Bounds on Reversible, Complement, Reversible-Complement, Constant Weight Sum Codes

Krishna Gopal Benerjee, and Adrish Banerjee
Department of Electrical Engineering,
Indian Institute of Technology Kanpur, India
Email: {kgopal, adrish}@iitk.ac.in

***Abstract*—For an even-size alphabet, a code is called Reversible, Complement, Reversible-Complement, Constant Weight Sum code (RCCWS code) if the code is closed under reversible, complement, and reversible-complement sequences, and for some given symbols in the alphabet, the total number of occurrence of those symbols in each sequence is identical. In this work, we have obtained an upper bound (Sphere Packing bound) and a lower bound (Gilbert-Varshamov bound) on the maximum size of an RCCWS code defined over the even-size alphabet. As a special case of these two bounds, we also have obtained an upper and lower bound on the maximum size of DNA codes that satisfies reversible, complement, reversible-complement, and GC-content constraints.**

I. Introduction

In coding theory, bounds on parameters are important to study the properties of codes. For any code with a given length and minimum Hamming distance, the Sphere Packing bound is an upper bound on the size of the code, and Gilbert-Varshamov bound is a lower bound on the maximum size of the code [1].

For any positive integers m, n, M and d_H, an (n, M, d_H) code $\mathscr{C}$ over an m-size alphabet is a set of M codewords each of length n such that the minimum Hamming distance $d_H = \min\{d_H(\mathbf{a}, \mathbf{b}) : \mathbf{a} \neq \mathbf{b} \text{ and } \mathbf{a}, \mathbf{b} \in \mathscr{C}\}$, where the Hamming distance between sequences $\mathbf{a}$ and $\mathbf{b}$ is $d_H(\mathbf{a}, \mathbf{b})$. For any code defined over an m-size alphabet, let $A_m(n, d_H)$ be the maximum number of codewords in a code of length n and the minimum distance d_H, then

- Sphere Packing bound:

$$A_m(n, d_H) \leq \frac{m^n}{\sum_{r=0}^{\lfloor (d_H-1)/2 \rfloor} \binom{n}{r}(m-1)^r} \tag{1}$$

- Gilbert-Varshamov bound:

$$A_m(n, d_H) \geq \frac{m^n}{\sum_{r=0}^{d_H-1} \binom{n}{r}(m-1)^r} \tag{2}$$

Further, the lower [2], and upper [3] bounds are studied for constant weight codes. Also, using bounds for spherical codes, upper bounds on the size of constant weight binary codes are studied in [4]. Asymptotic upper bounds are studied for generalized weights of a binary linear code of a given size [5]. Upper bounds for the small size binary codes are discussed in [6].

For different applications of DNA computing, such as DNA data storage, DNA codes are designed with some properties, such as reversible constraint (DNA sequences and their reversible sequences are sufficiently different), reversible-complement constraint (DNA sequences and their reversible-complement sequences are sufficiently different), and GC-content constraint (all the DNA sequences have the same total sum of Gs' and Cs') [7]–[9]. DNA codes with reversible and reversible-complement constraints are constructed in [10]–[12]. DNA codes with reversible-complement constraints are studied in [13] along with two additional properties that include avoiding homopolymers and secondary structures. A homopolymer of run-length ℓ in a DNA sequence is the repeating of symbols at ℓ-consecutive positions in the sequence. Also, the secondary structure is the structure formed in a DNA sequence by the folding of two polynucleotide chains around one another DNA codes with reversible, reversible-complement, and GC-content constraints are studied in [14], where the DNA codes avoid homopolymers and secondary structures. Also, a lower bound on DNA codes' size is established in [14]. Bounds on GC-balanced DNA codes are studied in [15], [16]. Bounds on DNA codes with reversible and reversible-complement constraints are discussed in [17]. Bounds on DNA codes with GC-content constraint and free from homopolymers are studied in [18].

The main contribution of this work is as follows:

- We have obtained Sphere Packing (SP) bound and Gilbert-Varshamov (GV) bound for reversible, complement, reversible-complement, and constant weight sum codes (RCCWS codes) over an even-size alphabet. To the best of our knowledge, bounds for reversible, complement, reversible-complement, and constant weight sum codes are not studied in the literature.
- As a special case, we have also obtained upper and lower bounds on the maximum size of GC-balanced DNA codes with reversible, complement, and reversible-complement constraints. As per the authors' knowledge, bounds for GC-balanced DNA codes that satisfy reversible, complement, and reversible-complement constraints are not studied in the literature.

Organization: The preliminaries along with notations are de-

fined in Section II. Properties of RCCWS codes are discussed in Section III. Also, SP and GV bounds for RCCWS codes are established in the same section. In particular, using those bounds for RCCWS codes, lower and upper bounds are derived for the GC-balanced DNA codes with reversible, complement, and reversible-complement constraints in Section IV.

II. Preliminaries

In this section, we have described the notations and definitions that are used in the work.

A. RCCWS Codes and Their Properties

For positive integers p and q, consider an alphabet

$$\mathcal{A}_{2(p+q)} = \{\alpha_1, \alpha_2, \ldots, \alpha_{2p}, \beta_1, \beta_2, \ldots, \beta_{2q}\}$$

of size $2(p+q)$ such that

1) $\alpha_i^C = \alpha_{2p-i+1}$ for $i = 1, 2, \ldots, 2p$, and
2) $\beta_j^C = \beta_{2q-j+1}$ for $j = 1, 2, \ldots, 2q$,

where the symbol x^C is the complement of the symbol $x \in \mathcal{A}_{2(p+q)}$. For any $a \in \mathcal{A}_{2(p+q)}$, $(a^C)^C = a$, and $a^C \neq a$. Again, for any a and b in the alphabet $\mathcal{A}_{2(p+q)}$, $a = b$ iff $a^C = b^C$.

***Definition** 1:* For any positive integer n, consider $\mathbf{z} = z_1 z_2 \ldots z_n$, where $z_i \in \mathcal{A}_{2(p+q)}$ for $i = 1, 2, \ldots, 2(p+q)$. Then,

- the reversible sequence is $\mathbf{z}^R = z_n z_{n-1} \ldots z_1$,
- the complement sequence is $\mathbf{z}^C = z_1^C z_2^C \ldots z_n^C$,
- the reversible-complement sequence is $\mathbf{z}^{RC} = z_n^C z_{n-1}^C \ldots z_1^C$, and
- the weight sum, $w(\mathbf{z})$, of the sequence $\mathbf{z}$ is the number of positions of α_i in $\mathbf{z}$ for $i = 1, 2, \ldots, 2p$, i.e., $w(\mathbf{z})$ is the size of the set $\{j : z_j \in \{\alpha_1, \alpha_2, \ldots, \alpha_{2p}\}\}$.

