On Constant-Weight Binary B_2-Sequences

Jin Sima, **Yun-Han Li**, **Ilan Shomorony** and **Olgica Milenkovic**
Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, USA
{jsima,yunhanl2,ilans,milenkov}@illinois.edu

***Abstract*—Motivated by applications in polymer-based data storage we introduced the new problem of characterizing the code rate and designing constant-weight binary B_2-sequences. Binary B_2-sequences are collections of binary strings of length n with the property that the real-valued sums of all distinct pairs of strings are distinct. In addition to this defining property, constant-weight binary B_2-sequences also satisfy the constraint that each string has a fixed, relatively small weight ω that scales linearly with n. The constant-weight constraint ensures low-cost synthesis and uniform processing of the data readout via tandem mass spectrometers. Our main results include upper bounds on the size of the codes formulated as entropy-optimization problems and constructive lower bounds based on Sidon sequences.**

I. Introduction

Binary B_2-sequences were introduced by Lindström in [1] and were subsequently studied in a number of follow-up works [2], [3], [4]. Binary B_2-sequences represent a set (codebook) of binary vectors of some fixed length such that the entry-wise real-valued sums of all pairs of codevectors from the set are distinct; hence, given the sum one can uniquely determine the vectors that were summed up. Since their introduction, these sequences have found many applications such as for search algorithms [5], [6], multiple access system design [7], [8], and data fingerprinting [9], [3].

Some more recent applications of binary B_2-sequences include *polymer-based data storage.* Nonvolatile storage systems based on DNA and other native macromolecules and synthetic polymers hold the promise of ultrahigh storage densities, long-term readout compatibility and exceptional durability [10], [11], [12], [13], [14], [15], [16], [17]. Synthetic polymers are binary molecular storage media that represent 0s and 1s using polymers of significantly different masses [18]. A user-defined binary string is created by stitching together the polymer symbols in the required order and it is read by measuring the masses of prefixes and suffixes (or all substrings) of the polymer strings [19], [20]. To ensure unique reconstruction of mixture of polymer strings based on their prefix and suffix compositions only, one needs to follow a more involved process, described in [21], [22]. There, binary B_h-sequences (codes) [1], [3], [4] are used to ensure that the sums of masses of prefixes of the same length uniquely determine the strings themselves. Since in practice the polymer used to represent 1s has a significantly higher mass than the polymer used to represent 0s, the mass discrepancy can lead to high fragmentation loss and significantly increased chemical synthesis cost, using binary B_2-sequences of relatively small weight is desirable. This motivates introducing the problem of *constant-weight B_2-sequence* design, complementing results from the well-studied area of constant-weight error-control coding, [23], [24], [25], [26], [27], [28], [29].

This work was supported by the NSF grants 1816913 and 2007597.

The main results of our work include information-theoretic, prefix-suffix splitting upper bounds on the size of constant-weight binary B_2-sequences for which the weight ω scales linearly with n. Unlike its unconstrained counterpart, the strongest upper bound is given in terms of an optimization problem that has to be solved numerically. In addition, we also provide constructive lower bounds based on Sidon sequences. The equally important problem of designing binary B_2-sequences of bounded, small weight, will be discussed elsewhere.

The paper is organized as follows. Section II presents the notation, relevant concepts and a generalization of the approach from [1] to the case of constant-weight binary strings. Section II-A contains our main result, the sharpest known upper bound on the size of binary constant-weight B_2-sequences. Constructive lower bounds are presented in Section III.

II. Preliminaries and Entropy Bounds

We denote sets by calligraphic upper-case letters and vectors by boldface lower-case letters. Cardinalities of sets are denoted by upper-case letters. We also use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. All logarithms, unless stated otherwise, are taken base-2.

A set $\mathcal{A}_n \subset \{0,1\}^n$ of binary vectors is called a B_2-sequence set if real-valued sums of all distinct pairs of strings $\mathbf{c}_1 + \mathbf{c}_2$, $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{A}_n$, are distinct. A B_2-sequence set $\mathcal{A}_n^\omega$ is said to have constant weight ω if every vector $\mathbf{c} = (c_1, \ldots, c_n) \in \mathcal{A}_n^\omega$ has Hamming weight $|\mathbf{c}| \triangleq \sum_{i=1}^n c_i = \omega$.

Let A_n^ω be the size of the largest constant-weight B_2-sequence set $\mathcal{A}_n^\omega$ of weight ω. We are interested in the asymptotic behavior of A_n^ω, for constant $\bar{\omega} = \frac{\omega}{n}$, or more precisely, in the asymptotic code rate $R_{\bar{\omega}} = \limsup_{n\to\infty} \frac{\log A_n^{\bar{\omega} n}}{n}$. The results established in [1], [3] imply that for unrestricted binary B_2-sequence sets, the asymptotic code rate satisfies ≤ 0.5753, which also establishes $R_{\bar{\omega}} \leq 0.5753$ for any $\bar{\omega} \in [0,1]$. We seek to improve this upper bound on $R_{\bar{\omega}}$ for $\bar{\omega} \in [0, 1/2)$.

A simple asymptotic upper bound on $R_{\bar{\omega}}$ can be derived using information-theoretic arguments (attributed to Katona for the case of unrestricted sequences), by assuming a *uniform probability distribution* on the set of all possible *ordered* pairs $(\boldsymbol{C}, \boldsymbol{C}')$ of codevectors and invoking the fact that each pair

results in a unique sum. Note that C and C' are chosen independently but are allowed to be equal. To obtain an upper bound on $R_{\bar{\omega}}$, we assume that the size of the constant-weight B_2 code of length n equals A_n^ω. Then, the entropy of all ordered pairs of codevectors equals $2\log(A_n^\omega)$. Let $G = \mathbb{1}(\boldsymbol{C} > \boldsymbol{C}')$ be an indicator random variable for the event that $\boldsymbol{c}$ is lexicographically ranked higher than $\boldsymbol{c}'$. Since the sums of all unordered pairs are all distinct, there is a bijection between the conditional probability space of the pair $(\boldsymbol{C}, \boldsymbol{C}')$ given G and the probability space of a sequence of random variables, $X_i = C_i + C_i'$, $i \in [n]$, representing the coordinates of the sum vector. Consequently,

$$\begin{aligned} 2\log(A_n^\omega) &= H(\boldsymbol{C}, \boldsymbol{C}'|G) + H(G) \\ &= H(X_1, X_2, \ldots, X_n|G) + H(G) \\ &\leq H(X_1) + H(X_2) + \ldots + H(X_n) + 1. \end{aligned} \tag{1}$$

