Estimates on the Size of Symbol Weight Codes

Yeow Meng Chee, *Senior Member, IEEE*, Han Mao Kiah, *Student Member, IEEE*, and Punarbasu Purkayastha, *Member, IEEE*

***Abstract*—The study of codes for powerline communications has garnered much interest over the past decade. Various types of codes such as permutation codes, frequency permutation arrays, and constant composition codes have been proposed over the years. In this paper, we study a type of code called bounded symbol weight codes which was first introduced by Versfeld *et al.* in 2005, and a related family of codes that we term constant symbol weight codes. We provide new upper and lower bounds on the size of bounded symbol weight and constant symbol weight codes. We also give direct and recursive constructions of codes for certain parameters.**

***Index Terms*—Asymptotic bounds, constant composition codes (CCCs), powerline communications, Reed–Solomon codes, symbol weight codes.**

I. Introduction

THE notion of transmitting data over powerlines has posed an interesting challenge for information and coding theory. The noise characteristics of such a communication channel include permanent narrowband noise, impulse noise, and white Gaussian noise. Communication over this channel also has an additional requirement that the power envelope be as close to constant as possible. Vinck [31] studied this channel and showed that M-ary frequency shift keying (M-FSK) modulation, in conjunction with the use of permutation codes, provides a constant power envelope, frequency spreading, and redundancy to correct errors resulting from the harsh noise pattern. This has since resulted in research on frequency permutation arrays (FPAs) and constant composition codes (CCCs) which retain the property of a constant power envelope (see [5]–[16], and [14] for a survey). Every codeword of an FPA or a CCC has the requirement that the frequency of each symbol is fixed by the parameters of the code. Versfeld *et al.* [29] introduced the notion of the "same-symbol weight" of a code by relaxing the requirement that every symbol must occur a fixed number of times in any codeword. In every codeword of a same-symbol weight code, the frequency of any symbol is bounded. Even with this relaxation, it is possible to detect permanent narrowband noise. Versfeld *et al.* [29], [30] used Reed–Solomon codes to design codes with specified same-symbol weight.

Manuscript received April 11, 2012; accepted July 30, 2012. Date of publication August 21, 2012; date of current version December 19, 2012. This work was supported in part by the National Research Foundation of Singapore under Research Grant NRF-CRP2-2007-03.

The authors are with the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371 (e-mail: ymchee@ntu.edu.sg; kiah0001@ntu.edu.sg; punarbasu@ntu.edu.sg).

Communicated by N. Kashyap, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2012.2214434

In this paper, we mostly study the asymptotic behavior of codes in the symbol weight space. We use the term bounded symbol weight (as opposed to same-symbol weight [29]) to denote all words in the Hamming space with bounded symbol weight, that is, any symbol in a codeword does not occur more than a fixed number of times, say r. This terminology is adopted in order to distinguish this space from the constant symbol weight space, in which every symbol in a word in the Hamming space occurs at most r times and there exists one symbol which occurs exactly r times. The constant symbol weight space is clearly a subset of the bounded symbol weight space. We also use the term symbol weight space to refer to either the bounded symbol weight or constant symbol weight space. The actual space being referred to is made clear from the context and notation.

As described in [29] and [30], the symbol weight determines whether the code can detect and correct narrowband noise in the powerline channel. An FPA or a CCC belongs to some constant symbol weight space. The constant symbol weight space also contains other compositions all of which have the same maximal part, that is, all such codes have the same fixed symbol weight. Thus, a code in the constant symbol weight space is larger than a CCC of a fixed composition, and is still relevant for correcting narrowband noise. The asymptotic behavior of FPAs have been studied in [4], [8], and [14]. In contrast, there are relatively fewer results on the asymptotic behavior of CCCs (see [23] and [28]). We consider familiar techniques used to derive classical bounds such as the Gilbert-Varshamov (GV) bound, the Johnson bound, and the Singleton bound, on codes in the symbol weight space. However, the derivation of these results is not immediate because of the lack of any reasonable structure in the symbol weight spaces. In particular, even the Hamming balls of a fixed radius in these spaces depend on the center of the ball. In Section V, we also study nonasymptotic bounds on codes in the symbol weight spaces by expressing them in terms of different CCCs. This also raises related combinatorial questions regarding the size and construction of optimal codes, which can be an interesting avenue of future research. In later sections, we show that there exists codes, which are subsets of Reed–Solomon codes, with high rate and relative distance, which are a subset of the constant symbol weight space.

Throughout this paper, we are mostly concerned with codes that have positive rate and positive relative distance. Hence, we do not study codes with very large distances, in the Plotkin region. We start with some basic definitions and notations in Section II. We devote Section III to deriving the exact and asymptotic size of the symbol weight spaces. The results in Section III allow us to determine which constant composition space contained within the symbol weight space contributes the most to the rate of a symbol weight space. These estimates

are used in Section IV to determine upper and lower bounds on bounded symbol weight and constant symbol weight codes. In particular, it is clear that asymptotically some CCCs determine the rate of a symbol weight code. An upper bound is readily obtained from either the Singleton bound or the Linear Programming (LP) bound in the Hamming space. In Section IV, we also provide a Johnson-type bound on codes in the constant symbol weight space and use this bound to derive an asymptotic improvement of the Singleton bound and the LP bound for certain ranges of the minimum distances and the symbol weight. In Section V, we provide nonasymptotic lower bounds on symbol weight codes. We introduce a new metric on the space of compositions of an integer and use this metric to lower bound the size of symbol weight codes by a sum of sizes of CCCs. Finally, in Section VI, we provide other constructions of constant symbol weight codes and, in particular, show that the asymptotic lower bound presented in Section III is tight for certain parameters, for subcodes of Reed–Solomon codes.

II. Preliminaries

Let $\mathbb{Z}_q = \{0, \ldots, q-1\}$ denote an alphabet set of q elements. We consider symbol weight codes in the Hamming space $\mathbb{Z}_q^n = \{0, \ldots, q-1\}^n$. The *symbol weight* of a word is defined as the maximum of the frequencies of occurrences of symbols in the word. For instance, the all-0 word has a symbol weight of n. The bounded symbol weight space with symbol weight r is the set of all words with symbol weight at most r. This space is denoted by $SW(n, q, \leq r)$. The bounded symbol weight space is termed as "same-symbol weight space" in the works of Versfeld *et al.* [29], [30]. We adopt this terminology to distinguish this space from the constant symbol weight space that we define next. In the constant symbol weight space, every word has a symbol weight of exactly r. This space is denoted as $SW(n, q, r)$. If every symbol occurs in each codeword we can use the Pigeonhole principle to get the lower bound $r \geq \lceil n/q \rceil$. Since any word with this lowest value of symbol weight contains the least repetition of any symbol, these words are considered as ones with the *optimal* symbol weight.

In this paper, we study codes in the bounded and constant symbol weight spaces. A bounded (respectively, constant) symbol weight code is a subset of the bounded (respectively, constant) symbol weight space. Let $A_q^{SW}(n, d, \leq r)$ (respectively, $A_q^{SW}(n, d, r)$) denote the maximum size of a bounded (respectively, constant) symbol weight code with distance d in $SW(n, q, \leq r)$ (respectively, $SW(n, q, r)$). We denote a composition of n into q nonnegative parts by $\boldsymbol{n} = [n_0, n_1, \ldots, n_{q-1}]$. The constant composition space with composition $\boldsymbol{n} = [n_0, n_1, \ldots, n_{q-1}]$ is a subset of $\mathbb{Z}_q^n$ in every word of which the ith symbol occurs exactly n_i times. A CCC is a subset of a constant composition space. We use the notation $A_q(\boldsymbol{n}, d)$ to denote the maximum size of a code in the constant composition space given by the composition $\boldsymbol{n}$ and minimum distance at least d. We use the notation $A_q(n, d)$ to denote the maximum size of a code with minimum distance at least d in the Hamming space. A code $\mathcal{C}$ of length n, size M, distance d, over $\mathbb{Z}_q$ is denoted by $\mathcal{C}(n, M, d)_q$. If $\mathcal{C}$ has a constant symbol weight r, it is denoted by $\mathcal{C}(n, M, d, r)_q$. If $\mathcal{C}$ is a linear code of dimension k over a finite field $\mathbb{F}_q$, it is denoted as $\mathcal{C}[n, k, d]_q$.

An FPA consists of vectors in which every symbol occurs a fixed number, say λ, of times. Hence, an FPA is a CCC with composition $\boldsymbol{n} = [\lambda, \ldots, \lambda]$. Thus, the FPA is a subset of the constant symbol weight space with symbol weight λ. Similarly, it can be seen that a CCC with composition $\boldsymbol{n} = [n_0, \ldots, n_{q-1}]$ is a subset of the constant symbol weight space with symbol weight $r = \max\{n_i : i = 0, \ldots, q-1\}$.

The coded modulation scheme introduced by Vinck [31] for the powerline channel considered M-FSK modulation along with the use of permutation codes. The demodulator considered is a hard-decision demodulator consisting of an envelope detector with a threshold. At every time instance, the demodulator provides a multivalued output consisting of all the symbols that correspond to frequencies at which the output of the envelope detector exceeds the threshold. A narrowband noise in this context results in the same symbol appearing at all time instances. As explained in [29], a linear code is less effective in this channel. For instance, the all-zero codeword cannot be distinguished from a narrowband noise. Hence, permutation codes, FPAs and CCCs are more suitable for communication in this channel. To understand why we study the constant symbol weight space, consider the following example.

Example 2.1: Consider a CCC in $\mathbb{Z}_4^8$ with composition $[1, 1, 3, 3]$ and minimum distance $d = 4$, that is suitable for correcting narrowband noise in a powerline channel. Let (0, 1, 2, 2, 2, 3, 3, 3) be a codeword in this CCC. Then, the vector (0, 0, 0, 1, 1, 1, 2, 3) is also a suitable vector for correcting narrowband noise. However, this vector belongs to a different constant composition space that contains vectors with composition $[3, 3, 1, 1]$. Both these constant composition spaces are a subset of the constant symbol weight space with symbol weight three. In Section V, we prove that since the compositions $[1, 1, 3, 3]$ and $[3, 3, 1, 1]$ have distance four in a specific metric that we define later, any vector from the constant composition space with composition $[1, 1, 3, 3]$ will be at a distance at least four from any vector of the other space with composition $[3, 3, 1, 1]$. Hence, we can increase the size of the code by including all the codewords from a CCC in the latter space.

It is clear that any constant symbol weight space with symbol weight r can be written as the union of different constant composition spaces, each of which contains vectors with the same symbol weight r. This relation to the constant composition space is used throughout this paper.

In Section III, we first determine the size of the constant symbol weight space and the bounded symbol weight space. This size is then used to determine a Gilbert-Varshamov-type (GV-type) bound on the symbol weight spaces. Unfortunately, the expression for the size of the symbol weight spaces is unwieldy and gives little insight into the behavior of the lower bounds. The symbol weight spaces are also not ball-homogeneous, that is, the size of a Hamming ball in the space depends on the center of the ball. For example, the bounded symbol weight space in $\mathbb{Z}_3^3$ with symbol weight r at most two has 24 vectors. The ball of radius one around the vector (1, 0, 0) contains six vectors, namely, (1, 0, 0), (2, 0, 0), (1, 1, 0), (1, 2, 0),

(1, 0, 1), and (1, 0, 2). In contrast, the ball of radius one around (2, 1, 0) contains seven vectors, namely, (2, 0, 0), (0, 1, 0), (1, 1, 0), (2, 1, 0), (2, 2, 0), (2, 1, 1), and (2, 1, 2). This fact makes it difficult to state decent lower bounds. Similar comments apply to the computation of the Hamming bound. Hence, in the following two sections, we instead study the asymptotic behavior of the symbol weight spaces and the rate of the corresponding symbol weight codes.