***Example** 1:* For $p = 2$ and $q = 1$, the alphabet is $\mathcal{A}_6 = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2\}$, where $\alpha_i^C = \alpha_{5-i}$ for $i = 1, 2, 3, 4$ and $\beta_j^C = \beta_3 - j$ for $j = 1, 2$. Now, for $n = 5$, consider $\mathbf{z} = \alpha_3\beta_1\alpha_1\alpha_1\alpha_2$. Then,

- the reversible sequence is $\mathbf{z}^R = \alpha_2\alpha_1\alpha_1\beta_1\alpha_3$,
- the complement sequence is $\mathbf{z}^C = \alpha_3^C\beta_1^C\alpha_1^C\alpha_1^C\alpha_2^C = \alpha_2\beta_2\alpha_4\alpha_4\alpha_3$,
- the reversible-complement sequence is $\mathbf{z}^{RC} = \alpha_2^C\alpha_1^C\alpha_1^C\beta_1^C\alpha_3^C = \alpha_3\alpha_4\alpha_4\beta_2\alpha_2$, and
- the weight sum of the sequence $\mathbf{z}$ is $w(\mathbf{z}) = 4$.

For any positive integers p, q, n, M and d_H, an (n, M, d_H) code $\mathscr{C}$ is a subset of $\mathcal{A}_{2(p+q)}^n$ of size M such that the minimum Hamming distance $d_H = \min\{d_H(\mathbf{a}, \mathbf{b}) : \mathbf{a} \neq \mathbf{b} \text{ and } \mathbf{a}, \mathbf{b} \in \mathscr{C}\}$, where the Hamming distance between sequences $\mathbf{a}$ and $\mathbf{b}$ is $d_H(\mathbf{a}, \mathbf{b})$.

***Definition** 2:* For any given positive integer w, an (n, M, d_H) code $\mathscr{C}$ is called (n, M, d_H, w) reversible, complement, reversible-complement, and constant weight sum code (RCCWS codes) if, for each $\mathbf{z} \in \mathscr{C}$,

- Reversible Property: $\mathbf{z}^R \in \mathscr{C}$, i.e., $\mathscr{C}$ is closed under reversible sequences,
- Complement Property: $\mathbf{z}^C \in \mathscr{C}$, i.e., $\mathscr{C}$ is closed under complement sequences,
- Reversible-Complement Property: $\mathbf{z}^{RC} \in \mathscr{C}$, i.e., $\mathscr{C}$ is closed under reversible-complement sequences, and
- Weight Sum Property: $w(\mathbf{z}) = w$, i.e., all the sequences in $\mathscr{C}$ have the same weight sum w.

***Example** 2:* For $p = 2$ and $q = 1$, consider an $(n = 4, M = 6, d_H = 3)$ code $\mathscr{C}$ defined over the alphabet $\mathcal{A}_6$, where $\mathscr{C} = \{\beta_1\beta_1\alpha_4\alpha_4, \alpha_4\alpha_4\beta_1\beta_1, \beta_2\beta_2\alpha_1\alpha_1, \alpha_1\alpha_1\beta_2\beta_2, \beta_2\alpha_2\alpha_3\beta_1, \beta_1\alpha_3\alpha_2\beta_2\}$. The code $\mathscr{C}$ is an $(n = 4, M = 6, d_H = 3, w = 2)$ reversible, complement, reversible-complement and constant weight sum code.

For any reversible, complement and constant weight sum code $\mathscr{C}$, if $\mathbf{z} \in \mathscr{C}$ then the reversible sequence $\mathbf{z}^R$, and complement sequence $\mathbf{z}^C$ are in $\mathscr{C}$, and therefore, the sequence $(\mathbf{z}^R)^C$ is in $\mathscr{C}$. But, $(\mathbf{z}^R)^C = (\mathbf{z}^{RC})$, and thus, the code $\mathscr{C}$ is closed under reversible-complement sequences, *i.e.*, $\mathbf{z}^{RC} \in \mathscr{C}$. Hence, the code $\mathscr{C}$ is an RCCWS code. Similarly, a complement, reversible-complement, and constant weight sum code is an RCCWS code, and also, a reversible, reversible-complement, and constant weight sum code is an RCCWS code. Note that the code given in Example 2 is a RCCWS code.

For any non-negative integers n, r and w ($0 \leq w \leq n$), consider a set, $\mathcal{B}_{2(p+q),w,n}$, of all sequences with the length n and the weight sum w over $\mathcal{A}_{2(p+q)}$. Then, the set

$$H_r(\mathbf{a}) = \{\mathbf{z} : d_H(\mathbf{a}, \mathbf{z}) = r \text{ for } \mathbf{z} \in \mathcal{B}_{2(p+q),w,n}\}$$

is called a Hamming sphere with the radius r, the center $\mathbf{a}$, and weight sum w. Also, we denote the size of the Hamming sphere $H_r(\mathbf{a})$ by $V_H(\mathbf{a}, r)$.

B. DNA Codes and their Properties

In this sub-section, we have listed definitions and properties of DNA sequences and DNA codes used in this work.

Any sequence defined on $\{A, C, G, T\}$ is known as a DNA sequence. For a DNA sequence $\mathbf{z} = z_1 z_2 \ldots z_n$ of length n,

- the reversible DNA sequence is $\mathbf{z}^R = z_n z_{n-1} \ldots z_1$,
- the complement DNA sequence is $\mathbf{z}^C = z_1^C z_2^C \ldots z_n^C$, and
- the reversible-complement DNA sequence is $\mathbf{z}^{RC} = z_n^C z_{n-1}^C \ldots z_1^C$,

where $A^C = T$, $T^C = A$, $C^C = G$ and $G^C = C$. The GC-weight of a DNA sequence is the total number of occurrences of Cs' and Gs' in the sequence. Any (n, M, d_H) code defined over the alphabet $\{A, C, G, T\}$ is called a DNA code. An (n, M, d_H) DNA code $\mathscr{C}_{DNA}$ satisfies

- Reversible Constraint: if $d_H(\mathbf{a}, \mathbf{b}^R) \geq d_H$ for any $\mathbf{a}, \mathbf{b} \in \mathscr{C}_{DNA}$ and $\mathbf{a} \neq \mathbf{b}^R$,
- Complement Constraint: if $d_H(\mathbf{a}, \mathbf{b}^C) \geq d_H$ for any $\mathbf{a}, \mathbf{b} \in \mathscr{C}_{DNA}$ and $\mathbf{a} \neq \mathbf{b}^C$, and
- Reversible-Complement Constraint: if $d_H(\mathbf{a}, \mathbf{b}^{RC}) \geq d_H$ for any $\mathbf{a}, \mathbf{b} \in \mathscr{C}_{DNA}$ and $\mathbf{a} \neq \mathbf{b}^{RC}$.