Assume that the probability of observing a 1 at the ith coordinate (i.e., C_i) of the B_2-sequence code equals p_i, $i = 1, \ldots, n$. Given that the weight of the binary vectors is ω, we have $\frac{1}{n}(p_1 + p_2 + \ldots + p_n) = \frac{\omega}{n} = \bar{\omega}$. The entropy of the ith coordinate of all possible 2-sums equals $H(X_i)$ and is equal to the entropy of a Binomial$(2, p_i)$ distribution, i.e., the distribution $\mathbf{p}^i$ with probabilities of 0, 1 and 2 equal to

$$p_0^i = (1-p_i)^2,\ p_1^i = 2\,p_i\,(1-p_i),\ p_2^i = p_i^2,\ i \in [n]$$

respectively. Note that $H(X_i) = H(\mathbf{p}_i)$ equals

$$\begin{aligned} &- p_i^2 \log p_i^2 - (1-p_i)^2 \log(1-p_i)^2 - 2p_i(1-p_i)\log p_i \\ &- 2p_i(1-p_i)\log(1-p_i) - 2p_i(1-p_i) \\ &= -\,2p_i \log p_i - 2(1-p_i)\log(1-p_i) - 2p_i + 2p_i^2 \triangleq H_{bin}(p_i). \end{aligned}$$

Since $\frac{d^2 H_{bin}(p_i)}{dp_i^2} = -\frac{2}{\ln(2)}(\frac{1}{p_i} + \frac{1}{1-p_i}) + 4 < 0$, the function $H_{bin}(p_i)$ is concave in p_i. Hence, we have

$$\begin{aligned} &H(X_1) + H(X_2) + \ldots + H(X_n) \\ &= H_{bin}(p_1) + H_{bin}(p_2) + \ldots + H_{bin}(p_n) \\ &\leq n H_{bin}\left(\frac{\sum_{i=1}^n p_i}{n}\right) = n H_{bin}\left(\bar{\omega}\right). \end{aligned} \tag{2}$$

Note that $H_{bin}(\bar{\omega})$ stands for the entropy of a Binomial$(2, \frac{\omega}{n})$ distribution, i.e., the distribution

$$p_0 = (1-\bar{\omega})^2,\ p_1 = 2\,\bar{\omega}\,(1-\bar{\omega}),\ p_2 = \bar{\omega}^2.$$

By (1) and (2) it follows that $2\log(A_n^\omega) \leq n H_{bin}(\bar{\omega}) + 1$ and

$$R_{\bar{\omega}} \leq \frac{1}{2} H_{bin}\left(\bar{\omega}\right). \tag{3}$$

While (3) provides a good starting upper bound for $R_{\bar{\omega}}$, we note that an alternative entropy bound can be obtained by considering the coordinates of the difference of codevector, rather than the sum of codevectors. Specifically, we assume a uniform distribution on the set of ordered pairs of codevectors $(\boldsymbol{C}, \boldsymbol{C}')$ in a constant-weight B_2 code, so that the entropy of the ordered pair is as before given by $2\log(A_n^\omega)$. Let $Y_i = C_i - C_i'$, $i \in [n]$, be the sequence of random variables representing the values of coordinates of the difference of vector in the ordered pair. Then, either one of the following holds: (1) $Y_i = 0$ for $i \in [n]$, whenever the two vectors in the ordered pair are equal; (2) there is a one to one mapping between $Y_1, Y_2, \ldots, Y_n$ and the ordered pair if the two vectors in the pair are not equal. This follows because if any two different pairs $\mathbf{c}_1, \mathbf{c}_2 \in A_n^\omega$ and $\mathbf{c}_1, \mathbf{c}_2 \in A_n^\omega$ have the same difference $\mathbf{c}_1 - \mathbf{c}_2 = \mathbf{c}_3 - \mathbf{c}_4$, then the two pairs $\mathbf{c}_1, \mathbf{c}_4$ and $\mathbf{c}_2, \mathbf{c}_3$ have the same sum $\mathbf{c}_1 + \mathbf{c}_4 = \mathbf{c}_2 + \mathbf{c}_3$, which violates the B_2 constraint. Next, let E be the event $\{(Y_1, \ldots, Y_n) = 0^n\}$. Then, we can upper-bound $2\log(A_n^\omega)$ by

$$\begin{aligned} &H(C, C'|Y_1, Y_2, \ldots, Y_n) + H(Y_1, Y_2, \ldots, Y_n) \\ &= \left(1 - \frac{1}{A_n^\omega}\right) H(C, C'|Y_1, Y_2, \ldots, Y_n, E^c) \\ &\quad + \frac{1}{A_n^\omega} H(C, C'|Y_1, Y_2, \ldots, Y_n, E) + H(Y_1, Y_2, \ldots, Y_n) \\ &= \frac{\log A_n^\omega}{A_n^\omega} + H(Y_1, Y_2, \ldots, Y_n) \\ &\leq H(Y_1) + H(Y_2) + \ldots + H(Y_n) + \frac{\log A_n^\omega}{A_n^\omega}, \end{aligned}$$

where $(1 - \frac{1}{A_n^\omega})$ equals the probability that the two codevectors in the pair are not equal. Assume that $P\{C_i = 1\} = q_i$. Then, as before, we have $\frac{1}{n}(q_1 + q_2 + \ldots + q_n) = \bar{\omega}$. The distribution $\mathbf{q}^i$ of Y_i is given by

$$q_0^i = q_i^2 + (1-q_i)^2,\ q_1^i = q_i(1-q_i),\ q_{-1}^i = q_i(1-q_i),\ i \in [n].$$

