefficient vector. Define $\widetilde{\mathrm{wt}}(f)$ to be the Hamming weight of $\widetilde{\varphi}(f)$ for any $f \in R$. The following lemma gives the relationship between $\widetilde{\mathrm{wt}}(f)$ and $\mathrm{wt}(f \bmod p)$.

▶ **Lemma 9.** *For any $f \in R$, $\mathrm{wt}(f \bmod p) \geq \min\{\widetilde{\mathrm{wt}}(f), k - \widetilde{\mathrm{wt}}(f)\}$.*

Proof. For any $f \in R$, let us write $f = \sum_{i=0}^{k} b_i x^i$. Then

$$f \bmod p = f - b_k p = \sum_{i=0}^{k-1} (b_i - b_k) x^i.$$

If $b_k = 0$, then $\mathrm{wt}(f \bmod p) = \widetilde{\mathrm{wt}}(f)$; if $b_k \neq 0$, then

$$\mathrm{wt}(f \bmod p) = |\{i : b_i \neq b_k\}| \geq k - \widetilde{\mathrm{wt}}(f).$$

So $\mathrm{wt}(f \bmod p) \geq \min\{\widetilde{\mathrm{wt}}(f), k - \widetilde{\mathrm{wt}}(f)\}$. ◀

▶ **Lemma 10.** *Given $\alpha \in \mathbb{F}_{q^k} \subseteq R$, suppose that for every $y \in R$ with $\widetilde{\mathrm{wt}}(y) \leq c(k)$ the condition*

$$c(k) - \widetilde{\mathrm{wt}}(y) \leq \widetilde{\mathrm{wt}}(\alpha y) \leq k - \left(c(k) - \widetilde{\mathrm{wt}}(y) \right)$$

holds, where the product αy is taken in R. Then $\Delta(C_{WE}^\alpha) \geq c(k)$, where $\Delta(C_{WE}^\alpha)$ denotes the distance of the code C_{WE}^α.

Proof. Take any non-zero $y \in \mathbb{F}_{q^k} \subseteq R$. Its corresponding codeword $C_{\mathrm{WE}}^\alpha(y) = (\varphi(y), \varphi(\alpha y \bmod p))$ has Hamming weight $\mathrm{wt}(y) + \mathrm{wt}(\alpha y \bmod p)$. By Lemma 9, the above condition implies

$$\mathrm{wt}(y) + \mathrm{wt}(\alpha y \bmod p) \geq \widetilde{\mathrm{wt}}(y) + \min\{\widetilde{\mathrm{wt}}(\alpha y), k - \widetilde{\mathrm{wt}}(\alpha y)\} \geq c(k).$$

Since C_{WE}^α is a linear code, $\Delta(C_{\mathrm{WE}}^\alpha) \geq c(k)$. ◀

We can now prove Theorem 7 by showing that α^* satisfies the condition of Lemma 10 with $c(k) = d$.

Proof of Theorem 7. Note that α is an element of $\mathbb{F}_{q^k}$ since it has degree at most $k - 2$ by construction. Let us check that the condition of Lemma 10 indeed holds for $\alpha^* = \sum_{a \in A} x^a$ and $c(k) = d$. For any $y \in R$ with $\widetilde{\mathrm{wt}}(y) = w$, we can write $y = \sum_{i=1}^{w} b_{s_i} x^{s_i}$ where $b_{s_i} \neq 0$ for all $1 \leq i \leq w$. We will denote $S = \{s_1, \ldots, s_w\} \subseteq [k']$ the set non-zero coefficient indices of y, where we define $[k'] = \{0, \ldots, k' - 1 = k\}$. The coefficients of the product $\alpha^* y = \sum_{j=0}^{k} c_j x^j$ are given by

$$c_j = \sum_{\substack{a \in A \\ s \in S}} \mathbf{1}\{a + s \equiv j \pmod{k'}\} b_s.$$