For any DNA codeword $\mathbf{x}$ in a DNA code $\mathscr{C}_{DNA}$,

- if $\mathbf{x}^R \in \mathscr{C}_{DNA}$ then the DNA code $\mathscr{C}_{DNA}$ satisfies reversible constraint,
- if $\mathbf{x}^C \in \mathscr{C}_{DNA}$ then the DNA code $\mathscr{C}_{DNA}$ satisfies complement constraint, and

- if $\mathbf{x}^{RC} \in \mathscr{C}_{DNA}$ then the DNA code $\mathscr{C}_{DNA}$ satisfies reversible-complement constraint.

For any (n, M, d_H) DNA code, if all the DNA sequences have the same GC-weight, then the DNA code is called GC-balanced DNA code. If the GC-weight is w then we represent the GC-balanced DNA code with the parameter (n, M, d_H, w). In particular, a GC-balanced DNA code is called a DNA code with GC-content constraint if the GC-weight of each DNA sequence is $\lfloor n/2 \rfloor$.

III. Bounds and Properties of RCCWS Codes

In this section, we have discussed the properties of RCCWS codes, and obtained bounds on the size of RCCWS codes.

A. Properties of Sequences

In this sub-section, we discussed the properties of sequences and RCCWS codes.

From the Definition 1, one can obtain Proposition 1, Proposition 2, Lemma 1, and Lemma 2 as follows:

***Proposition** 1:* For any sequence $\mathbf{a}$ of length n over the alphabet $\mathcal{A}_{2(p+q)}$, $\mathbf{a} \neq \mathbf{a}^C$.

***Proposition** 2:* For any sequences $\mathbf{a}$ and $\mathbf{b}$ each of length n over the alphabet $\mathcal{A}_{2(p+q)}$,

- $\mathbf{a} = \mathbf{b}^R$ if and only if $\mathbf{b} = \mathbf{a}^R$,
- $\mathbf{a} = \mathbf{b}^C$ if and only if $\mathbf{b} = \mathbf{a}^C$, and
- $\mathbf{a} = \mathbf{b}^{RC}$ if and only if $\mathbf{b} = \mathbf{a}^{RC}$.

***Lemma** 1:* For any sequence $\mathbf{a}$ of length n over the alphabet $\mathcal{A}_{2(p+q)}$, $d_H(\mathbf{a}, \mathbf{a}^C) = d_H(\mathbf{a}^R, \mathbf{a}^{RC}) = n$.

Proof: For any sequence $\mathbf{a} = a_1 a_2 \dots a_n$ over the alphabet $\mathcal{A}_{2(p+q)}$, recall from Definition 1, $\mathbf{a}^C = a_1^C a_2^C \dots a_n^C$, $\mathbf{a}^R = a_n a_{n-1} \dots a_1$, $\mathbf{a}^{RC} = a_n^C a_{n-1}^C \dots a_1^C$, and, for $j = 1, 2, \dots, n$, $d_H(a_j, a_j^C) = 1$. Then, from Proposition 1, for the Hamming case,

$$d_H(\mathbf{a}, \mathbf{a}^C) = \sum_{i=1}^{n} d_H(a_i, a_i^C) = n,$$

and

$$\begin{aligned} d_H(\mathbf{a}^R, \mathbf{a}^{RC}) &= \sum_{i=1}^{n} d_H(a_{n-i+1}, a_{n-i+1}^C) \\ &= \sum_{i=1}^{n} d_H(a_i, a_i^C) = d_H(\mathbf{a}, \mathbf{a}^C). \end{aligned}$$

Thus, the result follows. ∎

***Lemma** 2:* Consider any two sequences $\mathbf{a}$ and $\mathbf{b}$ each of length n over the alphabet $\mathcal{A}_{2(p+q)}$. Then,
$d_H(\mathbf{a}, \mathbf{b}) = d_H(\mathbf{a}^C, \mathbf{b}^C) = d_H(\mathbf{a}^R, \mathbf{b}^R) = d_H(\mathbf{a}^{RC}, \mathbf{b}^{RC})$,
$d_H(\mathbf{a}^R, \mathbf{b}) = d_H(\mathbf{a}^{RC}, \mathbf{b}^C) = d_H(\mathbf{a}, \mathbf{b}^R) = d_H(\mathbf{a}^C, \mathbf{b}^{RC})$,
$d_H(\mathbf{a}^C, \mathbf{b}) = d_H(\mathbf{a}, \mathbf{b}^C) = d_H(\mathbf{a}^{RC}, \mathbf{b}^R) = d_H(\mathbf{a}^R, \mathbf{b}^{RC})$, and
$d_H(\mathbf{a}^{RC}, \mathbf{b}) = d_H(\mathbf{a}^R, \mathbf{b}^C) = d_H(\mathbf{a}^C, \mathbf{b}^R) = d_H(\mathbf{a}, \mathbf{b}^{RC})$.