Since the entropy function $H(\mathbf{q}^i)$ is concave in $\mathbf{q}^i$,

$$\begin{aligned} &H(Y_1) + H(Y_2) + \ldots + H(Y_n) \\ &= H(\mathbf{q}^1) + H(\mathbf{q}^2) + \ldots + H(\mathbf{q}^n) \\ &\leq n H\left(\frac{\mathbf{q}^1 + \mathbf{q}^2 + \ldots + \mathbf{q}^n}{n}\right) \\ &= n\Big(-t\log t - (1-t)\log\left(\frac{1-t}{2}\right)\Big), \end{aligned}$$

where $t = \frac{\sum_{i=1}^n [q_i^2 + (1-q_i)^2]}{n}$. Note that $t \geq \bar{\omega}^2 + (1-\bar{\omega})^2 \geq \frac{1}{2}$ by Jensen's inequality, and that for $t \geq \frac{1}{2}$ the function $-t\log t - (1-t)\log(\frac{1-t}{2})$ decreases as t increases. Hence, by the previous inequality we have

$$H(Y_1) + H(Y_2) + \ldots + H(Y_n) \leq n H(\mathbf{q}),$$

where the distribution $\mathbf{q}$ is given by

$$q_0 = \bar{\omega}^2 + (1-\bar{\omega})^2,\ q_1 = \bar{\omega}(1-\bar{\omega}),\ q_{-1} = \bar{\omega}(1-\bar{\omega}).$$

As a result, $2\log(A_n^\omega) \leq n H(\mathbf{q}) + \frac{\log A_n^\omega}{A_n^\omega}$, and

$$R_{\bar{\omega}} \leq \frac{1}{2} H(\mathbf{q}). \tag{4}$$

It can easily be shown that the upper bound (3) is strictly better than (4). However, as we will see later, the bound (4) is more useful for deriving upper bounds than those obtained from direct information-theoretic arguments.

In Table I, we provide numerical values for the information-theoretic bounds derived in this section, along with those of the improved upper bounds to be described in the subsequent

exposition, for different values of weights ω that scale linearly with n. As may be seen, as ω decreases, all upper bounds become closer in value. Additional results included in the table are lower bounds, discussed in more detail in the last section.

TABLE I
UPPER AND LOWER BOUNDS ON THE SIZE OF BINARY, CONSTANT-WEIGHT B_2-SEQUENCES

The value of $\bar{\omega}$	0.5	0.4	0.345	0.2	0.1	0.05	0.02
Entropy bound (3)	0.75	0.731	0.704	0.562	0.379	0.239	0.122
Entropy bound (4)	0.75	0.739	0.723	0.612	0.43	0.274	0.139
Upper bound (5)	0.6	0.6	0.594	0.515	0.365	0.235	0.121
Upper bound (6)	0.6	0.59	0.575	0.487	0.349	0.228	0.12
Lower bound	0.25	0.259	0.263	0.232	0.166	0.108	0.056

A. Improved Upper Bounds

An improved upper bound on $R_{\bar{\omega}}$ for our constrained B_2 codebook design can be obtained by adapting and generalizing a recent approach from [4] for unconstrained B_2 vectors. The underlying proof combines entropy bounds with a prefix-suffix decomposition approach first reported in [2]. For completeness, we first describe how to extend the prefix-suffix decomposition approach for constant-weight B_2 codebooks. Afterwards, we improve the two bounds – the information-theoretic and prefix-suffix bound – by combining them [4]. The main differences between the approaches designed for general codebooks [4], [2] and our approach is that we use more elaborate entropy bounds, group the codevectors $\mathbf{c} \in \mathcal{A}_n^{\omega}$ based on the weight of their prefixes and invoke specialized counting techniques. Importantly, the approach in [4] does not improve the result that can be obtained purely through the use of prefix-suffix decompositions [2], while our scheme improves *both the entropy and prefix-suffix approach* for constant-weight codebooks. The proof is deferred to the appendix.

Lemma 1. *Let every codevector be split as $\mathbf{c} = \mathbf{ab} \in \mathcal{A}_n^{\omega}$, where $\mathbf{a} \in \{0,1\}^e$ and $\mathbf{b} \in \{0,1\}^{n-e}$. Let $e = \bar{e}n$, where $\bar{e}$ is a constant in $[0,1]$. Then, $R_{\bar{\omega}}$ can be upper bounded by*

$$\min_{\bar{e}\in[0,1]} \max_{\bar{\omega}'\in[\max\{0,\bar{\omega}-1+\bar{e}\},\min\{\bar{e},\bar{\omega}\}]} \max\left\{ H(\frac{\bar{\omega}'}{\bar{e}})\cdot\bar{e}, \frac{1}{2}\Big[H(\frac{\bar{\omega}'}{\bar{e}})\cdot\bar{e} + H(\frac{2\bar{\omega}''}{1-\bar{e}})\cdot(1-\bar{e}) + 2\bar{\omega}''\Big]\right\}, \quad (5)$$

where $\bar{\omega}'' = \min\{\bar{\omega} - \bar{\omega}', \frac{1-\bar{e}}{4}\}$. By optimizing over $\bar{e}$, numerical values for the bound can be found for different ω's. A sampling of the results is shown in Table I.

We describe next our main result, constituting a sharper asymptotic upper bound on constant-weight binary B_2-sequences.

Theorem 1. *The rate $R_{\bar{\omega}}$ is upper bounded by*

$$R_{\bar{\omega}} \le \min_{\bar{e}\in[0,1]} \max_{\bar{\omega}'\in[\max\{0,\bar{\omega}-1+\bar{e}\},\min\{\bar{e},\bar{\omega}\}]} \max\left\{\bar{e}\, H\left(\frac{\bar{\omega}'}{\bar{e}}\right), \frac{1}{2}\Big[\bar{e}\, H\left(\frac{\bar{\omega}'}{\bar{e}}\right) + (1-\bar{e})H(p_0,p_1,p_{-1})\Big]\right\}, \quad (6)$$

where $p_0 = \frac{(\bar{\omega}'')^2+(1-\bar{e}-\bar{\omega}'')^2}{(1-\bar{e})^2}$, $p_1 = p_{-1} = \frac{1-p_0}{2}$, and $\bar{\omega}'' = \bar{\omega} - \bar{\omega}'$.

In the following, we prove Theorem 1. Let

$$\mathcal{B}_n^{\omega'} = \{\mathbf{c} : \mathbf{c} = \mathbf{ab} \in \mathcal{A}_n^{\omega},\ \mathbf{a} \in \{0,1\}^e,\ |\mathbf{a}| = \omega'\}, \quad (7)$$

where $e = \bar{e}n$ and $\omega' = \bar{\omega}'n$ are constants s.t. $\bar{e}, \bar{\omega}' \in [0,1]$, be the set of codevectors in $\mathcal{A}_n^{\omega}$ whose prefixes of length e have weight $\omega' \in [\max\{0, \omega - n + e\}, \min\{e, \omega\}]$. For notational convenience, we also use $f = n - e$ to denote the length of the suffixes. Note that

$$A_n^{\omega} = \sum_{\omega'\in[\max\{0,\omega-f\},\min\{e,\omega\}]} |\mathcal{B}_n^{\omega'}|. \quad (8)$$

Hence, we need to establish an upper bound on $|\mathcal{B}_n^{\omega'}|$ for any n, e and $\omega' \in [\max\{0, \omega - f\}, \min\{e, \omega\}]$.