This motivates us to define the shifted set

$$(j - A)_k = \{(j - a_1) \bmod k', \ldots, (j - a_d) \bmod k'\}$$

which gives us

$$c_j = \sum_{s \in (j-A)_k \cap S} b_s.$$

We will denote by

$$J_m = \{j \in [k'] : |(j - A)_k \cap S| = m\}$$

the indices j such that $|(j - A)_k \cap S|$ has size m. It is evident that if $(j - A)_k \cap S = \emptyset$ then $c_j = 0$, and if $|(j - A)_k \cap S| = 1$ then $c_j \neq 0$. So

$$|J_1| \leq \widetilde{\text{wt}}(\alpha^* y) \leq k - |J_0|.$$

Looking at the conditions of Lemma 10, it would be sufficient to take $c(k)$ such that

$$c(k) - w \leq \min\{|J_0|, |J_1|\}$$

for all w. We make the following claims on lower bounds of $|J_0|$ and $|J_1|$ which will be proven in Section 3.3.

▷ **Claim 11.** J_1 has size at least $wd - w(w - 1)$.

▷ **Claim 12.** J_0 has size at least $k - wd$.

Assuming the claims, it suffices to take $c(k)$ such that

$$c(k) \leq \min_{1 \leq w \leq c(k)} wd - w(w - 1) + w = \min\{d + 1, (d - c(k) + 2)c(k)\},$$

$$c(k) \leq \min_{1 \leq w \leq c(k)} k - wd + w = k - (d - 1)c(k).$$

Solving the two inequalities, it suffices to take $c(k) \leq d$. So $\alpha^* = \sum_{a \in A} x^a$, $c(k) = d$ satisfy the condition of Lemma 10.   ◄

3.3   Proofs of Claim 11 and Claim 12

To show a lower bound on $|J_1|$, it is desired that the sets $J_2, .., J_w$ are small. The following lemma gives an upper bound on the sizes of $J_2, .., J_w$:

▶ **Lemma 13.** *The sets $J_2, \ldots, J_w$ defined in the proof of Theorem 7 satisfy:*

$$\sum_{m=2}^{w} (m - 1)|J_m| \leq \sum_{m=2}^{w} \binom{m}{2}|J_m| \leq 2\binom{w}{2} = w(w - 1).$$

Proof. It is evident that the first inequality holds. Let us now show that for any two distinct $s, s' \in S$, we have $\{s, s'\} \subseteq (j - A)_k$ for at most two choices of j. Take any distinct pair $s, s' \in S$. Without loss of generality, assume $s < s'$. Suppose $\{s, s'\} \subseteq (j - A)_k$, then we can write $s \equiv j - a_l \pmod{k'}$ and $s' \equiv j - a_{l'} \pmod{k'}$ for some $l, l' \in [d]$. Then $a_l - a_{l'} \equiv s' - s \pmod{k'}$. So

$$a_l - a_{l'} = \begin{cases} s' - s & \text{if } l > l', \\ s' - s - k' & \text{if } l < l'. \end{cases}$$

In both cases, by definition of Sidon sets, l, l' are uniquely determined. Thus $j \equiv a_l + s$ is also uniquely determined.

For any $j_m \in J_m$, $|(j_m - A) \cap S| = m$ so there are $\binom{m}{2}$ distinct pairs $\{s, s'\} \subseteq (j_m - A)_k$. So the total count of such distinct pairs for all $j \in \cup_{m=2}^{w} J_m$ is $\sum_{m=2}^{w} \binom{m}{2}|J_m|$, which must not exceed $2\binom{w}{2}$ since each pair can only occur in $(j - A)_k$ for two choices of j. This gives

$$\sum_{m=2}^{w} \binom{m}{2}|J_m| \leq 2\binom{w}{2} = w(w - 1)$$

which completes the proof.   ◄

We can now obtain a lower bound on the size of J_1.

▷ **Claim.** J_1 has size at least $wd - w(w-1)$.

Proof. For each $s \in S$, there are $|A| = d$ number of j such that $s \in (j - A)_k$. When $w = 1$, the lemma holds as $|J_1| = d$. When $w \geq 2$, $|J_1|$ equals $|S|d = wd$ minus the times we overcount:

$$|J_1| = wd - \sum_{m=2}^{w} (m-1)|J_m|.$$

The proof is complete via the bound of Lemma 13:

$$|J_1| = wd - \sum_{m=2}^{w} (m-1)|J_m| \geq wd - w(w-1). \qquad \triangleleft$$

▷ **Claim.** J_0 has size at least $k - wd$.