Proof: From Definition 1, for any two sequence $\mathbf{a} = a_1 a_2 \dots a_n$ and $\mathbf{b} = b_1 b_2 \dots b_n$ over the alphabet $\mathcal{A}_{2(p+q)}$, we have $\mathbf{a}^C = a_1^C a_2^2 \dots a_n^C$, $\mathbf{b}^C = b_1^C b_2^2 \dots b_n^C$, $\mathbf{a}^R = a_n a_{n-1} \dots a_1$, $\mathbf{b}^R = b_n b_{n-1} \dots b_1$, $\mathbf{a}^{RC} = a_n^C a_{n-1}^C \dots a_1^C$, and $\mathbf{b}^{RC} = b_n^C b_{n-1}^C \dots b_1^C$. From the complement property, one can observe that $d_H(a_j, b_j) = d_H(a_j^C, b_j^C)$ for $j = 1, 2, \dots, n$, and therefore, the Hamming distance $d_H(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{n} d_H(a_i, b_i) = \sum_{i=1}^{n} d_H(a_i^C, b_i^C) = d_H(\mathbf{a}^C, \mathbf{b}^C)$. Thus, $d_H(\mathbf{a}, \mathbf{b}) = d_H(\mathbf{a}^C, \mathbf{b}^C)$. Now, from the property of reversible sequences, the Hamming distance $d_H(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{n} d_H(a_i, b_i) = \sum_{j=1}^{n} d_H(a_{n-j+1}, b_{n-j+1}) = d_H(\mathbf{a}^R, \mathbf{b}^R)$. Thus, $d_H(\mathbf{a}, \mathbf{b}) = d_H(\mathbf{a}^R, \mathbf{b}^R)$. Again, from the property of reversible-complement sequences, the Hamming distance $d_H(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{n} d_H(a_i, b_i) = \sum_{j=1}^{n} d_H(a_{n-j+1}, b_{n-j+1}) = \sum_{j=1}^{n} d_H(a_{n-j+1}^C, b_{n-j+1}^C) = d_H(\mathbf{a}^{RC}, \mathbf{b}^{RC})$. Thus, $d_H(\mathbf{a}, \mathbf{b}) = d_H(\mathbf{a}^{RC}, \mathbf{b}^{RC})$. Therefore, $d_H(\mathbf{a}, \mathbf{b}) = d_H(\mathbf{a}^C, \mathbf{b}^C) = d_H(\mathbf{a}^R, \mathbf{b}^R) = d_H(\mathbf{a}^{RC}, \mathbf{b}^{RC})$. Now, from the result $d_H(\mathbf{a}, \mathbf{b}) = d_H(\mathbf{a}^C, \mathbf{b}^C)$ and the property $(\mathbf{a}^R)^C = \mathbf{a}^{RC}$, one can find that $d_H(\mathbf{a}^R, \mathbf{b}) = d_H((\mathbf{a}^R)^C, \mathbf{b}^C) = d_H(\mathbf{a}^{RC}, \mathbf{b}^C)$. Again, from the result $d_H(\mathbf{a}, \mathbf{b}) = d_H(\mathbf{a}^R, \mathbf{b}^R)$ and the property $(\mathbf{a}^R)^R = \mathbf{a}$, one can find that $d_H(\mathbf{a}^R, \mathbf{b}) = d_H((\mathbf{a}^R)^R, \mathbf{b}^R) = d_H(\mathbf{a}, \mathbf{b}^R)$. Further, from the result $d_H(\mathbf{a}, \mathbf{b}) = d_H(\mathbf{a}^{RC}, \mathbf{b}^{RC})$ and the property $(\mathbf{a}^R)^{RC} = \mathbf{a}^C$, one can find that $d_H(\mathbf{a}^R, \mathbf{b}) = d_H((\mathbf{a}^R)^{RC}, \mathbf{b}^{RC}) = d_H(\mathbf{a}^C, \mathbf{b}^{RC})$. Therefore, $d_H(\mathbf{a}^R, \mathbf{b}) = d_H(\mathbf{a}^{RC}, \mathbf{b}^C) = d_H(\mathbf{a}, \mathbf{b}^R) = d_H(\mathbf{a}^C, \mathbf{b}^{RC})$. Similarly, from the result $d_H(\mathbf{a}, \mathbf{b}) = d_H(\mathbf{a}^C, \mathbf{b}^C) = d_H(\mathbf{a}^R, \mathbf{b}^R) = d_H(\mathbf{a}^{RC}, \mathbf{b}^{RC})$, the result $d_H(\mathbf{a}^C, \mathbf{b}) = d_H(\mathbf{a}, \mathbf{b}^C) = d_H(\mathbf{a}^{RC}, \mathbf{b}^R) = d_H(\mathbf{a}^R, \mathbf{b}^{RC})$ follows from the properties $(\mathbf{a}^C)^C = \mathbf{a}$, $(\mathbf{a}^C)^R = \mathbf{a}^{RC}$, $(\mathbf{a}^C)^{RC} = \mathbf{a}^R$, and the result $d_H(\mathbf{a}^{RC}, \mathbf{b}) = d_H(\mathbf{a}^R, \mathbf{b}^C) = d_H(\mathbf{a}^C, \mathbf{b}^R) = d_H(\mathbf{a}, \mathbf{b}^{RC})$ follows from the properties $(\mathbf{a}^{RC})^C = \mathbf{a}^R$, $(\mathbf{a}^{RC})^R = \mathbf{a}^C$, $(\mathbf{a}^{RC})^{RC} = \mathbf{a}$. ∎

Relation between weight sums of sequences and their reversible, complement, and reversible-complement sequences is established in Lemma 3.

***Lemma** 3:* For any sequence $\mathbf{z}$ of length n over the alphabet $\mathcal{A}_{2(p+q)}$, the weight sum $w(\mathbf{z}) = w(\mathbf{z}^R) = w(\mathbf{z}^C) = w(\mathbf{z}^{RC})$.

Proof: For any $\mathbf{z} = z_1 z_2 \dots z_n$ over $\mathcal{A}_{2(p+q)} = \{\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_{2p}, \beta_1, \beta_2, \dots, \beta_{2q}\}$, $z_i \in \{\alpha_1, \alpha_2, \dots, \alpha_{2p}\}$ if and only if $z_i^C \in \{\alpha_1, \alpha_2, \dots, \alpha_{2p}\}$ for $i = 1, 2, \dots, n$, and thus, $w(\mathbf{z}) = w(\mathbf{z}^C)$ and $w(\mathbf{z}^R) = w(\mathbf{z}^{RC})$. Again, consider a sequence $\mathbf{a} = a_1 a_2 \dots a_n$ over $\mathcal{A}_{2(p+q)}$ such that $\mathbf{a} = \mathbf{z}^R$, and thus, $a_i = z_{n-i+1}$ for $i = 1, 2, \dots, n$. Now, consider $S \subseteq \{1, 2, \dots, n\}$ of size $w(\mathbf{z})$ such that $j \in S$ if and only if $z_j \in \{\alpha_1, \alpha_2, \dots, \alpha_{2p}\}$. So, $j \in S$ if and only if $a_{n-j+1} \in \{\alpha_1, \alpha_2, \dots, \alpha_{2p}\}$, and therefore, there are only $w(\mathbf{z})$ symbols in the sequence $\mathbf{a}$ (= $\mathbf{z}^R$) such that the symbols are from the set $\{\alpha_1, \alpha_2, \dots, \alpha_{2p}\}$. Thus, $w(\mathbf{z}) = w(\mathbf{z}^R)$ and $w(\mathbf{z}^C) = w(\mathbf{z}^{RC})$. So, the result follows from these cases. ∎

B. Enumerating Sequences and their Properties

In this section, we have enumerated sequences with a given length and weight sum, and discussed the properties of the sequences.

For a given length and weight sum, we have enumerated the total number of sequences in Lemma 4.

***Lemma** 4:* For any given integers n and w ($0 \leq w \leq n$), the number of distinct sequences each of length n and weight sum w over the alphabet $\mathcal{A}_{2(p+q)}$ is $\binom{n}{w} 2^n p^w q^{n-w}$, i.e., the size of the set $\mathcal{B}_{2(p+q),w,n}$ is $\binom{n}{w} 2^n p^w q^{n-w}$.