Lemma 2. *For any $\omega' \in [\max\{0,\omega-f\}, \min\{e,\omega\}]$, we have*

$$\log|\mathcal{B}_n^{\omega'}| \le \max\left\{ e\, H\left(\frac{\omega'}{e}\right) + \log n, \frac{1}{2}\Big[e\, H\left(\frac{\omega'}{e}\right) + fH(p_0,p_1,p_{-1})\Big] + 1\right\}, \quad (9)$$

where $p_0 = \frac{(\omega'')^2+(f-\omega'')^2}{f^2}$, $p_1 = p_{-1} = \frac{1-p_0}{2}$, and $\omega'' = \omega - \omega'$. The function $H(x) = -x\log x - (1-x)\log(1-x)$ stands for the binary Shannon entropy function, while the function $H(p_0,p_1,p_{-1}) = -p_0\log p_0 - p_1\log p_1 - p_{-1}\log p_{-1}$ stands for the entropy of a ternary random variable with the distribution (p_0,p_1,p_{-1}) described above.

Proof. Fix ω' and set $B_n^{\omega'} = |\mathcal{B}_n^{\omega'}|$. Let $\{\mathbf{a}_1,\ldots,\mathbf{a}_r\} = \{\mathbf{a} : \mathbf{ab} \in \mathcal{B}_n^{\omega'} \text{ for some } \mathbf{b}\}$ be the set of all possible prefixes of the codevectors in $\mathcal{B}_n^{\omega'}$. Note that each $\mathbf{a}_i$ has weight ω' and that there are at most r different such vectors, where $r \le \binom{e}{\omega'}$. Let $\mathcal{S}_i = \{\mathbf{b} : \mathbf{a}_i\mathbf{b} \in \mathcal{B}_n^{\omega'}\}$, $i \in [r]$, be the (possibly empty) set of suffixes of codevectors in $\mathcal{B}_n^{\omega'}$ that have prefix $\mathbf{a}_i$. Then,

$$B_n^{\omega'} = \sum_{i=1}^{r} |\mathcal{S}_i|. \quad (10)$$

Now consider the set of all pairs of suffixes that belong to the same group $\mathcal{S}_i$ for some $i \in [r]$, denoted by

$$\mathcal{D} = \cup_{i=1}^{r}\{(\mathbf{b}_1,\mathbf{b}_2) : \mathbf{b}_1,\mathbf{b}_2 \in \mathcal{S}_i\}. \quad (11)$$

Note that $\mathbf{b}_1$ and $\mathbf{b}_2$ are allowed to be the same and that $(\mathbf{b}_1,\mathbf{b}_2)$ and $(\mathbf{b}_2,\mathbf{b}_1)$ are considered two different pairs, provided $\mathbf{b}_1,\mathbf{b}_2 \in \mathcal{S}_i$, $i \in [r]$, are distinct. Hence, $\mathcal{D}$ is a multiset. Then

$$|\mathcal{D}| = \sum_{i=1}^{r} |\mathcal{S}_i|^2 \ge \frac{(\sum_i^r |\mathcal{S}_i|)^2}{r} = \frac{(B_n^{\omega'})^2}{r}, \quad (12)$$

where the bound follows from Cauchy-Schwarz's inequality. Furthermore, consider the differences between all pairs in $\mathcal{D}$,

$$\mathcal{Z} = \{\mathbf{b}_1 - \mathbf{b}_2 : (\mathbf{b}_1,\mathbf{b}_2) \in \mathcal{D}\}, \quad (13)$$

where $\mathcal{Z}$ is a multiset. The multiplicity of 0^f (the all 0 vector of length f) in $\mathcal{Z}$ is exactly $B_n^{\omega'}$. In addition, the multiplicity of any nonzero element in $\mathcal{Z}$ is exactly one. To see this, suppose on the contrary that there exist different pairs of unequal elements $(\mathbf{b}_1, \mathbf{b}_2), (\mathbf{b}_3, \mathbf{b}_4) \in \mathcal{D}$ satisfying $\mathbf{b}_1 - \mathbf{b}_2 = \mathbf{b}_3 - \mathbf{b}_4$. By definition of $\mathcal{D}$, we have that $\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{S}_i$ for some $i \in [r]$ and $\mathbf{b}_3, \mathbf{b}_4 \in \mathcal{S}_j$ for some $j \in [1, r]$. This implies that $\mathbf{a}_i\mathbf{b}_1, \mathbf{a}_i\mathbf{b}_2, \mathbf{a}_j\mathbf{b}_3, \mathbf{a}_j\mathbf{b}_4$ are codevectors in $\mathcal{B}_n^{\omega'}$. Then,

$$\mathbf{a}_i\mathbf{b}_1 - \mathbf{a}_i\mathbf{b}_2 = \mathbf{a}_j\mathbf{b}_3 - \mathbf{a}_j\mathbf{b}_4, \tag{14}$$

contradicting the fact that $\mathcal{B}_n^{\omega'}$ is a binary B_2-sequence.