Proof. For any $s \in S$, there are $|A| = d$ values of $j \in [k']$ such that $s \in (j - A)_k$. So there are at most $|S|d = wd$ indices j such that $|(j - A)_k \cap S| > 0$. $\qquad \triangleleft$

4   Higher Rate Construction

In the last section, we gave an explicit construction of rate $1/2$ Wozencraft ensemble code $C_{\mathrm{WE}}^{\alpha^*}$ achieving minimum distance $\Omega(\sqrt{k})$. In this section, we will show that appropriate puncturing of $C_{\mathrm{WE}}^{\alpha^*}$ will give us codes of rates $r \in (1/2, 1)$ and minimum distance at least $\Omega\left(\left(1 - \sqrt{2 - \frac{1}{r}}\right)\sqrt{k}\right)$.

▶ **Definition 14.** *Let $C_{WE}^{\alpha,r}$ be the rate r punctured code given by removing the last $(2 - \frac{1}{r})k$ message bits of each codeword in C_{WE}^{α}.*

Let φ_r denote the map sending any polynomial f to the first $(\frac{1}{r} - 1)k$ least significant coefficients. Let $\mathrm{wt}_r(f)$ denote the Hamming weight of $\varphi_r(f)$.

▶ **Lemma 15.** *For any $f \in R$, $\mathrm{wt}_r(f \bmod p) \geq \min\{\mathrm{wt}_r(f), (\frac{1}{r} - 1)k - \mathrm{wt}_r(f)\}$.*

Proof. Writing $f = \sum_{i=0}^{k} b_i x^i$, we have

$$f \bmod p = f - b_k p = \sum_{i=0}^{k-1} (b_i - b_k)x^i.$$

If $b_k = 0$, then $\mathrm{wt}_r(f \bmod p) = \mathrm{wt}_r(f)$; if $b_k \neq 0$, then

$$\mathrm{wt}_r(f \bmod p) = \left|\left\{i \in [(\frac{1}{r} - 1)k] \mid b_i \neq b_k\right\}\right| \geq (\frac{1}{r} - 1)k - \mathrm{wt}_r(f).$$

So $\mathrm{wt}_r(f \bmod p) \geq \min\{\mathrm{wt}_r(f), (\frac{1}{r} - 1)k - \mathrm{wt}_r(f)\}$. $\qquad \blacktriangleleft$

▶ **Lemma 16.** *Let $\alpha \in \mathbb{F}_{q^k} \subseteq R$. Suppose for every $y \in R$ with $\widetilde{\mathrm{wt}}(y) \leq c(k)$ the condition*

$$c(k) - \widetilde{\mathrm{wt}}(y) \leq \mathrm{wt}_r(\alpha y) \leq (\frac{1}{r} - 1)k - \left(c(k) - \widetilde{\mathrm{wt}}(y)\right)$$

holds, where the product αy is taken in R. Then $\Delta(C_{WE}^{\alpha,r}) \geq c(k)$.

Proof. Take any non-zero $y \in \mathbb{F}_{q^k} \subseteq R$. Its corresponding codeword $C_{\mathrm{WE}}^{\alpha,r}(y) = (\varphi(y),$ $\varphi_r(\alpha y \mod p))$ has hamming weight $\mathrm{wt}(y) + \mathrm{wt}_r(\alpha y \mod p)$. By Lemma 15, the above condition implies

$$\mathrm{wt}(y) + \mathrm{wt}(\alpha y \mod p) \geq \widetilde{\mathrm{wt}}(y) + \min\{\mathrm{wt}_r(\alpha y), (\frac{1}{r} - 1)k - \mathrm{wt}_r(\alpha y)\} \geq c(k).$$

Since $C_{\mathrm{WE}}^{\alpha,r}$ is a linear code, $\Delta(C_{\mathrm{WE}}^{\alpha,r}) \geq c(k)$.   ◄

The following theorem establishes a lower bound on the minimum distance of the punctured code $C_{\mathrm{WE}}^{\alpha^*,r}$ with α^* constructed using Sidon sets as outlined in Theorem 7.