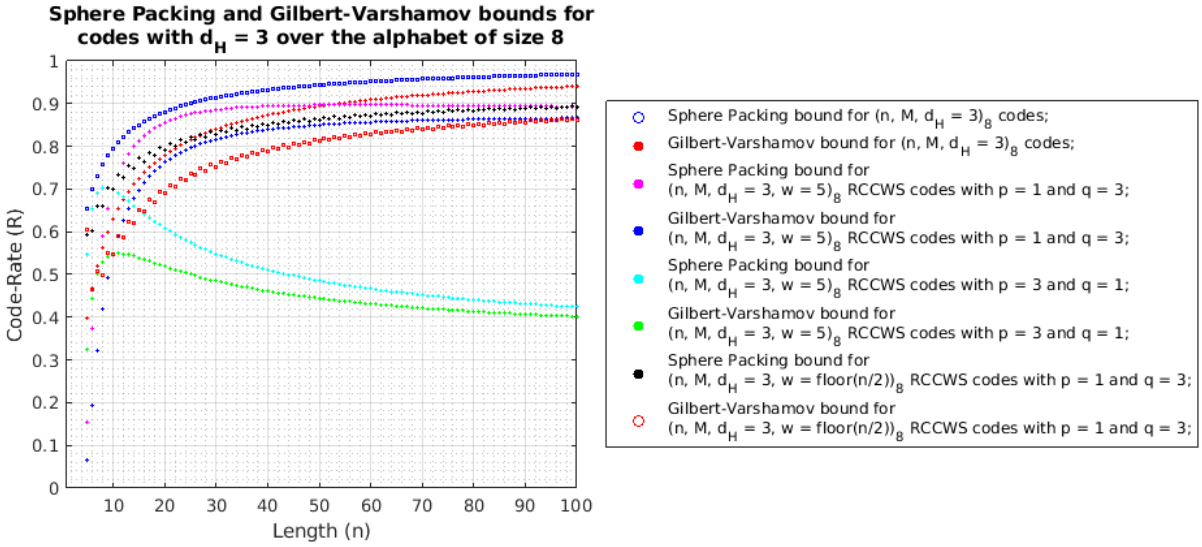

Figure 1. Bounds on codes over the alphabet $\mathcal{A}_8$.

Proof: For a sequence $\mathbf{z}$ of length n over the alphabet $\mathcal{A}_{2(p+q)}$, the weight sum $w(\mathbf{z})$ is w if and only if there are w positions with a symbol from $\{\alpha_1, \alpha_2, \ldots, \alpha_{2p}\}$ and $n-w$ positions with a symbol from $\{\beta_1, \beta_2, \ldots, \beta_{2q}\}$ in the sequence $\mathbf{z}$. Thus, there are $\binom{n}{w}(2p)^w(2q)^{n-w}$ $\left(=\binom{n}{w}2^n p^w q^{n-w}\right)$ distinct sequences each of length n and weight sum w. ■

Now, Lemma 4 can be extended as Lemma 5 for the set of sequences with reversible, complement, and reversible-complement properties.

***Lemma** 5:* For any integers n and w $(0 \leq w \leq n)$, $\mathbf{a} \in \mathcal{B}_{2(p+q),w,n}$ if and only if $\mathbf{a}^R, \mathbf{a}^C, \mathbf{a}^{RC} \in \mathcal{B}_{2(p+q),w,n}$.

Proof: The result follows from Lemma 3, and the fact that $\mathcal{B}_{2(p+q),w,n}$ is the set of all sequences each of length n and of weight sum w. ■

C. Properties of Hamming Spheres

In this section, we have obtained the properties of Hamming Spheres for such sequences.

Relations among Hamming sphere sizes for the reversible, complement, and reversible-complement sequences are established in Proposition 3.

***Proposition** 3:* For any given Hamming sphere $H_r(\mathbf{a}) = \{\mathbf{z} : d_H(\mathbf{a}, \mathbf{z}) = r \text{ for } \mathbf{z} \in \mathcal{B}_{2(p+q),w,n}\}$ of size $V_H(\mathbf{a}, r)$ defined over the alphabet $\mathcal{A}_{2(p+q)}$, for $\mathbf{z} \in \mathcal{B}_{2(p+q),w,n}$,

- the Hamming sphere $H_r(\mathbf{a}^C)$ is

$$\{\mathbf{z}^C : d_H(\mathbf{a}^C, \mathbf{z}^C) = r \text{ for } \mathbf{z} \in \mathcal{B}_{2(p+q),w,n}\}$$

  of size $V_H(\mathbf{a}^C, r) = V_H(\mathbf{a}, r)$,
- the Hamming sphere $H_r(\mathbf{a}^R)$ is

$$\{\mathbf{z}^R : d_H(\mathbf{a}^R, \mathbf{z}^R) = r \text{ for } \mathbf{z} \in \mathcal{B}_{2(p+q),w,n}\}$$

  of size $V_H(\mathbf{a}^R, r) = V_H(\mathbf{a}, r)$, and
- the Hamming sphere $H_r(\mathbf{a}^{RC})$ is

$$\{\mathbf{z}^{RC} : d_H(\mathbf{a}^{RC}, \mathbf{z}^{RC}) = r \text{ for } \mathbf{z} \in \mathcal{B}_{2(p+q),w,n}\}$$

  of size $V_H(\mathbf{a}^{RC}, r) = V_H(\mathbf{a}, r)$.

Proof: The proof follows from Definition 1, the fact $d_H(\mathbf{a}, \mathbf{b}) = d_H(\mathbf{a}^C, \mathbf{b}^C) = d_H(\mathbf{a}^R, \mathbf{b}^R) = d_H(\mathbf{a}^{RC}, \mathbf{b}^{RC})$ (as shown in Lemma 2), and Lemma 3. ■

From the triangle property of Hamming distance, one can obtain the Proposition 4, Proposition 5, and Proposition 6.

***Proposition** 4:* For any two distinct positive integers r_1 and r_2, consider two distinct sequences $\mathbf{a}$ and $\mathbf{b}$ each of length n and weight sum w over the alphabet $\mathcal{A}_{2(p+q)}$. If $d_H(\mathbf{a}, \mathbf{b}) > r_1 + r_2$ then both the Hamming spheres $H_{r_1}(\mathbf{a})$ and $H_{r_2}(\mathbf{b})$ are disjoint.

Proof: The proof follows from the triangular property for the Hamming distance. ■

***Proposition** 5:* For any two positive integers r_1 and r_2, consider a sequences $\mathbf{a}$ of length n and weight sum w over the alphabet $\mathcal{A}_{2(p+q)}$. If $r_1 \neq r_2$ then both the Hamming spheres $H_{r_1}(\mathbf{a})$ and $H_{r_2}(\mathbf{a})$ are disjoint.