Next, we generalize the derivation of an information-theoretic argument from [4]. Uniformly at random pick a pair from $\mathcal{D}$ and denote the outcome by a pair of random variables $(\boldsymbol{X}, \boldsymbol{Y})$. Then, the difference $\boldsymbol{X} - \boldsymbol{Y}$ is uniformly distributed over $\mathcal{Z} \backslash \{0^f\}$ (i.e, conditioned on $\boldsymbol{X} - \boldsymbol{Y} \neq 0^f$). Let E be the event $\{\boldsymbol{X} - \boldsymbol{Y} = 0^f\}$. Then, $H(\boldsymbol{X}, \boldsymbol{Y})$ equals

$$\begin{aligned} &H(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{X} - \boldsymbol{Y}) = H(\boldsymbol{X} - \boldsymbol{Y}) + H(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{X} - \boldsymbol{Y}) \\ &= H(\boldsymbol{X} - \boldsymbol{Y}) + \Pr(E) H(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{X} - \boldsymbol{Y}, E) \\ &+ \Pr(E^c) H(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{X} - \boldsymbol{Y}, E^c). \end{aligned} \tag{15}$$

Clearly, $H(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{X} - \boldsymbol{Y}, E^c) = 0$, since different nonzero elements in $\mathcal{Z}$ have multiplicity one. In addition,

$$H(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{X} - \boldsymbol{Y}, E) = \log B_n^{\omega'}, \;\; \Pr(E) = \frac{B_n^{\omega'}}{|\mathcal{D}|} \leq \frac{r}{B_n^{\omega'}},$$

where the inequality follows from (12). Therefore,

$$\Pr(E) H(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{X} - \boldsymbol{Y}, E) \leq \frac{r}{B_n^{\omega'}} \log B_n^{\omega'}. \tag{16}$$

Combining (15), (16) with $H(\boldsymbol{X}, \boldsymbol{Y}) = \log |\mathcal{D}| \geq \frac{(B_n^{\omega'})^2}{r}$, we obtain

$$\log(\frac{(B_n^{\omega'})^2}{r}) \leq H(\boldsymbol{X} - \boldsymbol{Y}) + \frac{r}{B_n^{\omega'}} \log B_n^{\omega'}. \tag{17}$$

In order to obtain an upper bound on $B_n^{\omega'}$, we need an upper bound on $H(\boldsymbol{X} - \boldsymbol{Y})$ specialized for constant-weight vectors.

Let n_{ij}, $i \in [f]$, $j \in [r]$, be the number of suffixes in $\mathcal{S}_j$ whose ith coordinate is 1. By subadditivity of the entropy function we have

$$H(\boldsymbol{X} - \boldsymbol{Y}) \leq \sum_{i=1}^{f} H(X_i - Y_i) \leq f H(\frac{\sum_{i=1}^{f} \mathbf{p}^i}{f}), \tag{18}$$

where $\mathbf{p}^i = (p_0^i, p_1^i, p_{-1}^i)$ is the distribution of $X_i - Y_i$,

$$\begin{aligned} &p_0^i = \frac{\sum_{j=1}^{r} [n_{ij}^2 + (|\mathcal{S}_j| - n_{ij})^2]}{\sum_{j=1}^{r} |\mathcal{S}_j|^2}, \;\; p_1^i = \frac{\sum_{j=1}^{r} n_{ij}(|\mathcal{S}_j| - n_{ij})}{\sum_{j=1}^{r} |\mathcal{S}_j|^2}, \\ &p_{-1}^i = \frac{\sum_{j=1}^{r} n_{ij}(|\mathcal{S}_j| - n_{ij})}{\sum_{j=1}^{r} |\mathcal{S}_j|^2}. \end{aligned} \tag{19}$$

We show next that the average distribution, denoted as

$$\mathbf{p}^* = \frac{\sum_{i=1}^{f} \mathbf{p}^i}{f}, \tag{20}$$

satisfies $p_0^* \geq \frac{(\omega'')^2 + (f - \omega'')^2}{f^2}$, where $\omega'' = \omega - \omega'$ is the weight for all suffixes of codevectors in $\mathcal{B}_n^{\omega'}$, i.e., the weight of vectors in $\mathcal{S}_j$, $j \in [r]$. From (19), it follows

$$\begin{aligned} p_0^* &= \frac{\sum_{i=1}^{f} p_0^i}{f} = \frac{\sum_{j=1}^{r} \left(\sum_{i=1}^{f} [n_{ij}^2 + (|\mathcal{S}_j| - n_{ij})^2] \right)}{f(\sum_{j=1}^{r} |\mathcal{S}_j|^2)} \\ &\geq \frac{\sum_{j=1}^{r} \left(\frac{(\sum_{i=1}^{f} n_{ij})^2}{f} + \frac{(\sum_{i=1}^{f} (|\mathcal{S}_j| - n_{ij})^2}{f} \right)}{f(\sum_{j=1}^{r} |\mathcal{S}_j|^2)} \\ &\overset{(a)}{=} \frac{\sum_{j=1}^{r} \left(|\mathcal{S}_j|^2 (\omega'')^2 + |\mathcal{S}_j|^2 (f - \omega'')^2 \right)}{f^2 (\sum_{j=1}^{r} |\mathcal{S}_j|^2)} \\ &= \frac{(\omega'')^2 + (f - \omega'')^2}{f^2}, \end{aligned} \tag{21}$$

where (a) follows from the fact that the weight of the vectors in $\mathcal{S}_j$ is fixed and equal to ω''. In addition, we have $p_1^* = p_{-1}^*$ since $p_1^i = p_{-1}^i$ from (19). Note that $p_0^* \geq \frac{1}{2}$ and that the entropy function $H(p_0^*, \frac{1-p_0^*}{2}, \frac{1-p_0^*}{2})$ is decreasing in p_0^* when $p_0^* \geq \frac{1}{2}$. Therefore, combined with (21) and (18), we have

$$H(\boldsymbol{X} - \boldsymbol{Y}) \leq f H(p_1, p_0, p_{-1}), \tag{22}$$

where $p_0 = \frac{(\omega'')^2 + (f - \omega'')^2}{f^2}$ and $p_1 = p_{-1} = \frac{1 - p_0}{2}$. Combining (17) (22), we obtain

$$\log B_n^{\omega'} \leq \frac{1}{2}(\log r + f H(p_0, p_1, p_{-1})) + \frac{r}{2B_n^{\omega'}} \log B_n^{\omega'}. \tag{23}$$

Finally, to prove (9), suppose to the contrary that

$$\begin{aligned} &\log B_n^{\omega'} > e\, H\left(\frac{\omega'}{e}\right) + \log n, \text{ and} \\ &\log B_n^{\omega'} > \frac{1}{2}\Big[e\, H\left(\frac{\omega'}{e}\right) + f H(p_0, p_1, p_{-1})\Big] + 1, \end{aligned} \tag{24}$$

From (23), (24), and the fact that $r \leq 2^{e H(\frac{\omega'}{e})}$, we have

$$1 < \frac{r}{2B_n^{\omega'}} \log B_n^{\omega'}, \tag{25}$$

as well as the inequality below which contradicts (25):

$$B_n^{\omega'} > n 2^{e\, H\left(\frac{\omega'}{e}\right)} \geq nr \geq r \log B_n^{w'}. \tag{26}$$