In Section III, we determine the asymptotic size of the symbol weight space. This enables us to determine which constant composition space, contained within the symbol weight space, has the largest size.

III. Asymptotic Size of the Symbol Weight Space

To determine the asymptotic size of the symbol weight space, we first state the expression for the nonasymptotic case. Each vector in the symbol weight space corresponds to a vector in some constant composition space. Hence, we introduce some basic definitions below to describe the size of the symbol weight space. Let us denote the set of all compositions of n into q nonnegative parts by $\mathcal{N}$, i.e.,

$$\mathcal{N} \stackrel{\triangle}{=} \left\{ \boldsymbol{n} \in \mathbb{Z}^q : \boldsymbol{n} \geq 0, \sum_{i=0}^{q-1} n_i = n \right\}$$

and define

$$\mathcal{N}(r) \stackrel{\triangle}{=} \{\boldsymbol{n} \in \mathcal{N} : \max\{n_0, \ldots, n_{q-1}\} = r\}$$
$$\mathcal{N}(\leq r) \stackrel{\triangle}{=} \{\boldsymbol{n} \in \mathcal{N} : \max\{n_0, \ldots, n_{q-1}\} \leq r\}.$$

Let $P(N, K, R)$ denote the compositions of N into K parts, each part bounded between 0 and R. An expression for the size of $P(N, K, R)$ is given by [13, p. 1037]

$$|P(N,K,R)| = \sum_i (-1)^i \binom{K}{i} \binom{K+N-(R+1)i-1}{K-1}. \tag{1}$$

Define k_0 to be $k_0 \stackrel{\triangle}{=} \max\{n-(r-1)q, 1\}$. The quantity k_0 corresponds to the smallest number of symbols that can occur with frequency exactly r in any vector with symbol weight r. The size of the set $\mathcal{N}(r)$ is given by the following lemma.

Lemma 3.1:

$$|\mathcal{N}(r)| = \sum_{k=k_0}^{\lfloor n/r \rfloor} \binom{q}{k} |P(n-rk, q-k, r-1)|$$

and

$$k_0 = \max\{n-(r-1)q, 1\} = \begin{cases} q-\Delta, & r = \lceil \frac{n}{q} \rceil = \frac{n+\Delta}{q} \\ 1, & \text{otherwise} \end{cases}$$

for some $\Delta \equiv \Delta(n,q)$ such that $0 \leq \Delta \leq q-1$.

Proof: If a vector $\boldsymbol{v} \in \mathbb{Z}_q^n$ has a composition $\boldsymbol{n}$ such that exactly k of the symbols in $\mathbb{Z}_q$ have composition r in $\boldsymbol{v}$, then the rest of the symbols must satisfy the inequality

$$n - rk \leq (q-k)(r-1).$$

This inequality, in conjunction with the requirement that at least one symbol must have composition r, determines the value of k_0. If a composition $\boldsymbol{n}$ has exactly k symbols with the value r, then these k symbols can be chosen in $\binom{q}{k}$ ways. The rest of the elements of $\boldsymbol{n}$ must correspond to a composition of $n-rk$ into $q-k$ parts, each part bounded between 0 and $r-1$.

Note that r must satisfy $r \geq \lceil n/q \rceil$. We have

$$n-(r-1)q \geq 1 \quad \Leftrightarrow \quad r \leq \frac{n+q-1}{q}.$$

There is exactly one integer r which satisfies $\lceil n/q \rceil \leq r \leq (n+q-1)/q$. This value of r is given by $r = (n+\Delta)/q$, for some Δ such that $0 \leq \Delta \leq q-1$. ■

The size of the constant symbol weight space $SW(n,q,r)$ can now be determined to be

$$|SW(n,q,r)| = \sum_{k=k_0}^{\lfloor n/r \rfloor} \binom{q}{k} \binom{n}{r,\ldots,r,n-rk} \times \sum_{\boldsymbol{x} \in P(n-rk,q-k,r-1)} \binom{n-rk}{x_1,\ldots,x_{q-k}} \tag{2}$$

where $\boldsymbol{x} = (x_1, \ldots, x_{q-k})$, and r is repeated k times in the multinomial coefficient $\binom{n}{r,\ldots,r,n-rk}$. The size of the bounded symbol weight space is a sum of the sizes of the different constant symbol weight spaces, as shown below

$$\begin{aligned} |SW(n,q,\leq r)| &= \sum_{s=\lceil n/q \rceil}^{r} \sum_{k=k_0(s)}^{\lfloor n/s \rfloor} \binom{q}{k} \binom{n}{s,\ldots,s,n-sk} \times \sum_{\boldsymbol{x} \in P(n-sk,q-k,s-1)} \binom{n-sk}{x_1,\ldots,x_{q-k}} \\ &= \sum_{\boldsymbol{y} \in P(n,q,r)} \binom{n}{y_1,\ldots,y_q} \end{aligned} \tag{3}$$

where $\boldsymbol{x} = (x_1, \ldots, x_{q-k})$, $\boldsymbol{y} = (y_1, \ldots, y_q)$, and $k_0(s) = \max\{n-(s-1)q, 1\}$.

The expressions in the equations above can be used to provide GV-type existence bounds on symbol weight codes. A GV bound on the size of a code $\mathcal{C}$ with minimum distance d in a space $\mathcal{S}$ is given as

$$|\mathcal{C}| \geq \frac{|\mathcal{S}|}{V(\mathcal{S}, d-1)}$$

where $V(\mathcal{S}, d-1)$ is the volume of a ball of radius $d-1$ in the space $\mathcal{S}$. Although the sizes of the constant and bounded symbol weight spaces are given by the above (2) and (3), respectively, there are several hurdles in applying the GV-type bound directly. First, the space itself lacks any suitable structure and is not even ball-homogeneous. Even for the special case of an FPA in which all the symbols occur equally often in every vector, the expression for the GV (and also the Hamming bound) is quite unwieldy because the size of the ball does not have a nice form; see Huczynska [15, Th. 2.7]. Second, the expressions for the sizes of the spaces are not suitable for expressing the bound in a simple form. We instead study the asymptotic form of this bound in the next section. To determine the asymptotic

results, we first need to understand the behavior of the sizes of the symbol weight spaces for large block length n.

The expression for the asymptotic size of the constant symbol weight space is given by the following theorem. A similar expression for the bounded symbol weight space can be readily derived from this theorem and is mentioned at the end of this section. The following theorem holds for any q such that q grows at most proportional to n. Note that all the asymptotics are with respect to n and so the term $o(1)$ below goes to zero as n goes to ∞.

Theorem 3.2: For any q, such that $q = \theta n^{\epsilon}$, where θ is a positive constant and $0 \leq \epsilon \leq 1$

$$\frac{1}{n}\log_q |SW(n,q,r)| = \begin{cases} h_q\left(1-\frac{r}{n}\right) - o(1), & r > \lceil \frac{n}{q} \rceil \\ 1 - o(1), & r = \lceil \frac{n}{q} \rceil. \end{cases}$$

We first give a brief outline of the proof of this theorem. As mentioned earlier, a constant symbol weight space with symbol weight r is a union of several constant composition spaces, each of which contains vectors of symbol weight r. We first show in Lemma 3.3 that the number of constant composition spaces does not contribute to the rate of the constant symbol weight space. This is not surprising and it is true even for the Hamming space, when considered as a union of constant composition spaces. Because of this Lemma, we now know that there is a constant composition space which dominates the expression for the rate. Lemmas 3.4 and 3.5 below help us determine this dominant term. It turns out that this dominant term comes from the constant composition space which has exactly k_0 symbols that occur exactly r times in any vector.

We continue with the proof of the theorem, by first establishing a sequence of lemmas presented below. Let $h_p(x)$ be the p-ary entropy function defined in the range $0 \leq x \leq 1$, as

$$h_p(x) \stackrel{\triangle}{=} -x\log_p \frac{x}{p-1} - (1-x)\log_p(1-x).$$

Lemma 3.3:

$$\frac{1}{n}\log_q |\mathcal{N}(r)| = o(1).$$

Proof: The number of terms in the summation over the range $k_0 \leq k \leq \lfloor n/r \rfloor$ is at most n. Hence, only one of the terms in the summation dominates in the asymptotics. We note that $|P(n-rk, q-k, r-1)| \leq |\mathcal{N}|$. Also, $|\mathcal{N}| = \binom{n+q-1}{q-1}$ (see [27, pp. 415]). For a constant q, it shows that $P(n-rk, q-k, r-1)$ grows at most polynomially in n and hence

$$\begin{aligned} \frac{1}{n}\log_q |\mathcal{N}(r)| &= \frac{1}{n}\log_q |P(n-rk, q-k, r-1)| + o(1) \\ &\leq \frac{1}{n}\log_q (an^{q-1}) + o(1) \\ &= o(1) \end{aligned}$$

for some positive constant a. For $q = \theta n^{\epsilon}$, $0 < \epsilon \leq 1$, and positive constant θ, we get

$$\frac{1}{n}\log_q \binom{q}{k} \leq \frac{q}{n} h_2\left(\frac{k}{q}\right)\log_q 2 + o(1) = o(1)$$

and hence

$$\frac{1}{n}\log_q |P(n-rk, q-k, r-1)| \leq \frac{n+q-1}{n} \times h_2\left(\frac{q-1}{n+q-1}\right)\log_q 2 + o(1) = o(1).$$

■

By the above lemma, we can conclude that in the asymptotics of (2) only one term $\binom{n-rk}{x_1,\ldots,x_{q-k}}$ from the inner summation dominates in the asymptotics, and similarly only one term from $\binom{q}{k}\binom{n}{r,\ldots,r,n-rk}$ is present in the asymptotics. The dominant multinomial terms are given by an optimal choice of k. First, we determine the dominating multinomial term from the inner summation in (2). We use the following lemma. Let $\Gamma(x)$ denote the Gamma function

$$\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt.$$

In particular, for an integer x, $\Gamma(x) = (x-1)!$.

Lemma 3.4 [25, p. 109]: Let $x_1, \ldots, x_K$ be nonnegative real numbers. Then

$$\prod_{i=1}^{K} \Gamma(x_i) \geq \left(\Gamma\left(\frac{\sum_i x_i}{K}\right)\right)^K.$$

This lemma immediately implies that $\binom{N}{x_1,\ldots,x_K} \leq \binom{N}{N/K,\ldots,N/K}$. Hence, the dominating term in the inner summation in (2) is given by $\binom{n-rk}{l,\ldots,l}$, where $l = (n-rk)/(q-k)$.[1] For large n, we obtain the following asymptotic expression for $|SW(n,q,r)|$:

$$\begin{aligned} \frac{1}{n}\log_q |SW(n,q,r)| = \log_q n - k\frac{r}{n}\log_q r - \frac{n-rk}{n} \times \\ \log_q(n-rk) + \frac{n-rk}{n}\log_q(q-k) + \\ \frac{1}{n}\log_q \binom{q}{k} - o(1). \end{aligned} \tag{4}$$

Neglecting the $o(1)$ term, the maximum of the expression in (4) over k yields the rate of the constant symbol weight space. Unfortunately, a closed-form expression for the optimizing value of k seems difficult to achieve, even if k is considered over reals instead of integers. We instead look at how the expression behaves for large n. The lemma below asserts that the maximum is achieved at $k^* = k_0$ as $n \to \infty$.