Proof: The result can be proved using the method of contradiction. For that, if possible, assume that $H_{r_1}(\mathbf{a}) \cap H_{r_2}(\mathbf{a}) \neq \emptyset$. So, there must exist at-least one sequence $\mathbf{b} \in H_{r_1}(\mathbf{a}) \cap H_{r_2}(\mathbf{a})$ of length n over $\mathcal{A}_{2(p+q)}$. And therefore, $\mathbf{b} \in H_{r_1}(\mathbf{a})$ and also $\mathbf{b} \in H_{r_2}(\mathbf{a})$. Now, if $\mathbf{b} \in H_{r_1}(\mathbf{a})$ then $d_H(\mathbf{a}, \mathbf{b}) = r_1$, and if $\mathbf{b} \in H_{r_2}(\mathbf{a})$ then $d_H(\mathbf{a}, \mathbf{b}) = r_2$. But, $r_1 \neq r_2$ and hence, it contradicts the uniqueness property of the distance. Thus, the given statement is true. ■

***Proposition** 6:* For any (n, M, d_H, w) RCCWS code $\mathscr{C}$, if $\mathbf{c}_1, \mathbf{c}_2 \in \mathscr{C}$ such that $\mathbf{c}_1 \neq \mathbf{c}_2$ then

1) $H_{r_1}(\mathbf{c}_1) \cap H_{r_2}(\mathbf{c}_2) = \emptyset$,
2) $H_{r_1}(\mathbf{c}_1^R) \cap H_{r_2}(\mathbf{c}_2^R) = \emptyset$,
3) $H_{r_1}(\mathbf{c}_1^C) \cap H_{r_2}(\mathbf{c}_2^C) = \emptyset$, and
4) $H_{r_1}(\mathbf{c}_1^{RC}) \cap H_{r_2}(\mathbf{c}_2^{RC}) = \emptyset$,

where r_1 and r_2 are positive integers such that $0 \leq r_1 + r_2 \leq d_H - 1$.

Proof: The proof follows from Definition 1 and the triangular property for the Hamming distance. ■

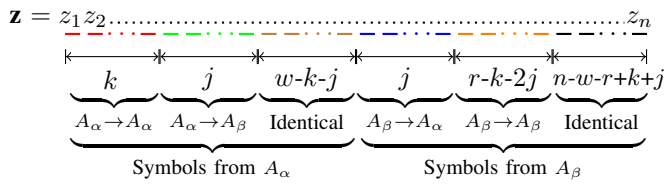

Figure 2. The number of positions is shown where the symbols on those positions in sequences $\mathbf{z}$ and $\mathbf{a}$ differ.

D. Hamming Sphere Size

In this section, we have obtained the size of Hamming Sphere for such sequences as follows.

***Lemma** 6:* For any sequence $\mathbf{z}$ of length n and weight sum w over the alphabet $\mathcal{A}_{2(p+q)}$, the size of the Hamming sphere $H_r(\mathbf{z})$ is $V_H(n, r, w)$ as given in Equation (3).

Proof: For any $\mathbf{z} = z_1 z_2 \dots z_n$ in $\mathcal{B}_{2(p+q),w,n}$, the w positions are filled by symbols from $A_\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_{2p}\}$, and the remaining $n - w$ positions are filled by symbols from $A_\beta = \{\beta_1, \beta_2, \dots, \beta_{2q}\}$ in $\mathbf{z}$. Also, in any sequence $\mathbf{a} = a_1 a_2 \dots a_n$ in $H_r(\mathbf{z})$, the w positions are filled by symbols from A_α, and the remaining $n - w$ positions are filled by symbols from A_β. Now, without loss of generalisation, assume that, in the sequence $\mathbf{z}$, symbols at first w positions are from A_α, and the remaining $n - w$ positions are from A_β, i.e., $z_i \in A_\alpha$ for $1 \le i \le w$, and $z_i \in A_\beta$ for $w + 1 \le i \le n$. For positive integers j and k, as illustrated in Figure 2, let

- for $1 \le i \le k$, symbols $z_i, a_i \in A_\alpha$ such that $z_i \ne a_i$,
- for $k + 1 \le i \le k + j$, symbols $z_i \in A_\alpha$ and $a_i \in A_\beta$, and thus, $z_i \ne a_i$, and
- for $k + j + 1 \le i \le w$, symbols $z_i, a_i \in A_\alpha$ such that $z_i = a_i$.

And therefore,

- from the property $w(\mathbf{z}) = w(\mathbf{a}) = w$, for $w + 1 \le i \le w + j$, $z_i \in A_\beta$ and $a_i \in A_\alpha$, and thus, $z_i \ne a_i$,
- from the Hamming distance property $d_H(\mathbf{z}, \mathbf{a}) = r$, for $w + j + 1 \le i \le r + w - k - j$, symbols $z_i, a_i \in A_\beta$ such that $z_i \ne a_i$, and
- for $r + w - k - j + 1 \le i \le n$, symbols $z_i, a_i \in A_\beta$ such that $z_i = a_i$.

Thus, for given $\mathbf{z}$, the number of distinct sub-sequences of $\mathbf{a}$ for each case are as follows.

- There are $\binom{w}{k}(2p - 1)^k$ distinct sub-sequences of length k and weight sum $w(z_1 z_2 \dots z_k) = w(a_1 a_2 \dots a_k) = k$.
- There are $\binom{w-k}{j}(2q)^j$ distinct sub-sequences of length j such that $w(z_{k+1} z_{k+2} \dots z_{k+j}) = j$ and $w(a_{k+1} a_{k+2} \dots a_{k+j}) = 0$.
- There are $\binom{n-w}{j}(2p)^j$ distinct sub-sequences of length j such that $w(z_{w+1} z_{w+2} \dots z_{w+j}) = 0$ and $w(a_{w+1} a_{w+2} \dots a_{k+j}) = j$.
- There are $\binom{n-w-j}{r-k-2j}(2q - 1)^{r-k-2j}$ distinct sub-sequences of length $r - k - 2j$ and of weight sum $w(z_{w+j+1} z_{w+j+2} \dots z_{w+r-k-j}) = w(a_{w+j+1} a_{w+j+2} \dots a_{w+r-k-j}) = 0$.
- There is only one sub-sequence for the remaining two cases.

Now, recall that, for integers a and b, the value $\binom{a}{b}$ exists only if $0 \le b \le a$. Therefore, from the cases, for given sequence $\mathbf{z}$, the total number of distinct sequences $\mathbf{a}$ each of length n and weight sum $w(\mathbf{a}) = w$ is equal to the value as given in Equation (3). Hence, it follows the result. ∎

In Lemma 6, it can be observed that the size $H_r(\mathbf{z}_1)$ and $H_r(\mathbf{z}_2)$ are the same for any $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{B}_{2(p+q),w,n}$. Therefore, we can denote the sphere sizes $V_H(\mathbf{z}_1, r)$ and $V_H(\mathbf{z}_2, r)$ with $V_H(n, r, w)$.

E. Bounds on RCCWS Codes

In this section, we have obtained Sphere Packing and Gilbert-Varshamov bounds for RCCWS codes.

1) Sphere Packing (SP) Bound: Sphere Packing bound is calculated for RCCWS codes in Lemma 7 and Theorem 1.

***Lemma** 7:* For any (n, M, d_H, w) RCCWS code $\mathscr{C}$,

$$M \le \left\lfloor \frac{\binom{n}{w} 2^n p^w q^{n-w}}{\sum_{r=0}^{\left\lfloor \frac{d_H - 1}{2} \right\rfloor} V_H(n, r, w)} \right\rfloor,$$

where $V_H(n, r, w)$ is given in Equation (3).