▶ **Theorem 17.** *For any rate $r > 1/2$, using the construction of α^* by Theorem 7, the condition of Lemma 16 is satisfied with $c(k) = \Omega\left(\left(1 - \sqrt{2 - \frac{1}{r}}\right)\sqrt{k}\right)$. Thus the punctured code $C_{WE}^{\alpha^*,r}$ has minimum distance at least $\Omega\left(\left(1 - \sqrt{2 - \frac{1}{r}}\right)\sqrt{k}\right)$.*

The proof is omitted[2] since it follows a similar outline as the proof of Theorem 7. We present two claims similar to Claim 11 and Claim 12 that substantiate this argument. For any $y = \sum_{s \in S} b_s x^s$ with weight $|S| = w$, denote $J_m^r = \{j \in [(\frac{1}{r} - 1)k] : |(j - A)_k \cap S| = m\}$ the non-punctured indices j such that $|(j - A)_k \cap S|$ has size m, where A is the Sidon set in the construction of α^*. Then analogous to Claim 11 and Claim 12, we can lower bound the size of J_1^r and J_2^r:

▷ Claim 18. J_1^r has size at least $w\left(\left(1 - \sqrt{(2 - \frac{1}{r})} - o(1)\right)\sqrt{k} - (2 - \frac{1}{r})^{\frac{1}{4}} k^{\frac{1}{4}} - 1\right) - w^2$.

▷ Claim 19. J_0^r has size at least $(\frac{1}{r} - 1)k - w\left(\sqrt{(\frac{1}{r} - 1)k} + (\frac{1}{r} - 1)^{\frac{1}{4}} k^{\frac{1}{4}} + 1\right)$.

To justify the claims, we will require the following theorem by Lindström [11] which bounds the order of any Sidon set via its length.

▶ **Theorem 20** (Lindström [11]). *For any Sidon set with length m, the order is at most $\sqrt{m} + m^{1/4} + 1$.*

The key observation is that for any $s \in S$, the shifted set $(s + A)_k$ is a Sidon set. Moreover, it is partitioned into two parts after puncturing: the remaining part $(s + A)_k \cap [(\frac{1}{r} - 1)k]$ and the removed part $(s + A)_k \cap ([k'] \setminus [(\frac{1}{r} - 1)k])$, where each part is itself a Sidon set. Applying Theorem 20 to the two parts and recalling that $|A| = (1 - o(1))\sqrt{k}$ by Remark 8, we can then obtain the results described in Claim 18 and Claim 19.

5   Open Questions

It is well known that Wozencraft ensemble codes C_{WE}^{α} satisfy the Gilbert-Varshamov bound for most α. Concretely,

$$\lim_{k \to \infty} \mathbf{Pr}_{\alpha \in S} [\Delta(C_{\mathrm{WE}}^{\alpha}) \geq d(k)] = 1,$$

where $S = \mathbb{F}_{q^k}^*$, $d(k) = \left(h_q^{-1}(\frac{1}{2}) - \epsilon\right) \cdot 2k$ with h_q the q-ary entropy function and $\epsilon > 0$ chosen arbitrarily. In this paper, we have proposed a construction of $S = \{\alpha^*\}$ such that the equation above holds with $d(k) = \Omega(\sqrt{k})$. It is of natural interest to reduce the size of the

---

[2] The proof appears in detail in the full version of the paper posted at https://arxiv.org/abs/2305.02484.

ensemble and find $S, d(k)$ satisfying the equation above with $|S|$ small and $d(k)$ large. For example, can one construct S such that the equation above is satisfied via $|S| = O(2^{o(k)})$ and $d(k) = \Omega(k)$, or $|S| = \text{poly}(k)$ and $d(k) = \Omega(k^c)$ with $c > \frac{1}{2}$? In addition, are there any barriers to such constructions, as in, would such constructions imply progress on some other explicit construction challenge?