□

By combining Lemma 2 and (8) we conclude that

$$\begin{aligned} \log A_n^{\omega} \leq \max_{\omega' \in [\max\{0, \omega - f\}, \min\{e, \omega\}]} \max \Big\{ &e\, H\left(\frac{\omega'}{e}\right), \\ &\frac{1}{2}\Big[e\, H\left(\frac{\omega'}{e}\right) + f H(p_0, p_1, p_{-1})\Big] \Big\} + \log n, \end{aligned} \tag{27}$$

where $p_0 = \frac{(\omega'')^2 + (f - \omega'')^2}{f^2}$, $p_1 = p_{-1} = \frac{1 - p_0}{2}$, and $\omega'' = \omega - \omega'$, for any choice of e and f such that $e + f = n$. Therefore, we have (6). Note that the bound (6) is smaller than the bound 0.6 in [2] whenever $\omega < \frac{n}{2}$, and is smaller than the best known upper bound 0.5753 for unconstrained binary B_2-sequences reported in [3] whenever $\omega \leq 0.345n$. See Table I for more details regarding the actual values of the upper bounds.

III. A Lower Bound

We describe next a construction for constant-weight binary B_2 codes $\mathcal{A}_n^\omega$ of size $(\frac{n}{\omega})^{\frac{\omega}{2}+o(\omega)}$ and

$$(\left\lfloor \frac{n}{\omega} \right\rfloor)^{\frac{\omega\lceil \frac{n}{\omega}\rceil - n}{2(\lceil \frac{n}{\omega}\rceil - \lfloor \frac{n}{\omega}\rfloor)}} (\left\lceil \frac{n}{\omega} \right\rceil)^{\frac{n-\omega\lfloor \frac{n}{\omega}\rfloor}{2(\lceil \frac{n}{\omega}\rceil - \lfloor \frac{n}{\omega}\rfloor)}} 2^{o(\omega)},$$

for the case that $\frac{n}{\omega}$ is an integer and a noninteger real value, respectively. The construction implies the following result.

Theorem 2. *The rate $R_{\bar{\omega}}$ is lower bounded by*

$$R_{\bar{\omega}} \geq \frac{\bar{\omega}}{2}\log(\frac{1}{\bar{\omega}})$$

whenever $\frac{1}{\bar{\omega}}$ is an integer, and

$$R_{\bar{\omega}} \geq \frac{\bar{\omega}\lceil \frac{1}{\bar{\omega}}\rceil - 1}{2(\lceil \frac{1}{\bar{\omega}}\rceil - \lfloor \frac{1}{\bar{\omega}}\rfloor)}\log(\left\lfloor \frac{1}{\bar{\omega}} \right\rfloor) + \frac{1-\bar{\omega}\lfloor \frac{1}{\bar{\omega}}\rfloor}{2(\lceil \frac{1}{\bar{\omega}}\rceil - \lfloor \frac{1}{\bar{\omega}}\rfloor)}\log(\left\lceil \frac{1}{\bar{\omega}} \right\rceil)$$

otherwise.

An important observation is that our construction, although conceptually simple, results in codes with rate at least $\frac{1}{4}$th of the largest possible rate of unconstrained constant-weight codes, $\binom{n}{\omega}$.

In what follows, we assume for simplicity that $\frac{n}{\omega}$ is an integer, and as before, we let $\omega \leq \frac{n}{2}$. The idea is to find a surjective linear mapping $F : \{0,1,2\}^n \to [0, (\frac{n}{\omega})^\omega - 1]$ that converts any length-n vector over the alphabet $\{0,1,2\}$ into an integer in $[0, (\frac{n}{\omega})^\omega - 1]$. More precisely, the mapping F is required to satisfy the following two properties:

(A) For any integer $i \in [0, (\frac{n}{\omega})^\omega - 1]$, there exists a vector $\mathbf{c} \in \{0,1\}^n$ of weight ω, such that $F(\mathbf{c}) = i$.

(B) For any vectors $\mathbf{c}, \mathbf{c}' \in \{0,1\}^n$, we have that $F(\mathbf{c}) + F(\mathbf{c}') = F(\mathbf{c}+\mathbf{c}')$. Note that $\mathbf{c}+\mathbf{c}' \in \{0,1,2\}^n$.

Given the mapping F, we construct an integer Sidon set [30] from the set $[0, (\frac{n}{\omega})^\omega - 1]$. By the Bose-Chawla construction, there exists a set of integers $\{i_1, \ldots, i_{(\frac{n}{\omega})^{\frac{\omega}{2}+o(\omega)}}\} \subset [0, (\frac{n}{\omega})^\omega - 1]$ of size $(\frac{n}{\omega})^{\frac{\omega}{2}+o(\omega)}$ such that the sums of any two integers in the set are distinct. Then from property (A) of the mapping F, for every i_j, $j \in [1, (\frac{n}{\omega})^{\frac{\omega}{2}+o(\omega)}]$, there exists a vector $\mathbf{c}_j \in \{0,1\}^n$ of weight ω such that $F(\mathbf{c}_j) = i_j$. Finally, by property (B) of the mapping F and the definition of the set $\{i_1, \ldots, i_{(\frac{n}{\omega})^{\frac{\omega}{2}+o(\omega)}}\}$, the set $\{\mathbf{c}_1, \ldots, \mathbf{c}_{(\frac{n}{\omega})^{\frac{\omega}{2}+o(\omega)}}\}$ is a binary B_2 codebook of weight ω.

For any integer $k \in [0, n-1]$, let $k = a_k\omega + b_k$, where $a_k = \lfloor \frac{k}{\omega} \rfloor$ and $b_k = k \bmod \omega$. For any $\mathbf{c} \in \{0,1,2\}^n$ define

$$F(\mathbf{c}) \triangleq \sum_{i=1}^{n} a_{i-1}\left(\frac{n}{\omega}\right)^{b_{i-1}} c_i. \tag{28}$$

It is obvious that F satisfies property (B). To show that F satisfies (A), we note that any integer $m \in [0, (\frac{n}{\omega})^\omega - 1]$ has a $\frac{n}{\omega}$-ary representation $m = \sum_{i=0}^{\omega-1} m_i \left(\frac{n}{\omega}\right)^i$, where $m_i \in [0, \frac{n}{\omega} - 1]$. Let $\mathbf{c}_m$ be a vector in $\{0,1\}^n$, whose indices of the 1 bits are given by $\{m_i\omega + i : i \in [0, \omega-1]\}$. Then, $\mathbf{c}_m$ has weight ω and $F(\mathbf{c}_m) = \sum_{i=0}^{\omega-1} a_{m_i\omega+i}(\frac{n}{\omega})^{b_{m_i\omega+i}} = m$. Hence F satisfies property (B).