Proof: For any (n, M, d_H, w) RCCWS code $\mathscr{C}$, if $\mathbf{c} \in \mathscr{C}$ then $\mathbf{c}^R, \mathbf{c}^C, \mathbf{c}^{RC} \in \mathscr{C}$. Now, if $\mathbf{c}_1, \mathbf{c}_2 \in \mathscr{C}$ then

- $H_{r_1}(\mathbf{c}_1) \cap H_{r_2}(\mathbf{c}_2) = \emptyset$ for $\mathbf{c}_1 \ne \mathbf{c}_2$,
- $H_{r_1}(\mathbf{c}_1) \cap H_{r_2}(\mathbf{c}_2^R) = \emptyset$ for $\mathbf{c}_1 \ne \mathbf{c}_2^R$,
- $H_{r_1}(\mathbf{c}_1) \cap H_{r_2}(\mathbf{c}_2^C) = \emptyset$ for $\mathbf{c}_1 \ne \mathbf{c}_2^C$, and
- $H_{r_1}(\mathbf{c}_1) \cap H_{r_2}(\mathbf{c}_2^{RC}) = \emptyset$ for $\mathbf{c}_1 \ne \mathbf{c}_2^{RC}$,

where r_1 and r_2 are positive integers such that $0 \le r_1 + r_2 \le d_H - 1$. Thus,

$$\sum_{\mathbf{c} \in \mathscr{C}} \sum_{j=0}^{\left\lfloor \frac{d_H - 1}{2} \right\rfloor} V_H(\mathbf{c}, j, w) \le \binom{n}{w} 2^n p^w q^{n-w}.$$

For any non-negative integers r and any $\mathbf{c}_1, \mathbf{c}_2 \in \mathscr{C}$, from Lemma 6, $V_H(\mathbf{c}_1, r, w) = V_H(\mathbf{c}_2, r, w) = V_H(n, r, w)$. Therefore,

$$M \le \frac{\binom{n}{w} 2^n p^w q^{n-w}}{\sum_{r=0}^{\left\lfloor \frac{d_H - 1}{2} \right\rfloor} V_H(n, r, w)}.$$

Thus, the result follows from the fact that the size M is a positive integer. ∎

***Theorem** 1:* For any given non-negative integers p, q, n, d_H and w ($\le n$), the maximum size of RCCWS code with given length n and weight sum w over $\mathcal{A}_{2(p+q)}$ is

$$A_{2(p+q)}^{R,C,RC,W}(n, d_H, w) \le \left\lfloor \frac{\binom{n}{w} 2^n p^w q^{n-w}}{\sum_{r=0}^{\left\lfloor \frac{d_H - 1}{2} \right\rfloor} V_H(n, r, w)} \right\rfloor,$$

where $V_H(n, r, w)$ is given in Equation (3).

Proof: The proof follows from Lemma 7 for the $(n, A_{2(p+q)}^{R,C,RC,W}(n, d_H, w), d_H, w)$ RCCWS code. ∎

$$V_H(n,r,w) = \sum_{k=0}^{\min\{r,w\}} \left(\sum_{j=\max\{0,w+r-n-k\}}^{\min\{w-k,n-w,\lfloor (r-k)/2 \rfloor\}} \binom{w}{k}\binom{w-k}{j}\binom{n-w}{j}\binom{n-w-j}{r-k-2j} 2^{2j} p^j q^j (2p-1)^k (2q-1)^{r-k-2j} \right) \tag{3}$$

$$V(n,r,w) = \sum_{k=0}^{\min\{r,w\}} \left(\sum_{j=\max\{0,w+r-n-k\}}^{\min\{w-k,n-w,\lfloor (r-k)/2 \rfloor\}} \binom{w}{k}\binom{w-k}{j}\binom{n-w}{j}\binom{n-w-j}{r-k-2j} 2^{2j} \right) \tag{4}$$

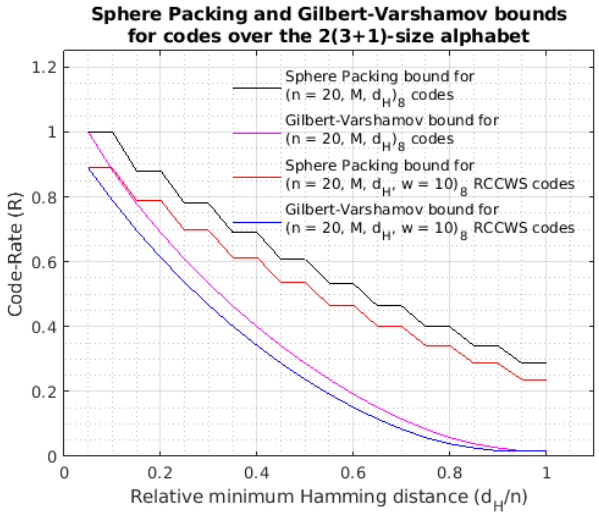

Figure 3. For $p = 3$ and $q = 1$, bounds are plotted for codes over $\mathcal{A}_8$.

2) Gilbert-Varshamov (GV) Bound: Gilbert-Varshamov bound is calculated for RCCWS codes in Theorem 2.

***Theorem** 2:* For any given non-negative integers p, q, n, d_H and w ($\leq n$), the maximum size of RCCWS code with length n and weight sum w over $\mathcal{A}_{2(p+q)}$ is

$$A^{R,C,RC,W}_{2(p+q)}(n,d_H,w) \geq \left\lceil \frac{\binom{n}{w}2^n p^w q^{n-w}}{\sum_{r=0}^{d_H-1} V_H(n,r,w)} \right\rceil,$$

where $V_H(n,r,w)$ is given in Equation (3).

Proof: For given n, d and w, the maximum size of RCCWS code $\mathscr{C}$ is denoted by $A^{R,C,RC,W}_{2(p+q)}(n,d_H,w)$. Therefore, for any $\mathbf{a} \in \mathcal{B}_{2(p+q),w,n}$, there exists a sequence $\mathbf{c} \in \mathscr{C}$ such that $\mathbf{a} \in H_{d-1}(\mathbf{c})$. But, for each $\mathbf{c} \in \mathscr{C}$, the size of the Hamming Sphere $H_{d-1}(\mathbf{c})$ is $V_H(n,d-1,w)$, where $V_H(n,d-1,w)$ is given in Equation (3) with $r = d-1$. Thus,

$$A^{R,C,RC,W}_{2(p+q)}(n,d_H,w) \cdot V_H(n,d-1,w) \geq \binom{n}{w}2^n p^w q^{n-w}.$$

Now, the result follows from Lemma 4 and the fact that the size $A^{R,C,RC,W}_{2(p+q)}(n,d_H,w)$ is an integer. ∎

Bounds on various codes over $\mathcal{A}_8$ are plotted in Figure 1. In this figure, for code length $n = 5,6,\ldots,100$ and the minimum Hamming distance $d_H = 3$, we have plotted following.