Lemma 3.5: Let $\{f_n(x)\}_{n=1}^{\infty}$ be a family of bounded, strictly monotonic decreasing functions in x, defined over the domain $x \in [x_0, X_0]$, such that $\lim_{n\to\infty} f_n(x) = f(x)$. Let $\{g_n(x)\}_{n=1}^{\infty}$ be a family of nonnegative functions such

[1] We ignore the fact that the ratios may not be integers. This argument can be made more rigorous, but cumbersome, by taking the composition to be $l_0 = \left\lfloor \frac{n-rk}{q-k} \right\rfloor$ for $(q-k)(1-\{\frac{n-rk}{q-k}\})$ times and $l_1 = \lceil \frac{n-rk}{q-k} \rceil$ for $(q-k)\{\frac{n-rk}{q-k}\}$ times, where $\{x\}$ denotes the fractional part of a real number x.

that $0 \leq g_n(x) \leq C_n$, where C_n depends only on n and $\lim_{n\to\infty} C_n = 0$. Then

$$\max_{x\in[x_0,X_0]} \lim_{n\to\infty} f_n(x) + g_n(x) = f(x_0) = \lim_{n\to\infty} \max_{x\in[x_0,X_0]} f_n(x) + g_n(x).$$

Proof: The strict monotonicity $f_n(x) > f_n(y)$ for any x, y, $x_0 \leq x < y \leq X_0$, implies that $f(x_0) \geq f(x)$ for all $x \in [x_0, X_0]$. Now

$$\max_{x\in[x_0,X_0]} f_n(x) + C_n \geq \max_{x\in[x_0,X_0]} f_n(x) + g_n(x) \geq \max_x f_n(x)$$
$$\Rightarrow f_n(x_0) + C_n \geq \max_{x\in[x_0,X_0]} f_n(x) + g_n(x) \geq f_n(x_0)$$
$$\Rightarrow \lim_{n\to\infty} \max_{x\in[x_0,X_0]} f_n(x) + g_n(x) = f(x_0).$$

We also get

$$\max_{x\in[x_0,X_0]} \lim_{n\to\infty} f_n(x) + g_n(x) = \max_{x\in[x_0,X_0]} f(x) = f(x_0).$$

■

This lemma implies that we can determine the asymptotic optimum of $f_n(x) + g_n(x)$ by simply taking the limit of the sequence of numbers $f_n(x_0)+g_n(x_0)$, which converges to $f(x_0)$.

Proof of Theorem 3.2: We apply Lemma 3.5 as follows. Let

$$F_n(k) = \log_q n - k\frac{r}{n}\log_q r - \frac{n-rk}{n}\log_q(n-rk) + \frac{n-rk}{n}\log_q(q-k),$$
$$G_n(k) = \frac{1}{n}\log_q \binom{q}{k}$$

be defined over integer $k \in [k_0, \lfloor n/r \rfloor]$. $G_n(k)$ can be upper bounded by a term independent of k

$$G_n(k) \leq \frac{1}{n}\log_q \binom{q}{\lfloor q/2 \rfloor}$$

and $\lim_{n\to\infty} \frac{1}{n}\log_q \binom{q}{\lfloor q/2 \rfloor} = 0$. We now note that for every $n \neq rq$, $F_n(k)$ is strictly monotonically decreasing. To establish this, we relax k to reals and consider the derivative $F_n'(k)$. We get

$$nF_n'(k) = -r\log_q r + r\log_q \frac{n-rk}{q-k} + r - \frac{n-rk}{q-k}$$
$$= r\left(\log_q \frac{n-rk}{r(q-k)} - \left(\frac{n-rk}{r(q-k)} - 1\right)\right)$$
$$\leq 0$$

where the last line follows because of the fact that $n - rk \leq r(q-k)$, and that $\log x \leq (x-1)$ for $0 < x \leq 1$, with equality at $x = 1$. We also note that $n - rk < r(q-k)$ if and only if $n \neq rq$. Hence, $F_n(k)$ is strictly monotonic decreasing for $n \neq rq$. For $n = rq$, $F_n(k)$ is a constant independent of k and $k_0 = q$, and hence, the optimal value of $F_n(k)$ is at $k = q$. Since Lemma 3.5 is applicable to $F_n(k)+G_n(k)$, we concentrate only on determining the asymptotics of $F_n(k_0)$. For $r > \lceil n/q \rceil$ we get $k_0 = 1$ and

$$F_n(1) = -\frac{r}{n}\log_q \frac{r}{n} - \left(1-\frac{r}{n}\right)\log_q\left(1-\frac{r}{n}\right) + \left(1-\frac{r}{n}\right)\log_q(q-1) - o(1)$$
$$= h_q\left(1-\frac{r}{n}\right) - o(1).$$

For $r = \lceil n/q \rceil = (n+\Delta)/q$, we have $k_0 = q - \Delta = n - (r-1)q$, and

$$F_n(q-\Delta) = \frac{\Delta(r-1)}{n}\log_q \frac{r}{r-1} - \log_q \frac{n+\Delta}{nq}$$
$$= 1 - o(1).$$

This proves Theorem 3.2. ■

The exponent of the asymptotic size of the bounded symbol weight space $SW(n, q, \leq r)$ is always $n(1 - o(1))$ since it contains $SW(n, q, \lceil n/q \rceil)$.

IV. Asymptotic Size of Symbol Weight Codes

In this section, we provide estimates on the rate of symbol weight codes for all $q = \theta n^{\epsilon}$, for any positive constant θ, and for $0 \leq \epsilon \leq 1$. We considered the asymptotic behavior of the symbol weight spaces because of the difficulty in determining reasonable expressions for fixed n. Below, we determine upper and lower bounds on the rate of a symbol weight code. First, we determine a GV-type bound in Theorem 4.3 below. The Singleton and LP upper bounds on codes in the Hamming space are applicable to the symbol weight codes too. In Theorem 4.4, we show that for constant symbol weight codes, the Singleton and LP upper bounds can be improved substantially for a specific range of the symbol weight.

The following lemma is immediate and it shows that the rate of symbol weight codes can be given in terms of the rate of a CCC.

Lemma 4.1:

$$\frac{1}{n}\log_q A_q^{SW}(n,d,r) = \frac{1}{n}\max_{\boldsymbol{n}\in\mathcal{N}(r)} \log_q A_q(\boldsymbol{n},d) + o(1),$$
$$\frac{1}{n}\log_q A_q^{SW}(n,d,\leq r) = \frac{1}{n}\max_{\boldsymbol{n}\in\mathcal{N}(\leq r)} \log_q A_q(\boldsymbol{n},d) + o(1). \quad (5)$$

Proof: Note that we clearly have the following upper and lower bounds on $A_q^{SW}(n,d,r)$:

$$\max_{\boldsymbol{n}\in\mathcal{N}(r)} A_q(\boldsymbol{n},d) \leq A_q^{SW}(n,d,r) \leq |\mathcal{N}(r)| \max_{\boldsymbol{n}\in\mathcal{N}(r)} A_q(\boldsymbol{n},d). \quad (6)$$

The lemma now follows from an application of Lemma 3.3. The second expression in (5) can be determined similarly. ■

We state the LP upper bound on codes in the Hamming space from Aaltonen [2].

Theorem 4.2 [2]:

$$\limsup_{n\to\infty}\frac{1}{n}\log_q A_q(n,d)\le h_q(k_q(\delta)),\quad 0\le\delta\le\frac{q-1}{q},$$

where $k_q(x)=\frac{q-1}{q}-\frac{q-2}{q}x-\frac{2}{q}\sqrt{(q-1)x(1-x)}, 0\le x\le 1$.

An upper bound on symbol weight codes is readily obtained by an upper bound on codes in the Hamming space, since

$$A_q^{SW}(n,d,r)\le A_q^{SW}(n,d,\le r)\le A_q(n,d).$$

Thus for constant q the LP bound is also an upper bound on symbol weight codes. For q growing with n, the Singleton bound is an upper bound on symbol weight codes. Below, we provide asymptotic estimates of symbol weight codes.

Theorem 4.3: Let $q=\theta n^{\epsilon}$, where $0<\theta$ is a constant, and $0\le\epsilon\le 1$. Let $d/n\to\delta$ and $r/n\to\rho$ as $n\to\infty$, where $0<\delta\le\frac{q-1}{q}$. Then, for q constant, i.e., $\epsilon=0$

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\log_q A_q^{SW}(n,d,r)&\ge h_q(1-\rho)-h_q(\delta)\\ \liminf_{n\to\infty}\frac{1}{n}\log_q A_q^{SW}(n,d,\le r)&\ge 1-h_q(\delta).\end{aligned}\tag{7}$$

For q increasing with n, one can use the Singleton bound. Thus, for $0<\epsilon\le 1$, we get

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\log_q A_q^{SW}(n,d,r)&\ge 1-\rho-\delta,\quad r=\rho n\\ \lim_{n\to\infty}\frac{1}{n}\log_q A_q^{SW}(n,d,r)&=1-\delta,\quad r=o(n)\\ \lim_{n\to\infty}\frac{1}{n}\log_q A_q^{SW}(n,d,\le r)&=1-\delta,\quad \text{any } r.\end{aligned}\tag{8}$$

Remark: Note that for q increasing with n the following limits can be inferred:

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log_q A_q^{SW}(n,d,r)&=\lim_{n\to\infty}\frac{1}{n}\log_q A_q(n,d),\ r=o(n)\\ \lim_{n\to\infty}\frac{1}{n}\log_q A_q^{SW}(n,d,\le r)&=\lim_{n\to\infty}\frac{1}{n}\log_q A_q(n,d),\ \text{any } r.\end{aligned}$$

Proof: We use the following lower bound on the constant symbol weight space, which is actually an Elias-type bound on the Hamming space (see [19]). This is followed by using the GV bound in the Hamming space

$$\begin{aligned}A_q(n,d)&\le\frac{q^n}{|SW(n,q,r)|}A_q^{SW}(n,d,r)\\ \Rightarrow\frac{1}{n}\log_q A_q^{SW}(n,d,r)&\ge\frac{1}{n}\log_q A_q(n,d)+h_q\left(1-\frac{r}{n}\right)\\ &\quad-1-o(1)\\ &\ge h_q\Big(1-\frac{r}{n}\Big)-\frac{1}{n}\log_q V(\mathbb{Z}_q^n,d-1)\\ &\quad-o(1)\\ &=h_q\Big(1-\frac{r}{n}\Big)-h_q\left(\frac{d-1}{n}\right)-o(1)\end{aligned}$$

where $h_q(x)$ is the q-ary entropy function and $V(\mathbb{Z}_q^n,d-1)$ is the volume of the ball of radius $d-1$ in the Hamming space. Similarly, for the bounded symbol weight space, we obtain

$$\frac{1}{n}\log_q A_q^{SW}(n,d,\le r)\ge 1-h_q\left(\frac{d-1}{n}\right)-o(1).$$

For a constant q, the asymptotics of these expressions are as given in (7).

For q growing with n, the upper bound on the symbol weight codes is provided by the Singleton bound

$$A_q^{SW}(n,d,r)\le A_q^{SW}(n,d,\le r)\le A_q(n,d)\le q^{n-d+1}.$$

Using the fact that $h_q(x)=x$ in the limit as $q\to\infty$, we get the results as stated in the theorem. In particular for $r=o(n)$, $\lim_n h_q(1-r/n)=1$ and $\lim_n h_q((d-1)/n)=\delta$. Since $A_q^{SW}(n,d,\le r)$ is greater than $A_q^{SW}(n,d,\lceil n/q\rceil)$, it gives the result stated in (8). ■

The lower bound (7) in the theorem may be interpreted as a GV-type bound in the symbol weight space that can be obtained if the volume of a ball of radius $d-1$ in the symbol weight space is upper bounded by the volume of a ball of radius $d-1$ in the Hamming space. Since the symbol weight space is not ball-homogeneous, that is, the size of the balls of radius $d-1$ depends on the center, we adopt the above method to derive the GV-type lower bound.[2]

In the following theorem, we provide an improvement on the upper bound for a *constant* symbol weight code with symbol weight r.