Appendix

We first group the vectors $\mathbf{c} \in \mathcal{A}_n^\omega$ based on the weight of $\mathbf{a}$, and use the definition $\mathcal{B}^{\omega'} = \{\mathbf{c} : \mathbf{c} = \mathbf{ab} \in \mathcal{A}_n^\omega,\ \mathbf{a} \in \{0,1\}^e,\ |\mathbf{a}| = \omega'\}$. The first part of the proof is similar to that of Lemma 2. We set $\{\mathbf{a}_1, \ldots, \mathbf{a}_r\} = \{\mathbf{a} : \mathbf{a} \in \{0,1\}^e, \mathbf{ab} \in \mathcal{B}_n^{\omega'} \text{ for some } \mathbf{b}\}$, $\mathcal{S}_i = \{\mathbf{b} : \mathbf{a}_i\mathbf{b} \in \mathcal{B}_n^{\omega'}\}$, and $\mathcal{D} = \cup_{i=1}^r \{(\mathbf{b}_1, \mathbf{b}_2) : \mathbf{b}_1, \mathbf{b}_2 \in \mathcal{S}_i\}$. This establishes (12). The remainder of the proof differs from the one provided for Lemma 2, as we use combinatorial arguments [2].

Denote $f = n - e$ and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_{|\mathcal{D}|}\} = \{\mathbf{b}_1 - \mathbf{b}_2 : (\mathbf{b}_1, \mathbf{b}_2) \in \mathcal{D}\}$ be the set of differences of the pairs of suffixes in $\mathcal{D}$, with multiplicities. Then, the multiplicity of 0^f in $\{\mathbf{v}_1, \ldots, \mathbf{v}_{|\mathcal{D}|}\}$ is $B_n^{\omega'}$ and the multiplicity of each nonzero vector in $\{\mathbf{v}_1, \ldots, \mathbf{v}_{|\mathcal{D}|}\}$ is one. For $i \in [|\mathcal{D}|]$, $j \in [1, f]$, let $h_{ij} = 1$, if the the j-th bit of $\mathbf{v}_i$ is 0, and $h_{ij} = -1$ otherwise. Then, $\sum_{i=1}^{|\mathcal{D}|} h_{ij} \geq 0$ for $j \in [1, f]$ and $\sum_{i=1}^{|\mathcal{D}|}\sum_{j=1}^{f} h_{ij} \geq 0$.

Since $\sum_{j=1}^{f} h_{ij} = f - 2k$ for each $\mathbf{v}_i$ with k non-zero entries, we have $\sum_{j=1}^{f} h_{ij} \leq 0$ for any $\mathbf{v}_i$ having at least $\frac{f}{2}$ non-zero entries. In addition, the number of possible difference vectors with k non-zero entries is at most $\binom{f}{k}2^k$. Let $|\mathbf{v}_i|$ denote the number of nonzero entries in $\mathbf{v}_i$, $i \in [|\mathcal{D}|]$. Then,

$$\begin{aligned} 0 \leq \sum_{i=1}^{|\mathcal{D}|}\sum_{j=1}^{f} h_{ij} &\overset{(a)}{\leq} fB_n^{\omega'} + \sum_{i:|\mathbf{v}_i|\geq 1}(\sum_{j=1}^{f} h_{ij}) \\ &\overset{(b)}{\leq} fB_n^{\omega'} + \sum_{i:1\leq|\mathbf{v}_i|\leq\min\{\frac{f}{2}, 2(\omega-\omega')\}}(\sum_{j=1}^{f} h_{ij}) \\ &- \sum_{i:|\mathbf{v}_i|>\frac{f}{2}} 1 \overset{(c)}{\leq} fB_n^{\omega'} + 2f\omega''\binom{f}{2\omega''}2^{2\omega''} - \sum_{i:|\mathbf{v}_i|>\frac{f}{2}} 1, \end{aligned} \tag{29}$$

where $\omega'' = \min\{\omega - \omega', \frac{f}{4}\}$, (a) follows from the fact that the multiplicity of 0^f in $\{\mathbf{v}_1, \ldots, \mathbf{v}_s\}$ is $B_n^{\omega'}$, (b) follows from the facts that $\sum_{j=1}^{f} h_{ij} < 0$ when $|\mathbf{v}_i| > \frac{f}{2}$ and that $|\mathbf{v}_i| \leq 2(\omega - \omega')$, and (c) follows from the facts that the number of $\mathbf{v}_i$ with $|\mathbf{v}_i| = k$ is at most $\binom{f}{k}2^k$ and that this number increases with k when $k \leq \min\{\frac{f}{2}, 2(\omega-\omega')\}$. Eq. (29) implies that

$$\begin{aligned} |\mathcal{D}| &\leq (f+1)|\mathcal{B}_n^{\omega'}| + 2(f+1)\omega''\binom{f}{2\omega''}2^{2\omega''} \\ &\leq (f+1)|\mathcal{B}_n^{\omega'}| + 2^{fH(\frac{2\omega''}{f})+2\omega''+\log(n(n+1))} \end{aligned} \tag{30}$$

Combining (12) and (30), we have

$$|\mathcal{B}_n^{\omega'}|^2 \leq (f+1)|\mathcal{B}_n^{\omega'}|r + 2^{fH(\frac{2\omega''}{f})+2\omega''+\log(n(n+1))}r.$$

Since $r \leq 2^{eH(\frac{\omega'}{e})}$, we then have

$$\begin{aligned} \log|\mathcal{B}_n^{\omega'}| \leq \max\Bigg\{ & H(\frac{\omega'}{e})\cdot e + \log(n+1) + 1, \\ & \frac{1}{2}\Big[H(\frac{\omega'}{e})\cdot e + \log(n+1) + H(\frac{2\omega''}{f})\cdot f \\ & + 2\omega'' + \log(n(n+1))\Big] + 1\Bigg\}, \end{aligned} \tag{31}$$

which implies (5) together with (8).