In 1973, Weldon [20] proposed an ensemble of codes, a code that generalizes Wozencraft ensemble codes. The Weldon ensemble was used by him, and later Shen [15], to construct explicit concatenated codes achieving the Zyablov bound for rates less than 0.3, thus improving upon Justesen codes [10] for low rates. Weldon codes[3] of rate $1/(t+1)$ are indexed by $\alpha_1, \alpha_2, \ldots, \alpha_t \in \mathbb{F}_{q^k}$ and defined as

$$C_{\text{WE}}^{\alpha_1,\ldots,\alpha_t} = (\varphi(x), \varphi(\alpha_1 x), \ldots, \varphi(\alpha_t x)).$$

For some fixed $t > 2$, can one find explicit α_i's where the distance of $C_{\text{WE}}^{\alpha_1,\ldots,\alpha_t}$ is asymptotically larger than $\sqrt{k}$?

References

1 Wallace C Babcock. Intermodulation interference in radio systems frequency of occurrence and control by channel selection. *The Bell System Technical Journal*, 32(1):63–73, 1953.

2 Roger C Baker, Glyn Harman, and János Pintz. The difference between consecutive primes, ii. *Proceedings of the London Mathematical Society*, 83(3):532–562, 2001.

3 R. C. Bose. An affine analogue of singer's theorem. *The Journal of the Indian Mathematical Society*, 6(0), 1942. URL: `https://www.i-scholar.in/index.php/JIMSIMS/article/view/151305`.

4 Raj Chandra Bose and Sarvadaman Chowla. Theorems in the additive theory of numbers. Technical report, North Carolina State University. Dept. of Statistics, 1960.

5 Pafnutij Lvovič Čebyšev. Mémoire sur les nombres premiers, 1850.

6 Apostolos Dimitromanolakis. Analysis of the golomb ruler and the sidon set problems and determination of large near-optimal golomb rulers. *Diploma thesis, Department of Electronic and Computer Engineering, Technical University of Crete*, 2002.

7 Paul Erdos and Pál Turán. On a problem of sidon in additive number theory, and on some related problems. *J. London Math. Soc*, 16(4):212–215, 1941.

8 Philippe Gaborit and Gilles Zemor. Asymptotic improvement of the gilbert–varshamov bound for linear codes. *IEEE Transactions on Information Theory*, 54(9):3865–3872, 2008.

9 Tao Jiang and Alexander Vardy. Asymptotic improvement of the Gilbert-Varshamov bound on the size of binary codes. *IEEE Trans. Inf. Theory*, 50(8):1655–1664, 2004.

10 Jørn Justesen. Class of constructive asymptotically good algebraic codes. *IEEE Transactions on Information Theory*, 18(5):652–656, 1972.

11 Bernt Lindström. On b2-sequences of vectors. *Journal of number Theory*, 4(3):261–265, 1972.

12 Dick Lipton. An error correcting code from the book. `https://rjlipton.wpcomstaging.com/2011/08/08/an-error-correcting-code-from-the-book/`, August 2011. URL: `https://rjlipton.wpcomstaging.com/2011/08/08/an-error-correcting-code-from-the-book/`.

13 James L Massey. Threshold decoding, 1963.

14 Aditya Potukuchi and Yihan Zhang. Improved efficiency for covering codes matching the sphere-covering bound. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 102–107. IEEE, 2020.

---

[3] The ensemble codes which Weldon [20] proposed can have any rate of $n_1/(n_1 + n_2)$ with n_1, n_2 positive integers. However, he only used codes of rate $1/(t+1)$ to construct the concatenated code.

15   B-Z Shen. A justesen construction of binary concatenated codes that asymptotically meet the zyablov bound for low rate. *IEEE Transactions on Information Theory*, 39(1):239–242, 1993.

16   Amir Shpilka. Capacity achieving two-write wom codes. In *Latin American Symposium on Theoretical Informatics*, pages 631–642. Springer, 2012.

17   Amir Shpilka. New constructions of wom codes using the wozencraft ensemble. *IEEE Transactions on Information Theory*, 59(7):4520–4529, 2013.

18   Simon Sidon. Ein satz über trigonometrische polynome und seine anwendung in der theorie der fourier-reihen. *Mathematische Annalen*, 106(1):536–539, 1932.

19   Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 238–251, 2017.

20   E.J. Weldon. Justesen's construction–the low-rate case (corresp.). *IEEE Transactions on Information Theory*, 19(5):711–713, 1973. `doi:10.1109/TIT.1973.1055068`.