Theorem 4.4: Let $\lceil n/q\rceil\le r\le 2n/3, q=\theta n^{\epsilon}$ with $0\le\epsilon\le 1$. Let d satisfy $r\le d$. For $n\to\infty$, let $r/n\to\rho$, and $d/n\to\delta$. Then, for constant q

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log_q A_q^{SW}(n,d,r)&\le h_q\left(1-\frac{3}{2}\rho\right)\\ &\quad-(1-\rho)h_q\left(\frac{1-\frac{3}{2}\rho}{1-\rho}\right)+1-\frac{3}{2}\rho.\end{aligned}$$

For q growing with n, we get

$$\limsup_{n\to\infty}\frac{1}{n}\log_q A_q^{SW}(n,d,r)\le 1-\frac{3}{2}\rho.$$

The proof of this theorem relies on a Johnson-type upper bound, and a lemma given below. We follow some elements of the derivation of the Singleton bound in [26]. However, our purpose is to improve the Singleton bound by using the parameters of the constant symbol weight space. The improvement mainly stems from the following lemma.

Lemma 4.5: Let $d\ge r>2n/3$. Then, $A_q^{SW}(n,d,r)=q$.

Proof: We claim that if there are two codewords $\boldsymbol{x},\boldsymbol{y}$ both with symbol weight r, then the symbol which repeats r times must be different in the two codewords. Suppose not. Then, the two codewords $\boldsymbol{x},\boldsymbol{y}$ must have at least $n-2(n-r)$ coordinates which contain the same symbol. Thus, the distance between the codewords is at most $d\le 2(n-r)$ which implies $d<2n/3$,

[2]For certain parameters, better lower bounds on $A_q(n,d)$, for instance from algebraic geometry codes, can improve on this GV bound on $A_q^{SW}(n,d,r)$.

since $r > 2n/3$. This is a contradiction. We get $A_q^{SW}(n,d,r) \leq q$. To show the opposite inequality, let $\boldsymbol{x}$ be a word with symbol weight r, $r > 2n/3$. Then, $\boldsymbol{x} + \alpha\mathbf{1}$, $\alpha \in \mathbb{Z}_q$, where $\mathbf{1}$ is the all-one codeword, are also codewords with symbol weight r. This establishes that $A_q^{SW}(n,d,r) \geq q$. ■

We next give the Johnson-type bound.

Lemma 4.6:

$$A_q^{SW}(n,d,r) \leq \left\lfloor \frac{nq}{n-r} A_q^{SW}(n-1,d,r) \right\rfloor.$$

Proof: Consider the code-matrix of the constant symbol weight code with parameters $(n,M,d,r)_q$. Any row of the code-matrix has at least one symbol of frequency r. Fix one symbol of frequency r in each row. There are a total $M(n-r)$ symbols in the code-matrix which do not contribute to the symbol weight in any codeword. The average number of symbols, averaged over the n columns, with frequency at most r is then $M(n-r)/n$. The average number per symbol, averaged over n columns and q symbols is $M(n-r)/(qn)$. Thus, there exists at least one symbol α and at least one column m such that the subcode consisting of the symbol α in column m has size at least $M(n-r)/(nq)$. Discarding the coordinate corresponding to m gives us the bound as stated in the Lemma. ■

Proof of Theorem 4.4: We now proceed to prove the theorem. Apply Lemma 4.6 recursively $l+1$ times to get

$$A_q^{SW}(n,d,r) \leq \left\lfloor \frac{nq}{n-r} \cdots \left\lfloor \frac{(n-l)q}{n-l-r} \times A_q^{SW}(n-l-1,d,r) \right\rfloor \cdots \right\rfloor.$$

The recursion stops for l such that $r = \lceil 2(n-l-1)/3 \rceil$ and for $d \geq r$. For this value of r and d, $A_q^{SW}(n,d,r) = q$, by Lemma 4.5. The condition $l > 0$ implies $r \leq 2n/3$. Constraints on l are obtained from the inequalities

$$2(n-l-1)/3 \leq r = \lceil 2(n-l-1)/3 \rceil \leq 2(n-l)/3.$$

This gives us the upper bound

$$A_q^{SW}(n,d,r) \leq \frac{n \cdots (n-l)}{(n-r) \cdots (n-l-r)} q^{n-3r/2+1} = \frac{\binom{n}{l+1}}{\binom{n-r}{l+1}} q^{n-3r/2+1}.$$

In the asymptotics as $n \to \infty$, we get $l/n \to 1 - 3/2\rho$. This gives us the upper bounds as stated in the theorem. ■

In the case of q growing with n, this theorem improves on the Singleton bound $1-\delta$ for $\delta < 3\rho/2$. The upper bound in Theorem 4.4 for constant q improves on the LP bound for certain range of parameters. The improvements are possible only for large ρ and for $q \geq 5$. For $q = 2, 3$, the restrictions $\delta \geq \rho$ and $\rho \leq 2/3$ do not leave room for improvement. For constant q, the upper bound is in fact concave in shape. This can be verified by taking the second derivative with respect to ρ, which results in the negative expression $-1/(\rho(1-\rho)\ln q)$. Since this bound does not depend on δ, it seems that further improvements might be possible.

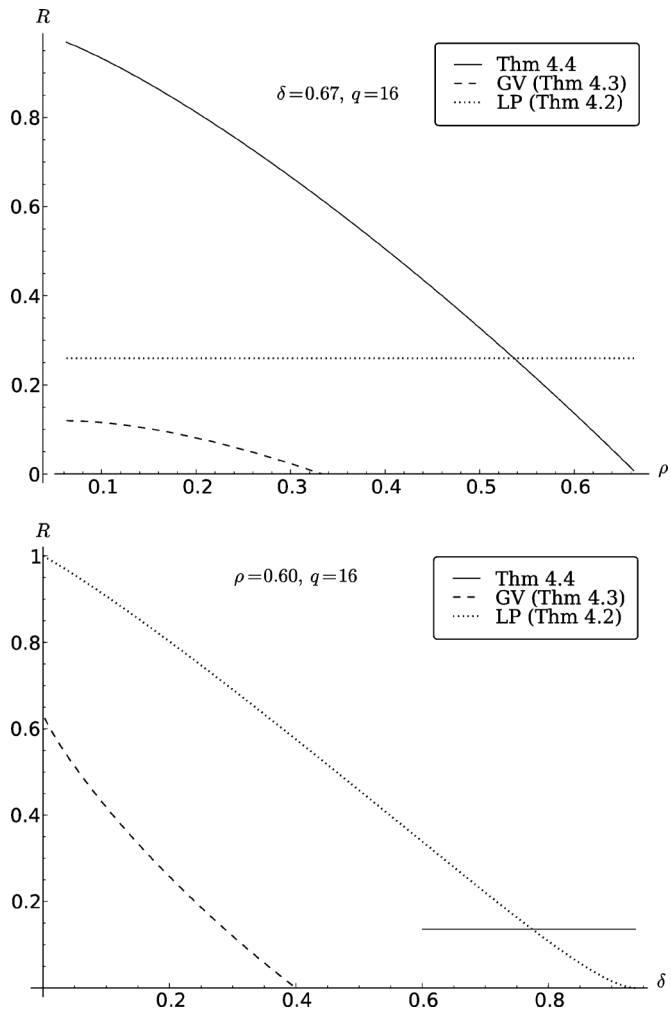

Fig. 1. Plots for $\delta = 2/3$ and for $\rho = 0.6$, respectively, under $q = 16$.

An example plot of all the bounds are provided in Figs. 1 and 2. Since the improvements are for larger q, we show the bounds for $q = 16$. In Fig. 1, the first plot is obtained at a particular value of δ and the second plot is obtained at a particular value of ρ. The improvements (over LP) are obtained in the regions $0.536 \leq \rho \leq 0.67$ and $0.60 \leq \delta \leq 0.774$, respectively. Fig. 2 shows the plots when q is increasing with n. In this case, we compare against the Singleton upper bound. It shows improvements in the region $\rho \leq \delta \leq \frac{3}{2}\rho$. Construction of codes which meet this upper bound for any parameters is an open problem.

Remarks:

1) For $q > n$ and $r = 1$, Dukes [12] provides a Singleton bound, $A_q^{SW}(n,d,1) \leq q(q-1)\cdots(q-n+d)$. Not surprisingly, for $q = \theta n$, $\theta > 1$ this reduces to $1-\delta$ in the asymptotics.
2) Missing from the list of bounds above is a Hamming-type bound on the symbol weight codes. The lack of a simple expression for the size of the ball is the main reason behind this omission.

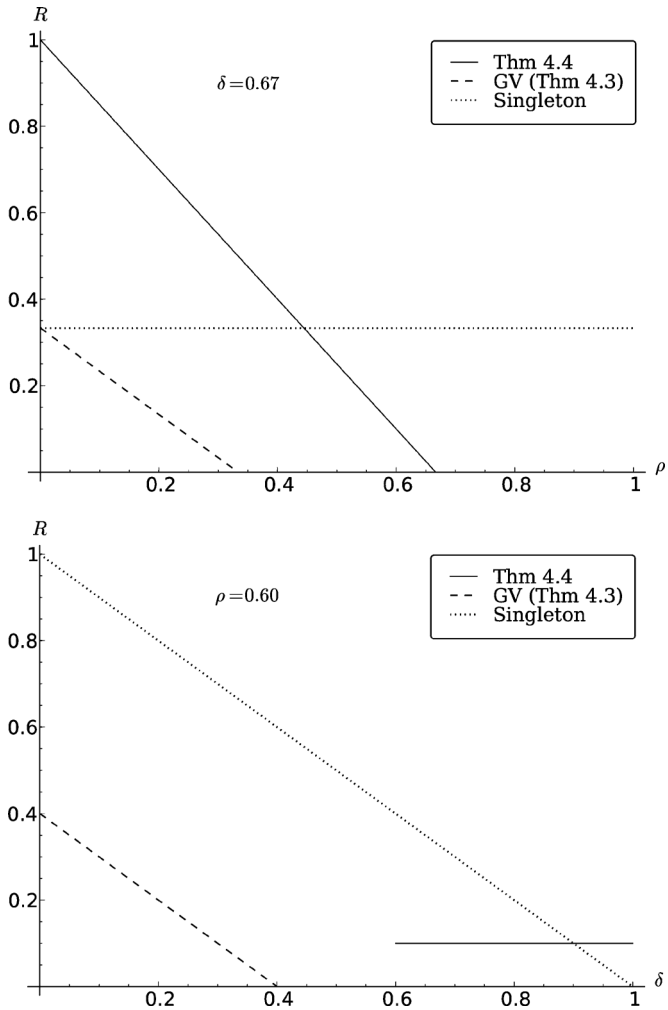

Fig. 2. Plots for $\delta = 2/3$ and for $\rho = 0.6$, respectively.

V. Lower Bound on Symbol Weight Codes

As mentioned in the previous sections, the traditional means of determining the GV-type bounds is not very useful for nonasymptotic block lengths n. In this section, we adopt a different approach to determine lower bounds on the size of symbol weight codes. The lower bounds are obtained by using corresponding constructions and lower bounds on CCCs. There is a wide body of literature on CCCs that can be used to determine these lower bounds (see [5]–[16]). To the best of our knowledge, most of the work in the literature on CCCs has focused on determining constructions and bounds for either very large distances (in the region where the Johnson bound or Plotkin bound is applicable) or for very small distances such as $d = 2, 3, 4$. For FPA and permutation codes, there do exist constructions with distances in the ranges in between (see [4], [8], and [15]). We note three papers in this connection. Sidorenko [28] provides an asymptotic upper bound on CCCs; this is not useful for this section since we look at lower bounds. In a very recent work, Luo and Helleseth [23] construct CCCs of almost uniform composition and with relative distances very close to the Plotkin limit $(d/n \approx (q-1)/q)$. The bound that we provide below requires a relatively large alphabet size so that codes from different constant composition spaces can be combined. Hence, the CCCs from [23] are not useful in this context. Chu *et al.* [8] provide some lower bounds for CCCs for large distances. We use some of the constructions from this latter work in this section to provide examples of lower bounds on symbol weight codes.