REFERENCES

[1] B. Lindström, "Determination of two vectors from the sum," *Journal of Combinatorial Theory*, vol. 6, no. 4, pp. 402–407, 1969.

[2] ——, "On B2-sequences of vectors," *Journal of number Theory*, vol. 4, no. 3, pp. 261–265, 1972.

[3] G. Cohen, S. Litsyn, and G. Zémor, "Binary B2-sequences: a new upper bound," *Journal of Combinatorial Theory, Series A*, vol. 94, no. 1, pp. 152–155, 2001.

[4] S. Della Fiore and M. Dalai, "A note on $\overline{2}$ separable codes and B2 codes," *Discrete Mathematics*, vol. 345, no. 3, p. 112751, 2022.

[5] A. G. D'yachkov, "Lectures on designing screening experiments," *arXiv preprint arXiv:1401.7505*, 2014.

[6] M. Bouvel, V. Grebinski, and G. Kucherov, "Combinatorial search on graphs motivated by bioinformatics applications: A brief survey," in *International Workshop on Graph-Theoretic Concepts in Computer Science*. Springer, 2005, pp. 16–27.

[7] V. Gritsenko, G. Kabatiansky, V. Lebedev, and A. Maevskiy, "Signature codes for noisy multiple access adder channel," *Designs, Codes and Cryptography*, vol. 82, no. 1, pp. 293–299, 2017.

[8] Y. Polyanskiy, "Information theoretic perspective on massive multiple-access," *Short Course (slides) Skoltech Inst. of Tech., Moscow, Russia*, 2018.

[9] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Transactions on Information Theory*, vol. 44, no. 5, pp. 1897–1905, 1998.

[10] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, and E. Birney, "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA," *Nature*, vol. 494, no. 7435, p. 77, 2013.

[11] G. M. Church, Y. Gao, and S. Kosuri, "Next-generation digital information storage in DNA," *Science*, vol. 337, no. 6102, pp. 1628–1628, 2012. [Online]. Available: http://science.sciencemag.org/content/337/6102/1628

[12] R. N. Grass, R. Heckel, M. Puddu, D. Paunescu, and W. J. Stark, "Robust chemical preservation of digital information on DNA in silica with error-correcting codes," *Angewandte Chemie International Edition*, vol. 54, no. 8, pp. 2552–2555, 2015.

[13] S. H. T. Yazdi, Y. Yuan, J. Ma, H. Zhao, and O. Milenkovic, "A rewritable, random-access DNA-based storage system," *Scientific reports*, vol. 5, no. 1, pp. 1–10, 2015.

[14] S. H. T. Yazdi, R. Gabrys, and O. Milenkovic, "Portable and error-free dna-based data storage," *Scientific reports*, vol. 7, no. 1, pp. 1–6, 2017.

[15] S. K. Tabatabaei, B. Wang, N. B. M. Athreya, B. Enghiad, A. G. Hernandez, C. J. Fields, J.-P. Leburton, D. Soloveichik, H. Zhao, and O. Milenkovic, "DNA punch cards for storing data on native DNA sequences via enzymatic nicking," *Nature communications*, vol. 11, no. 1, pp. 1–10, 2020.

[16] C. Pan, K. Tabatabaei, S. H. T. Yazdi, A. G. Hernandez, C. M. Schroeder, and O. Milenkovic, "Rewritable two-dimensional DNA-based data storage with machine learning reconstruction," *bioRxiv*, 2021.

[17] S. K. Tabatabaei, B. Pham, C. Pan, J. Liu, S. Chandak, S. A. Shorkey, A. G. Hernandez, A. Aksimentiev, M. Chen, C. M. Schroeder *et al.*, "Expanding the molecular alphabet of dna-based data storage systems with neural network nanopore readout processing," *Nano letters*, vol. 22, no. 5, pp. 1905–1914, 2022.

[18] C. Laure, D. Karamessini, O. Milenkovic, L. Charles, and J.-F. Lutz, "Coding in 2D: Using intentional dispersity to enhance the information capacity of sequence-coded polymer barcodes," *Angewandte Chemie*, vol. 128, no. 36, pp. 10 880–10 883, 2016. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ange.201605279

[19] J. Acharya, H. Das, O. Milenkovic, A. Orlitsky, and S. Pan, "String reconstruction from substring compositions," *SIAM Journal on Discrete Mathematics*, vol. 29, no. 3, pp. 1340–1371, 2015.

[20] S. Pattabiraman, R. Gabrys, and O. Milenkovic, "Reconstruction and error-correction codes for polymer-based data storage," in *2019 IEEE Information Theory Workshop (ITW)*. IEEE, 2019, pp. 1–5.

[21] R. Gabrys, S. Pattabiraman, and O. Milenkovic, "Reconstructing mixtures of coded strings from prefix and suffix compositions," in *2020 IEEE Information Theory Workshop (ITW)*. IEEE, 2021, pp. 1–5.

[22] ——, "Reconstruction of sets of strings from prefix/suffix compositions," *IEEE Transactions on Communications*, 2022.

[23] R. Graham and N. Sloane, "Lower bounds for constant weight codes," *IEEE Transactions on Information Theory*, vol. 26, no. 1, pp. 37–43, 1980.

[24] E. Agrell, A. Vardy, and K. Zeger, "Upper bounds for constant-weight codes," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2373–2395, 2000.

[25] S. Johnson, "Upper bounds for constant weight error correcting codes," *Discrete Mathematics*, vol. 3, no. 1-3, pp. 109–124, 1972.

[26] A. E. Brouwer, J. B. Shearer, N. J. Sloane, and W. D. Smith, "A new table of constant weight codes," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1334–1380, 2006.

[27] T. Etzion and A. Vardy, "A new construction for constant weight codes," in *2014 International Symposium on Information Theory and its Applications*. IEEE, 2014, pp. 338–342.

[28] Y. M. Chee and S. Ling, "Constructions for q-ary constant-weight codes," *IEEE Transactions on Information Theory*, vol. 53, no. 1, pp. 135–146, 2006.

[29] F.-W. Fu, A. H. Vinck, and S.-Y. Shen, "On the constructions of constant-weight codes," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 328–333, 1998.

[30] I. Z. Ruzsa, "An infinite sidon sequence," *Journal of Number Theory*, vol. 68, no. 1, pp. 63–71, 1998.