We first describe a method to determine the size of a symbol weight code in terms of CCCs. This method may be viewed as a generalization of an elementary bound in [3, Eq. (2)] on binary bounded weight codes to q-ary spaces.

Let $\boldsymbol{n} = [n_0, \ldots, n_{q-1}]$ and $\boldsymbol{n}' = [n'_0, \ldots, n'_{q-1}]$ denote two different compositions of n. The aim here is to lower bound the size of a symbol weight code by the sum of all possible different CCCs which have the same symbol weight. Thus, we first need to determine the condition on two different compositions $\boldsymbol{n}$ and $\boldsymbol{n}'$ such that any vector $\boldsymbol{c}$ with composition $\boldsymbol{n}$ is at least distance d away from a vector $\boldsymbol{c}'$ with composition $\boldsymbol{n}'$. It can be seen that $\min\{n_i, n'_i\}$ is the maximum number of coordinates in $\boldsymbol{c}$ and $\boldsymbol{c}'$ where the ith symbol is common to both. Thus, the Hamming distance $d_H(\boldsymbol{c}, \boldsymbol{c}')$ satisfies

$$d_H(\boldsymbol{c}, \boldsymbol{c}') \geq n - \sum_{i=0}^{q-1} \min\{n_i, n'_i\}.$$

A sufficient condition for $d_H(\boldsymbol{c}, \boldsymbol{c}') \geq d$ to hold is

$$n - \sum_{i=0}^{q-1} \min\{n_i, n'_i\} \geq d. \tag{9}$$

Let

$$d_+(\boldsymbol{n}, \boldsymbol{n}') \stackrel{\triangle}{=} n - \sum_{i=0}^{q-1} \min\{n_i, n'_i\}. \tag{10}$$

Then, we obtain

Lemma 5.1: $d_+(\cdot, \cdot)$ is a distance function on $\mathcal{N}$.

Proof: $d_+(\boldsymbol{n}, \boldsymbol{n}')$ is clearly symmetric. To show the triangle inequality, we note that we can rewrite

$$\begin{aligned} d_+(\boldsymbol{n}, \boldsymbol{n}') &= \sum_{i=0}^{q-1} n_i - \min\{n_i, n'_i\} \\ &= \sum_i (n_i - n'_i)^+ \end{aligned}$$

where $(x)^+ \stackrel{\triangle}{=} \max\{x, 0\}$. Also, for any nonnegative real numbers x, y, z, it can be readily verified that

$$(x-y)^+ + (y-z)^+ \geq (x-z)^+.$$

Thus, we get

$$\begin{aligned} d_+(\boldsymbol{n}, \boldsymbol{n}') + d_+(\boldsymbol{n}', \boldsymbol{n}'') &= \sum_i (n_i - n'_i)^+ + \sum_i (n'_i - n''_i)^+ \\ &= \sum_i (n_i - n'_i)^+ + (n'_i - n''_i)^+ \\ &\geq \sum_i (n_i - n''_i)^+ \\ &= d_+(\boldsymbol{n}, \boldsymbol{n}''). \end{aligned}$$

Since $(n_i - n_i')^+ \geq 0$, we get that $d_+(\boldsymbol{n}, \boldsymbol{n}') = 0$ if and only if $n_i = n_i'$ for all $i = 0, \ldots, q-1$. ■

Let $\mathcal{N}(r,d) \subset \mathcal{N}(r)$ (respectively, $\mathcal{N}(\leq r, d) \subset \mathcal{N}(\leq r)$) be such that for any distinct $\boldsymbol{n}$, $\boldsymbol{n}' \in \mathcal{N}(r,d)$ (respectively, $\mathcal{N}(\leq r, d)$) we have $d_+(\boldsymbol{n}, \boldsymbol{n}') \geq d$. We can now readily give a lower bound on the size of symbol weight codes in terms of the CCCs

$$A_q^{SW}(n,d,r) \geq \sum_{\boldsymbol{n} \in \mathcal{N}(r,d)} A_q(\boldsymbol{n}, d)$$
$$A_q^{SW}(n,d,\leq r) \geq \sum_{\boldsymbol{n} \in \mathcal{N}(\leq r,d)} A_q(\boldsymbol{n}, d). \tag{11}$$

For large n and q, the size of the set $\mathcal{N}(r)$ becomes very large. Hence, finding all the compositions in $\mathcal{N}(r)$ which are separated by distance at least d is difficult. We instead seek lower bounds on $\mathcal{N}(r,d)$ so that the size of the symbol weight codes can be more easily expressed in terms of the sizes of either one or a few CCCs.

Remark: Note that $d_+(\cdot,\cdot)$ is a metric on a "simplex" which intersects each axis at (Euclidean) distance n from the origin. In particular, the components of $\boldsymbol{n}$ need not be restricted to integers for $d_+(\cdot,\cdot)$ to become a metric. Also, n need not be restricted to be an integer.

A. Lower Bounds on $|\mathcal{N}(r,d)|$

In this section, we determine lower bounds to the size of $\mathcal{N}(r,d)$. To get these lower bounds, we first obtain a relation between the Hamming distance between two compositions and the distance between two compositions as given by (10).

Lemma 5.2: For any two compositions $\boldsymbol{n}$, $\boldsymbol{n}' \in \mathcal{N}$, if $d_H(\boldsymbol{n}, \boldsymbol{n}') = 2d$, then $d_+(\boldsymbol{n}, \boldsymbol{n}') \geq d$.

Proof: Define two sets $I^+ = \{i : n_i > n_i'\}$ and $I^- = \{i : n_i < n_i'\}$. Clearly, in the rest of the coordinates, $n_i = n_i'$. Then, we get the following set of equalities:

$$\begin{aligned}
&\sum_{i=0}^{q-1} n_i = \sum_{i=0}^{q-1} n_i' \\
\Leftrightarrow &\sum_{i \in I^+} n_i + \sum_{i \in I^-} n_i = \sum_{i \in I^+} n_i' + \sum_{i \in I^-} n_i' \\
\Leftrightarrow &\sum_{i \in I^+} (n_i - n_i') = \sum_{i \in I^-} (n_i' - n_i) \\
\Leftrightarrow &\sum_{i \in I^+ \cup I^-} (n_i - n_i')^+ = \sum_{i \in I^+ \cup I^-} (n_i' - n_i)^+.
\end{aligned} \tag{12}$$

Note that the LHS and RHS of the last equation are both equal to $d_+(\boldsymbol{n}, \boldsymbol{n}')$. Let $|I^+| = x$, then since $d_H(\boldsymbol{n}, \boldsymbol{n}') = 2d$, we get $|I^-| = 2d - x$. Using the fact that the difference $n_i - n_i' \geq 1$, for $i \in I^+$ and $n_i' - n_i \geq 1$ for $i \in I^-$, we get

$$\begin{aligned}
d_+(\boldsymbol{n}, \boldsymbol{n}') &\geq \max\{x, 2d - x\} \\
&\geq \min_{1 \leq x \leq 2d-1} \max\{x, 2d - x\} \\
&= d.
\end{aligned}$$

■

This lemma immediately allows us to use existing GV bounds in various spaces (under Hamming distance) to derive lower bounds on $|\mathcal{N}(r,d)|$.

1) Lower Bound From a Permutation Code on S_r: Let the alphabet set be $\{0, \ldots, r-1\}$, that is, $q = r$. In every word of length $n = r(r+1)/2$, let all the r symbols occur such that the frequencies of the symbols are in the set $\{1, \ldots, r\}$ and all the frequencies occur. Because of this construction, given r the values of n, q are restricted as given above. Using the GV lower bound on the permutation code with Hamming distance at least $2d$ between two codewords gives us the lower bound

$$|\mathcal{N}(r,d)| \geq \frac{r!}{V(2d-1, S_r)} \tag{13}$$

where $V(2d-1, S_r)$ is the volume of the ball of radius $2d-1$ in S_r.

2) Lower Bound for General q: As explained in the proof of Theorem 4.3, a lower bound on constant symbol weight codes is provided by an Elias-type bound in the Hamming space. We can obtain another lower bound on constant symbol weight codes by considering lower bounds on $\mathcal{N}(r,d)$. A lower bound on $\mathcal{N}(r,d)$ is obtained by letting k of the symbols $\{0, \ldots, q-1\}$ repeat r times in every codeword of the constant symbol weight code and the remaining $q-k$ symbols satisfy the condition that there are $(q-k)\{\frac{n-rk}{q-k}\}$ symbols with composition $l_1 = \lceil \frac{n-rk}{q-k} \rceil$ and $(q-k)(1-\{\frac{n-rk}{q-k}\})$ symbols with composition $l_0 = \lfloor \frac{n-rk}{q-k} \rfloor$. Denote this composition by $\boldsymbol{n} = \boldsymbol{n}(l_0, l_1, k, r)$, that is

$$\boldsymbol{n}(l_0, l_1, k, r) = [\underbrace{r, \ldots, r}_{k}, \underbrace{l_0, \ldots, l_0}_{(q-k)\{\frac{n-rk}{q-k}\}}, \underbrace{l_1, \ldots, l_1}_{(q-k)(1-\{\frac{n-rk}{q-k}\})}].$$

Because of the above choice of the repetitions of each symbol we seek a "binary constant weight code" in $\mathcal{N}(r,d)$ where $q-k$ coordinates have the value l_0 or l_1 and the rest k coordinates have the value r. We also want the Hamming distance between distinct codewords to be at least $2d$. Denote the maximum size of a binary constant weight code of length n, weight w and minimum distance d by $A_2(n,d,w)$.

The GV bound under the above constraints is

$$|\mathcal{N}(r,d)| \geq A_2(q, 2d, k) \geq \frac{\binom{q}{k}}{\sum_{i=0}^{2d-2} \binom{k}{i}\binom{q-k}{i}}.$$

Note that we get $2d-2$ in the denominator (instead of $2d-1$) since the binary constant weight space affords only even distances. From [18], we know that the lower bound is significant and grows exponentially as 2^{qT}, for some constant $T \equiv T(q,k,d)$ only when the following conditions are satisfied:

$$\frac{q}{2}\left(1 - \sqrt{1 - \frac{4d}{q}}\right) \leq k \leq \frac{q}{2}\left(1 + \sqrt{1 - \frac{4d}{q}}\right)$$
$$d \leq k\left(1 - \frac{k}{q}\right).$$

The above lower bounds on $|\mathcal{N}(r,d)|$ give lower bounds on symbol weight codes as follows.

B. Lower Bounds on Codes

The lower bound on $A_q^{SW}(n,d,r)$ can be stated as follows.

Theorem 5.3: We get the following results for different compositions.

1) For $\boldsymbol{n} = [1,\ldots,r]$, we get

$$A_q^{SW}(n,d,r) \geq \frac{r!}{V(2d-1,S_r)} A_q(\boldsymbol{n},d).$$

2) For $\boldsymbol{n} = \boldsymbol{n}(l_0,l_1,k,r)$, we get

$$\begin{aligned} A_q^{SW}(n,d,r) &\geq A_2(q,2d,k) A_q(\boldsymbol{n},d) \\ &\geq \frac{\binom{q}{k}}{\sum_{i=0}^{2d-2}\binom{k}{i}\binom{q-k}{i}} A_q(\boldsymbol{n},d). \end{aligned}$$

3) Let $k_1 \geq k_0 = \max\{n-(r-1)q, 1\}$, and $b \equiv b(k_1) = \left\lfloor \frac{\lfloor n/r \rfloor - k_1}{2d} \right\rfloor$. Then

$$\begin{aligned} A_q^{SW}(n,d,r) \geq \max_{k_0 \leq k_1 \leq \lfloor n/r \rfloor} \sum_{i=0}^{b} A_2(q,2d,k_1+2di) \times \\ A_q(\boldsymbol{n}(l_0,l_1,k_1+2di,r),d). \end{aligned} \tag{14}$$

Proof: The first two results follow immediately from the lower bounds on $|\mathcal{N}(r,d)|$. In part 1, each codeword in the permutation code corresponds to a rearrangement of the composition $\boldsymbol{n} = [1,\ldots,r]$. In part 2, each codeword in the binary constant weight code corresponds to a rearrangement of the composition in $\boldsymbol{n} = \boldsymbol{n}(l_0,l_1,k,r)$.

For the third result, we include a larger range of CCCs. The expression is obtained by taking CCCs from separate constant composition spaces $\boldsymbol{n}(l_0,l_1,k_1+2di,r)$, whose compositions are separated by a Hamming distance of at least $2d$. Two different compositions $\boldsymbol{n}(l_0,l_1,k_1+2di,r)$ and $\boldsymbol{n}(l_0,l_1,k_1+2d(i+1),r)$ correspond to taking binary constant weight codes with weights separated by $2d$. Note that this choice of separate compositions corresponds to a binary bounded weight code, as studied in [3]. ■

There is a tradeoff between the size of the constant composition space with composition $\boldsymbol{n}(l_0,l_1,k,r)$ and the size of the constant weight code in $\mathbb{Z}_2^q$. The size of the constant weight code in $\mathbb{Z}_2^q$ is substantial only for large k around $q/2$. On the other hand, the size of the constant composition space is large for small k, thus potentially allowing for a larger CCC.

The lower bound in (14) is in fact useful in the case of a bounded symbol weight code with symbol weight at most r. It is unclear how to combine codes of different symbol weights s, where $\lceil n/q \rceil \leq s \leq r$, such that we can obtain a computable expression. We instead use (14) and optimize over the different symbol weights s and the smallest weight k_1. Note that for a given symbol weight s, the quantity k_1 corresponds to the minimum number of symbols with frequency s that we include in our estimate.

Theorem 5.4:

$$\begin{aligned} A_q^{SW}(n,d,\leq r) &\geq \max_{\lceil \frac{n}{q} \rceil \leq s \leq r} A_q^{SW}(n,d,s) \\ &\geq \max_{\lceil \frac{n}{q} \rceil \leq s \leq r} \max_{k_1 \leq \lfloor \frac{n}{s} \rfloor} \sum_{i=0}^{b(s)} A_2(q,2d,k(i,s)) \\ &\quad \times A_q\big(\boldsymbol{n}(l_0(i,s),l_1(i,s),k(i,s),s),d\big) \end{aligned}$$

where $k_1 \equiv k_1(s) \geq \max\{n-(s-1)q,1\}$, $k \equiv k(i,s) = k_1(s)+2di$, $b(s) = \left\lfloor \frac{\lfloor n/s \rfloor - k_1}{2d} \right\rfloor$, $l_0(i,s) = \lfloor \frac{n-sk}{q-s} \rfloor$, and $l_1(i,s) = \lceil \frac{n-sk}{q-s} \rceil$.

C. Numerical Examples

We consider some numerical examples in order to show how the expressions in the previous section can be used to obtain lower bounds on symbol weight codes. We adopt the exponential notation of Chu *et al.* [8] to denote a composition in a compact form. The notation $n_0^{t_0} n_1^{t_1} \cdots n_{q-1}^{t_{q-1}}$ is used to denote the composition

$$[\underbrace{n_0,\ldots,n_0}_{t_0},\underbrace{n_1,\ldots,n_1}_{t_1},\ldots,\underbrace{n_{q-1},\ldots,n_{q-1}}_{t_{q-1}}].$$

We also recall the notion of a refinement of a composition from the same work. The composition $\boldsymbol{n} = [n_0,\ldots,n_{q-1}]$ is called a *refinement* of a composition $\boldsymbol{m} = [m_0,\ldots,m_{p-1}]$ if there is a partition $I_0,\ldots,I_{p-1}$ of $\{0,\ldots,q-1\}$ such that $\sum_{i\in I_j} n_i = m_j$, for every j. We write $\boldsymbol{n} \preccurlyeq \boldsymbol{m}$ if $\boldsymbol{n}$ is a refinement of $\boldsymbol{m}$. This notion is important because of the following inequality (see [8]):

$$A_q([n_0,\ldots,n_{q-1}],d) \geq A_p([m_0,\ldots,m_{p-1}],d). \tag{15}$$

Below, we use lower bounds on FPAs, where the lower bound is taken from [8]. Using (15), lower bounds on CCCs are obtained from the lower bounds on the FPAs. The lower bound on FPA mentioned below rely on the existence of certain (generalized) distance preserving mappings from $\mathbb{Z}_q^n$ to the permutation space S_n (see [8]). The distances between compositions used in this section are all taken in the $d_+(\cdot,\cdot)$ metric, unless mentioned otherwise.

Example 5.5: In this example, we show how Theorem 5.3 and (11) can be used. We know from [8, Example 3.7] that $A_4(6^4,7) \geq 2^{12}$. Since $1^4 5^4 \preccurlyeq 6^4$, we immediately obtain that $A_8(1^4 5^4,7) \geq 2^{12}$. In this case, $q=8$ and $d=7$ and the number of symbols occurring with frequency five is $k=4$. Hence, Theorem 5.3 is not applicable. But (11) can be applied directly. For instance, the compositions $1^4 5^4$ and $5^4 1^4$ satisfy $d_+(1^4 5^4, 5^4 1^4) = 4(1-5)^+ + 4(5-1)^+ = 16$ which is greater than seven. Thus, $A_8^{SW}(24,7,5) \geq 2 \cdot 2^{12}$.

In fact, the compositions $\boldsymbol{n} = 1^4 5^4$ and $\boldsymbol{n}' = 5^4 1^4$ have the special property that if a symbol i has different compositions n_i, n_i' then $|n_i - n_i'| = 4$. We can exploit this property to get a variant of Lemma 5.2 below.

Lemma 5.6: For two compositions $\boldsymbol{n}$, $\boldsymbol{n}' \in \mathcal{N}$ let $d_H(\boldsymbol{n}, \boldsymbol{n}') = D$. For $i = 0, \ldots, q-1$, if either $|n_i - n_i'| \geq a > 0$ or it is zero, then $d_+(\boldsymbol{n}, \boldsymbol{n}') \geq Da/2$.

Proof: The proof is very similar to the proof of Lemma 5.2. Let I^+, I^- be as defined in the proof of that lemma, and let $|I^+| = x$, $|I^-| = D - x$. Finally, use (12) to get

$$\begin{aligned} d_+(\boldsymbol{n}, \boldsymbol{n}') &\geq a \max\{x, D-x\} \\ &\geq a \min_{1 \leq x \leq D-1} \max\{x, D-x\} \geq Da/2. \end{aligned}$$

■

Returning to this example, we know that if two compositions differ in a symbol, then the difference is four, and so $a = 4$. We will have the distance between two compositions at least seven if we ensure (using Lemma 5.6) that $Da/2 \geq 7$, that is, $D \geq \lceil 14/4 \rceil = 4$. Using $q = 8$, $D = 4$, $k = 4$, we get that $A_8^{SW}(24, 7, 5) \geq A_2(8, 4, 4) A_8(1^4 5^4, 7)$. The size of the binary code is obtained from [1]: $A_2(8, 4, 4) = 14$. This gives the much improved lower bound $A_8^{SW}(24, 7, 5) \geq 14 \cdot 2^{12}$.

Example 5.7: We continue with the previous example and a finer refinement $1^8 2^8 \preccurlyeq 6^4$. We show in this example that directly using (11) can lead to a better bound, compared to Theorem 5.3. In this case, we apply Theorem 5.3 with $k = 8$, $2d = 14$, $q = 16$, and $\boldsymbol{n}(l_0, l_1, k, r) = 1^8 2^8$. From the table of constant weight codes in [1], we get $A_2(16, 14, 8) = 2$. This gives $A_{16}^{SW}(24, 7, 2) \geq 2 \cdot 2^{12}$. However, this bound can be improved by using (11). A greedy search through the compositions with the maximum value of each part being two shows that the compositions $0^3 1^2 2^{11}$, $1^1 2^4 0^3 1^1 2^7$, $1^1 2^8 0^3 1^1 2^3$, $2^1 1^1 2^3 1^1 2^3 1^1 2^3 0^2 1^1$, and $2^3 0^1 2^4 0^1 2^3 0^1 2^2 0^1$ are mutually at distance at least seven from one another. Since each of them is a refinement of 6^4, we get that $A_{16}^{SW}(24, 7, 2) \geq 5 \cdot 2^{12}$.

Finally, we look at an example which considers the bounded symbol weight and demonstrate the use of Theorem 5.4.

Example 5.8: Consider the refinement $3^8 \preccurlyeq 6^4$. We can consider a code in this space to be embedded in the constant composition space with 16 symbols. Thus, we get $A_{16}(3^8 0^8, 7) \geq 2^{12}$. We use the fact that if we consider all compositions containing eight symbols occurring with frequency three, then the difference of frequency between two symbols from different compositions is either zero or three. Thus, $a = 3$, $k = 8$, and $q = 16$. Using Lemma 5.6, we need to ensure that $Da/2 \geq 7 \Rightarrow D \geq 5$. We get the lower bound $A_{16}(24, 7, 3) \geq A_2(16, 5, 8) A_{16}(3^8 0^8, 7)$. From [1], we have $A_2(16, 5, 8) = A_2(16, 6, 8) \geq 120$. Hence, $A_{16}^{SW}(24, 7, \leq 3) \geq \max\{A_{16}^{SW}(24, 7, 3), A_{16}^{SW}(24, 7, 2)\} \geq 120 \cdot 2^{12}$. In this particular example, using the lower bound in (14) on the constant symbol weight codes with $k_1 = 1$ does not yield a better bound, primarily due to the absence of a known good lower bound on the corresponding CCC. The improvement mainly stems from the fact that we use a large binary constant weight code with weight $k_1 = q/2$.

As is evident from the above examples, Theorems 5.3 and 5.4 help in actual computation of the bounds. In Example 5.7, the number of ordered partitions of $n = 24$, into $q = 16$ parts with each part taking values between zero and two, inclusive, is $|P(24, 16, 2)| = 258570$. A nonexhaustive greedy search could only find five compositions. It is computationally difficult to search exhaustively in such a large space. Similarly, for larger lengths and alphabet sizes, finding the size of, and compositions in, $\mathcal{N}(r, d)$ is difficult. Instead, by relying on known bounds on binary constant weight codes and CCCs, we compute the sizes of the symbol weight codes more easily.

VI. Constructions of Symbol Weight Codes

In this section, we determine constructions of symbol weight codes. Versfeld *et al.* [29], [30] provided constructions of bounded symbol weight codes from Reed–Solomon codes. We seek to obtain constant symbol weight codes with positive rate and positive relative distance. The following two constructions provide us with such codes with positive rate and positive relative distance, given that we already have a constant symbol weight code with positive rate and positive relative distance.

$\boldsymbol{u}|\boldsymbol{v}$ *Construction:* Let $\mathcal{C}$ be a constant symbol weight code with parameters $\mathcal{C}(n, M, d, r)_q$ over $\mathbb{Z}_q$. Let $\mathcal{C}'$ be an FPA over $\mathbb{Z}_q$ with parameters $\mathcal{C}'(r'q, M', d')$. Then the $\boldsymbol{u}|\boldsymbol{v}$ construction results in a code $\mathcal{D}$. It is obtained by taking all codewords as follows:

$$\mathcal{D} = \{(\boldsymbol{u}, \boldsymbol{v}) : \boldsymbol{u} \in \mathcal{C}, \boldsymbol{v} \in \mathcal{C}'\}.$$

The code $\mathcal{D}$ has parameters $\mathcal{D}(n + r'q, MM', \min\{d, d'\}, r + r')_q$ over $\mathbb{Z}_q$. In particular if the code $\mathcal{C}$ had the minimum symbol weight $r = \lceil n/q \rceil$ then so does the code $\mathcal{D}$, that is, $r + r' = \lceil (n + r'q)/q \rceil$.

Concatenated Construction: Let $\mathcal{C}$ be a code over $\mathbb{Z}_q$ with parameters $\mathcal{C}(n, M, d)_q$. Let $\mathcal{C}'$ be an FPA with parameters $\mathcal{C}'(rp, M', d')_p$ over $\mathbb{Z}_p$, such that $M' \geq q$. The concatenated code $\mathcal{D}$ with $\mathcal{C}'$ as the inner code and $\mathcal{C}$ as the outer code has parameters $\mathcal{D}(nrp, M, dd', rn)_p$. It is obtained by replacing every q-ary symbol of $\mathcal{C}$ with a codeword from $\mathcal{C}'$. In particular, the resulting code $\mathcal{D}$ has the minimum symbol weight $npr/p = nr$.

A. Constructions From Reed–Solomon Codes

In this section, we use the symbol k to denote the dimension of the Reed–Solomon code. Let $\mathcal{C}[n, k, d]_q$ be a Reed–Solomon code over a finite field $\mathbb{F}_q$ with $d = n - k + 1$ and $n = q - 1$. Versfeld *et al.* [29], [30] showed that aside from the Reed–Solomon codewords which correspond to a constant polynomial, the Reed–Solomon code has maximum symbol weight of $n - d = k - 1$. It is also established in the same works that there exists a coset of the Reed–Solomon code such that the maximum symbol weight of any codeword in the code is at most $n - d + 1$. These codes belong to the bounded symbol weight space $SW(n, q, \leq r)$. By the Singleton bound, these are optimal codes.

In this section, we establish several results which show that subsets of Reed–Solomon codes or their cosets can achieve the GV-type lower bound in Theorem 4.3. First, Lemma 6.1 below shows that for any *constant* symbol weight r, there exists a coset of the Reed–Solomon code that attains the GV bound asymptotically. In Theorem 6.4, we provide a more explicit description

of a subset of the Reed–Solomon code itself that has the constant symbol weight r for $r \geq n/2$, such that it attains the GV bound asymptotically. Since $k-1 \geq r$, this also means that this latter result holds only for Reed–Solomon codes with rate more than $1/2$.

We first show by an averaging argument that the GV bound can be achieved by subcodes of cosets of Reed–Solomon codes.

Lemma 6.1: Let $\mathcal{C}[n,k,d]_q$ be a family of Reed–Solomon codes. For $n \to \infty$ let $r/n \to \rho$, and $d/n \to \delta$. Then, there exists a family of subcodes $\mathcal{C}'$ which is a subset of some coset of the code $\mathcal{C}$ such that

$$\liminf_{n\to\infty} \frac{1}{n} \log_q |\mathcal{C}'| \geq 1 - \rho - \delta.$$

Proof: Let $\mathcal{C}_1, \ldots, \mathcal{C}_{q^{n-k}}$ denote the cosets of the Reed–Solomon code. Since the cosets of the Reed–Solomon code are disjoint and they partition the Hamming space, we have

$$\sum_{i=1}^{q^{n-k}} |\mathcal{C}_i \cap SW(n,q,r)| = |SW(n,q,r)|.$$

Thus, the average size of the intersection of a coset with the space $SW(n,q,r)$ is $|SW(n,q,r)|/q^{n-k}$. Hence, there exists at least one coset whose intersection with $SW(n,q,r)$ has size at least this average. In the asymptotics for n, $q \to \infty$, we get the result stated in the Lemma. ■

In the remaining part of this section, we give a more explicit description of a subcode of the Reed–Solomon code with rate equal to the GV-type bound. The derivation of this result uses a lemma and a proposition stated below. The Proposition 6.2 below states that asymptotically the rate of the constant symbol weight code with symbol weight r that is a subset of the Reed–Solomon code, can not exceed the rate of the subcode formed by all the codewords of weight $n-r$. Lemma 6.3 gives an upper bound on the size of the number of codewords of weight $n-r$. The combination of this proposition and the lemma imply that the rate of the constant symbol weight code, which is a subset of the Reed–Solomon code, with symbol weight $r = \rho n$, can not exceed the GV-type lower bound $1-\rho-\delta$ that we obtained in Theorem 4.3, *for any* ρ, $0 < \rho < 1$. Theorem 6.4 below shows that this rate can be attained for any ρ satisfying $1/2 \leq \rho < 1$. We state the proposition and the lemma first, and defer their proofs to after the proof of the theorem.

Proposition 6.2: Let $\mathcal{C}[n,k,d]_q$ be a family of Reed–Solomon codes with parameters $n = q-1$, $d = n-k+1$. Let $S(r)$ denote the set of vectors with symbol weight exactly r, for $1 \leq r \leq k-1$, and let B_{n-r} denote the number of vectors of weight $n-r$. Then

$$|S(r)| \leq q(q-1)B_{n-r}.$$

Lemma 6.3: The weight distribution $\{B_w : w = d, \ldots, n\}$ of a linear maximum distance separable code with parameters $[n,k,d]_q$ satisfies

$$B_{n-r} \leq \binom{n}{n-r}(q^{k-r}-1),$$

for $0 \leq r \leq k-1$.

The main theorem in this section is now stated below.

Theorem 6.4: Let $\mathcal{C}[n,k,d]_q$ denote the family of Reed–Solomon codes with $n = q-1$ and $d = n-k+1$. Let $k-1 \geq r \geq n/2$. For $n \to \infty$, let $r/n \to \rho$ and $d/n \to \delta$. There exists a family of subcodes $\mathcal{C}'$ of $\mathcal{C}$ of symbol weight exactly r such that

$$\lim_{n\to\infty} \frac{1}{n} \log_q |\mathcal{C}'| = 1 - \rho - \delta.$$

Proof: Every codeword of the Reed–Solomon code consists of coordinates which are the evaluations at all the nonzero points of $\mathbb{F}_q$, of a polynomial of degree at most $k-1$. Let $f(x) = f_0 + f_1 x + \cdots + f_{k-1}x^{k-1}$ be a polynomial in $\mathbb{F}_q$. Let $\mathbb{F}_q^* = \mathbb{F}_q \setminus \{0\}$. If $f(x)$ has symbol weight r then it implies that $f(x) = \alpha$ for some $\alpha \in \mathbb{F}_q$ and for r different values of x in $\mathbb{F}_q^*$. In other words, $f(x) - \alpha$ has exactly r distinct roots. Note that r is restricted to be $r \leq k-1$ since the polynomials cannot have more than $k-1$ roots.

Let $f(x)$ be a polynomial of degree $k-1$ such that it has exactly r nonzero distinct roots $\alpha_1, \ldots, \alpha_r$ in $\mathbb{F}_q^*$. Then $f(x)$ can be written as

$$f(x) = \beta(x-\alpha_1) \times \cdots \times (x-\alpha_r) \times g(x),$$

where $\beta \in \mathbb{F}_q^*$ and $g(x)$ is a product of monic irreducible polynomials, each of degree at least two. The total degree of $g(x)$ is $k-1-r$. Since $r \geq n/2$, the polynomial $f(x)$ can not attain the value α, where $\alpha \in \mathbb{F}_q^*$, at more than r different points $x \in \mathbb{F}_q^*$ since there are $q-1-r \leq n/2$ points at which the function is nonzero. Hence, the symbol weight of the codeword represented by $f(x)$ is exactly r. We seek the asymptotic exponent of the number of such polynomials $f(x)$. This number is dominated by the number of possible monic irreducible polynomials $g(x)$. To describe this number, we recall the definition of the Möbius function $\mu(t)$

$$\mu(t) \stackrel{\triangle}{=} \begin{cases} 1, & \text{if } t = 1 \\ (-1)^s, & \text{if } t \text{ has } s \text{ distinct prime factors} \\ 0, & \text{if } p^2 | t \text{ for some prime } p. \end{cases}$$

The number of monic irreducible polynomials of degree t is given by the sum $\frac{1}{t}\sum_{s|t, s\geq 1} \mu(s) q^{t/s}$ (see [20, Th. 3.25]). In particular, for large t, this sum is dominated by just the first term $\frac{1}{t}\mu(1)q^t = \frac{1}{t}q^t$. As a consequence, asymptotically the number of polynomials $f(x)$ is described by the number of monic irreducible polynomials $g(x)$ of degree $k-1-r$. The asymptotic exponent of this count is approximately $\lim_{n\to\infty}(k-1-r)/n = 1-\delta-\rho$. Thus, the subset $\mathcal{C}'$ of $\mathcal{C}$ consists of all the

codewords obtained from at least these polynomials. Hence, the count above provides a lower bound on the rate of $\mathcal{C}'$

$$\liminf_{n\to\infty} \frac{1}{n}\log_q |\mathcal{C}'| \geq 1-\delta-\rho.$$

The upper bound on the rate of $\mathcal{C}'$ is obtained by applying both Lemma 6.3 and Proposition 6.2. We get

$$\begin{aligned} |\mathcal{C}'| \leq |S(r)| &\leq q(q-1)B_{n-r} \\ &< q(q-1)\binom{n}{n-r}q^{k-r} \end{aligned}$$

and

$$\limsup_{n\to\infty} \frac{1}{n}\log_q |S(r)| \leq \lim_{n\to\infty} \frac{k-r}{n} = 1-\delta-\rho.$$

■

Proof of Proposition 6.2: Let $\boldsymbol{c} = (c_1, \ldots, c_n)$ be a codeword in the Reed–Solomon code. Then, $\boldsymbol{c}$ is the image of a polynomial $c(x)$ evaluated at all points of $\mathbb{F}_q^* = \mathbb{F}_q \setminus \{0\}$. We write $\boldsymbol{c} = (c(x))_{x\in\mathbb{F}_q^*}$. If $\boldsymbol{c}$ resulting from $c(x)$ has symbol weight exactly r then so do the codewords obtained from the polynomials $\gamma c(x) + \beta$, for $\gamma \in \mathbb{F}_q^*$, $\beta \in \mathbb{F}_q$. Consider the subset $S'(r)$ of $S(r)$ that is obtained by retaining exactly one monic polynomial from the set $\{\gamma c(x) + \beta : \gamma \in \mathbb{F}_q^*, \beta \in \mathbb{F}_q\}$ for any polynomial $c(x)$. Thus, the size of $S'(r)$ satisfies $|S'(r)| = |S(r)|/(q(q-1))$.

We claim that $|S'(r)| \leq B_{n-r}$. To show this, we claim that there exists an injection mapping from $S'(r)$ to the set of all vectors of weight $n-r$. Since any $c(x)$ in $S'(r)$ has symbol weight exactly r, there exists a $\beta \in \mathbb{F}_q$ such that $c(x) - \beta$ has exactly r distinct roots. Thus, the codeword $(c(x)-\beta)_{x\in\mathbb{F}_q^*}$ has Hamming weight exactly $n-r$. This is the only such vector. If there exists $e(x) \in S'(r)$ and $\alpha \in \mathbb{F}_q$ such that $(e(x) - \alpha)_{x\in\mathbb{F}_q^*} = (c(x)-\beta)_{x\in\mathbb{F}_q^*}$, then the two polynomials $c(x)$ and $e(x)$ must satisfy the relation $c(x) - \beta = e(x) - \alpha$ since they are the same on $n = q-1$ points and their degrees are at most $k-1 < n$. Thus, $c(x) = e(x) - \alpha + \beta$, which is not possible since $S'(r)$ contains exactly one polynomial of this form. ■

Proof of Lemma 6.3: The expression for the weight distribution satisfies (see [24, Chapter 11])

$$B_{n-r} = \binom{n}{n-r}\sum_{j=0}^{k-r-1}(-1)^j\binom{n-r}{j}(q^{k-r-j}-1).$$

Retaining only the first term gives the required upper bound on B_{n-r}. The above expression can be rewritten as

$$\begin{aligned} B_{n-r} = \binom{n}{n-r}\Bigg\{(q^{k-r}-1) - \sum_{i=1}^{\lfloor (k-r-1)/2\rfloor}\Bigg[\binom{n-r}{2i-1}\times \\ (q^{k-r-(2i-1)}-1) - \binom{n-r}{2i}(q^{k-r-2i}-1)\Bigg] - \\ I(2\nmid k-r-1)\binom{n-r}{k-r-1}(q-1)\Bigg\} \end{aligned}$$

where $I(2 \nmid k-r-1)$ is an indicator function that is 1 if 2 does not divide $k-r-1$ and 0 otherwise. We show that each of the terms in the summation above is positive, and hence, we can upper bound B_{n-r} by only the first term. This is proved via the following sequence of inequalities. For any j, $j = 0, \ldots, k-r-1$, we have

$$\begin{aligned} &\binom{n-r}{j}(q^{k-r-j}-1) > \binom{n-r}{j+1}(q^{k-r-j-1}-1) \\ \Leftrightarrow\quad & q^{k-r-j}-1 > \frac{n-r-j}{j+1}(q^{k-r-j-1}-1) \\ \Leftrightarrow\quad & q > \frac{n-r-j}{j+1}\,\frac{1-q^{-(k-r-j-1)}}{1-q^{-(k-r-j)}}. \end{aligned}$$

We use the inequality $q > (n-r)/(1-q^{-1})$. This expression is greater than the RHS of the above because of the inequalities $1-q^{-1} \leq 1-q^{-(k-r-j)}$, $1 \geq 1-q^{-(k-r-j-1)}$, and $n-r \geq (n-r-j)/(j+1)$. This proves the lemma. ■

B. Discussion

For correcting narrowband noise in the powerline channel, it is desirable that the symbol weight be close to the minimum possible value of $\lceil n/q \rceil$. It remains open to determine a large subset of the Reed–Solomon code for the case $\rho < 1/2$. It follows from Proposition 6.2 that the rate of this subset cannot exceed the GV-type bound $1-\rho-\delta$. The more interesting question is whether this bound can be achieved, especially in the cases where r is small. The work of Konyagin and Pappalardi [17] gives an affirmative answer to this question in the case of $r = 1$, and for low relative distance δ. They show that the number of permutation polynomials of degree at most $q-1-d$ is approximately $q!/q^d$ for $d \leq 0.03983q$. This is asymptotically, $\frac{1}{q}\log_q(q!/q^d) \simeq 1-\delta$, where $d = \delta n$.

For the other ranges of r, when r is growing with n, we believe that it should be possible to attain the GV bound. For instance, it would be interesting to prove that for large r, $r \geq (k-1)/2$, and for any irreducible polynomial $g(x)$ of degree $k-r-1$, there exists r distinct and nonzero points $\alpha_1, \ldots, \alpha_r$ in $\mathbb{F}_q^*$ such that the polynomial $(x-\alpha_1)\cdots(x-\alpha_r)g(x)$ has symbol weight exactly r. Since most of the count of polynomials attaining constant symbol weight r comes from the count of irreducible polynomials, proving this will show that the rate attains the GV-type bound. We have obtained no counterexample on performing an exhaustive computer search over all such irreducible polynomials $g(x)$ of degree $k-1-r$, with $(k-1)/2 \leq r < k < n$ in all finite fields up to $\mathbb{F}_{17}$. We are unable to verify for larger fields because the computations become prohibitive. We believe that the conjecture is not true for very small r (when r does not grow with n). For instance, for $r = 1$, the polynomial $(x-\alpha)g(x)$, where $g(x) = x^3 + 2$ is an irreducible polynomial in $\mathbb{F}_7$, has symbol weight greater than one for every choice of $\alpha \in \mathbb{F}_7^*$.

VII. Conclusion

We derive the asymptotic estimates of the sizes of symbol weight codes. We also provide means of obtaining lower bounds on such codes and show that it is possible to provide symbol

weight codes with the minimal possible symbol weight via recursive constructions, given we start with a known such code. Finally, we provided constructions of asymptotically good constant symbol weight codes. It remains open to determine families of codes with positive rate and positive relative distance with symbol weights that are optimal or close to the optimal value of $\lceil n/q \rceil$.

ACKNOWLEDGMENT

We thank the Associate Editor N. Kashyap for pointing out an error in an earlier version of the manuscript, and for his comments, which helped us improve the presentation of this article.

REFERENCES

[1] E. Agrell, A. Vardy, and K. Zeger, "Upper bounds for constant-weight codes," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2373–2395, Nov. 2000.

[2] M. Aaltonen, "A new upper bound on nonbinary block codes," *Discrete Math.*, vol. 83, no. 2–3, pp. 139–160, 1990.

[3] C. Bachoc, V. Chandar, G. Cohen, P. Solé, and A. Tchamkerten, "On bounded weight codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6780–6787, Oct. 2011.

[4] I. F. Blake, G. Cohen, and M. Deza, "Coding with permutations," *Inf. Control*, vol. 43, pp. 1–19, 1979.

[5] Y. M. Chee, G. Ge, and A. C. H. Ling, "Group divisible codes and their application in the construction of optimal constant-composition codes of weight three," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3552–3564, Aug. 2008.

[6] Y. M. Chee, S. H. Dau, A. C. H. Ling, and S. Ling, "Linear size optimal q-ary constant-weight codes and constant-composition codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 140–151, Jan. 2010.

[7] W. Chu, C. J. Colbourn, and P. Dukes, "Constructions for permutation codes in powerline communications," *Des. Codes Cryptogr.*, vol. 32, pp. 51–64, 2004.

[8] W. Chu, C. J. Colbourn, and P. Dukes, "On constant composition codes," *Discrete Appl. Math.*, vol. 154, no. 6, pp. 912–929, 2006.

[9] C. J. Colbourn, T. Kløve, and A. C. H. Ling, "Permutation arrays for powerline communication and mutually orthogonal Latin squares," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1289–1291, Jun. 2004.

[10] C. Ding and J. Yin, "Combinatorial constructions of optimal constant-composition codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3671–3675, Oct. 2005.

[11] C. Ding and J. Yin, "A construction of optimal constant composition codes," *Des. Codes Cryptogr.*, vol. 40, pp. 157–165, 2006.

[12] P. J. Dukes, "Coding with injections," *Des. Codes Cryptogr.*, 2011, DOI 10.1007/s10623-011-9547-4.

[13] I. M. Gessel and R. P. Stanley, "Algebraic enumeration," in *Handbook of combinatorics*. Amsterdam, The Netherlands: Elsevier, 1995, vol. 2, pp. 1021–1061.

[14] S. Huczynska, "Powerline communication and the 36 officers problem," *Phil. Trans. R. Soc. A*, vol. 364, pp. 3199–3214, 2006.

[15] S. Huczynska and G. L. Mullen, "Frequency permutation arrays," *J. Combinatorial Des.*, vol. 14, pp. 463–478, 2006.

[16] S. Huczynska, "Equidistant frequency permutation arrays and related constant composition codes," *Des. Codes Cryptogr.*, vol. 54, pp. 109–120, 2010.

[17] S. Konyagin and F. Pappalardi, "Enumerating permutation polynomials over finite fields by degree II," *Finite Fields Appl.*, vol. 12, pp. 26–37, 2006.

[18] V. I. Levenshtein, "Upper bound estimates for fixed weight codes," *Problemy Peredachi Informacii*, vol. 7, no. 4, pp. 3–12, 1970.

[19] V. I. Levenshtein, "Methods for obtaining bounds in metric problems of coding theory," in *Proc. IEEE-USSR Joint Workshop Inf. Theory*, 1976, pp. 126–143.

[20] R. Lidl and H. Niederreiter, "Finite fields," in *Encyclopedia of Mathematics and its Applications*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1997, vol. 20.

[21] J. Lin, J. Chang, R. Chen, and T. Kløve, "Distance-preserving and distance-increasing mappings from ternary vectors to permutations," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1334–1339, Mar. 2008.

[22] Y. Luo, F. Fu, A. J. H. Vinck, and W. Chen, "On constant-composition codes over $\mathbb{Z}_q$," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 3010–3016, Nov. 2003.

[23] J. Luo and T. Helleseth, "Constant composition codes as subcodes of cyclic codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7482–7488, Nov. 2011.

[24] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North Holland, 1991.

[25] A. W. Marshall, B. C. Arnold, and I. Olkin, *Inequalities: Theory of Majorization and its Applications*, 2nd ed. New York: Springer, 2011.

[26] R. Omrani and P. V. Kumar, "Improved constructions and bounds for 2-D optical orthogonal codes," in *Proc. Int. Symp. Inf. Theory*, Adelaide, Australia, 2005, pp. 127–131.

[27] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding in discrete memoryless channels I," *Inf. Control*, vol. 10, pp. 65–103, 1977.

[28] V. R. Sidorenko, "An upper bound on the length of q-ary codes," (in Russian) *Problemy Peredachi Informacii*, vol. 11, no. 3, pp. 14–20, 1975.

[29] D. J. J. Versfeld, A. J. H. Vinck, and H. C. Ferreira, "Reed-Solomon coding to enhance the reliability of M -FSK in a power line environment," in *Proc. Int. Symp. Power Line Commun. Appl.*, Vancouver, BC, Canada, Apr. 2005, pp. 100–104.

[30] D. J. J. Versfeld, A. J. H. Vinck, J. N. Ridley, and H. C. Ferreira, "Constructing coset codes with optimal same-symbol weight for detecting narrowband interference in M -FSK systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6347–6353, Dec. 2010.

[31] A. J. H. Vinck, "Coded modulation for power line communications," *AEU Int. J. Electron. Commun.*, vol. 54, pp. 45–49, 2000.

Yeow Meng Chee (SM'08) received the B.Math. degree in computer science and combinatorics and optimization and the M.Math. and Ph.D. degrees in computer science, from the University of Waterloo, Waterloo, ON, Canada, in 1988, 1989, and 1996, respectively.

Currently, he is an Associate Professor at the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. Prior to this, he was Program Director of Interactive Digital Media R&D in the Media Development Authority of Singapore, Postdoctoral Fellow at the University of Waterloo and IBM's Zürich Research Laboratory, General Manager of the Singapore Computer Emergency Response Team, and Deputy Director of Strategic Programs at the Infocomm Development Authority, Singapore. His research interest lies in the interplay between combinatorics and computer science/engineering, particularly combinatorial design theory, coding theory, extremal set systems, and electronic design automation.

Han Mao Kiah (S'12) received the B.Sc.(Hon) degree in mathematics from the National University of Singapore, Singapore in 2006. Currently, he is working towards his Ph.D. degree at the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore.

His research interest lies in the application of combinatorics to engineering problems in information theory. In particular, his interests include combinatorial design theory, coding theory and power line communications.

Punarbasu Purkayastha (M'10) received the B.Tech. degree in electrical engineering from Indian Institute of Technology, Kanpur, India in 2004, and the Ph.D. degree in electrical engineering from University of Maryland, College Park, in 2010.

Currently, he is a Research Fellow at the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. His research interests include coding theory, combinatorics, information theory and communication theory.