1) From Equation (1), Sphere Packing bound on code-rate
$$\frac{1}{n}\log_8\left(8^n \Big/ \sum_{r=0}^{\lfloor (d_H-1)/2 \rfloor}\binom{n}{r}7^r\right)$$
for codes over the alphabet $\mathcal{A}_8$.
2) From Equation (2), Gilbert-Varshamov bound on code-rate
$$\frac{1}{n}\log_8\left(8^n \Big/ \sum_{r=0}^{d_H-1}\binom{n}{r}7^r\right)$$
for codes over the alphabet $\mathcal{A}_8$.
3) From Theorem 1, Sphere Packing bound on code-rate
$$\frac{1}{n}\log_8\left(\left\lfloor \binom{n}{w}2^n p^w q^{n-w} \Big/ \sum_{r=0}^{\lfloor (d-1)/2 \rfloor} V(n,r,w)\right\rfloor\right)$$
for RCCWS codes over the alphabet $\mathcal{A}_8$ with
   a) $p = 3$ and $q = 1$ and
   b) $p = 1$ and $q = 3$.
4) From Theorem 2, Gilbert-Varshamov bound on code-rate
$$\frac{1}{n}\log_8\left(\left\lfloor \binom{n}{w}2^n p^w q^{n-w} \Big/ \sum_{r=0}^{d-1} V(n,r,w)\right\rfloor\right)$$
for RCCWS codes over the alphabet $\mathcal{A}_8$ with
   a) $p = 3$ and $q = 1$ and
   b) $p = 1$ and $q = 3$.

In Figure 3, we have plotted code-rate vs. relative minimum Hamming distance for codes over $\mathcal{A}_8$ with $p = 3$, and $q = 1$, where the relative minimum Hamming distance is d_H/n. Now, let us look at the loss in code rate as a result of these constraints for asymptotically large n.

***Remark** 1:* For given length n and weight sum w, the loss on RCCWS code-rate (compared to unconstrained code) is given by

$$\begin{aligned} 1 - \frac{1}{n}\log_{2(p+q)} A^{R,C,RC,W}_{2(p+q)}(n,d_H=1,w) &= 1 - \frac{1}{n}\log_{2(p+q)}\left(\binom{n}{w}2^n p^w q^{n-w}\right) \\ &= \log_{2(p+q)}\left(\frac{p+q}{q}\right) + \frac{w}{n}\log_{2(p+q)}\left(\frac{q}{p}\right) \\ &\quad - \frac{1}{n}\log_{2(p+q)}\binom{n}{w}. \end{aligned}$$

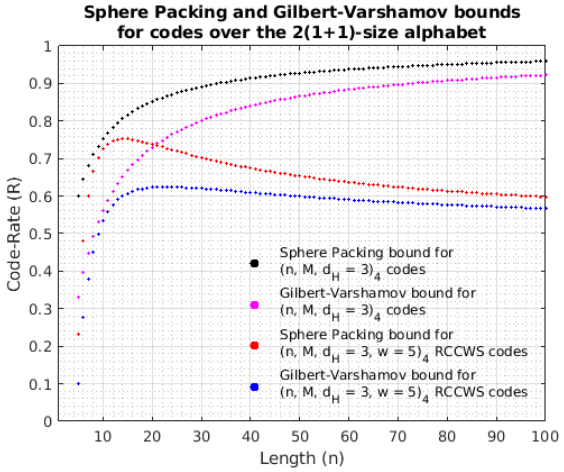

Figure 4. For $n = 5, 6, \ldots, 100$, and $d_H = 3$, Sphere Packing and Gilbert-Varshamov bounds are plotted for DNA codes.

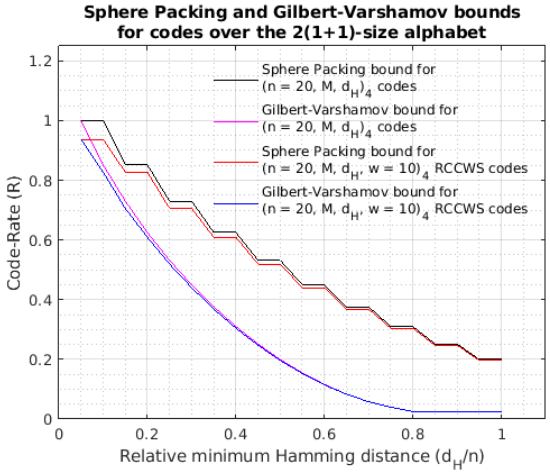

Figure 5. For $n = 20$, and $d_H = 1, 2, \ldots, n$, Sphere Packing and Gilbert-Varshamov bounds for DNA codes are illustrated.

For given w and large n, the loss on RCCWS code-rate (compared to unconstrained code) is

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\log_{2(p+q)}A^{R,C,RC,W}_{2(p+q)}(n,d_H=1,w)\right)$$
$$=1-\frac{1}{n}\log_{2(p+q)}\left(\binom{n}{w}2^n p^w q^{n-w}\right)$$
$$=\log_{2(p+q)}\left(\frac{p+q}{q}\right).$$

More precisely, for given length n, weight sum w and the minimum Hamming distance d_H (≥ 1), the loss on RCCWS code-rate (compared to unconstrained code) is

$$1-\frac{1}{n}\log_{2(p+q)}A^{R,C,RC,W}_{2(p+q)}(n,d_H,w)$$
$$=\log_{2(p+q)}\left(\frac{p+q}{q}\right)+\frac{w}{n}\log_{2(p+q)}\left(\frac{q}{p}\right)$$
$$-\frac{1}{n}\log_{2(p+q)}\binom{n}{w}+\frac{1}{n}\log_{2(p+q)}\sum_{r=0}^{\lfloor\frac{d_H-1}{2}\rfloor}V_H(n,r,w).$$

For given weight sum w and given minimum Hamming distance d_H, the loss on RCCWS code-rate (compared to unconstrained code) is

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\log_{2(p+q)}A^{R,C,RC,W}_{2(p+q)}(n,d_H,w)\right)$$
$$=\log_{2(p+q)}\left(\frac{p+q}{q}\right).$$

IV. Bounds on DNA codes with constraints

For $p = 1$ and $q = 1$, the alphabet $\mathcal{A}_4 = \{\alpha_1, \alpha_2, \beta_1, \beta_2\}$, where $\alpha_1^C = \alpha_2$, $\alpha_2^C = \alpha_1$, $\beta_1^C = \beta_2$ and $\beta_2^C = \beta_1$. Now, a correspondence can be established between the alphabet $\mathcal{A}_4$ and the DNA alphabet $\{G, C, A, T\}$ such that the nucleotides C and G are corresponds to the symbols α_1 and α_2, and the nucleotides A and T are corresponds to symbols β_1 and

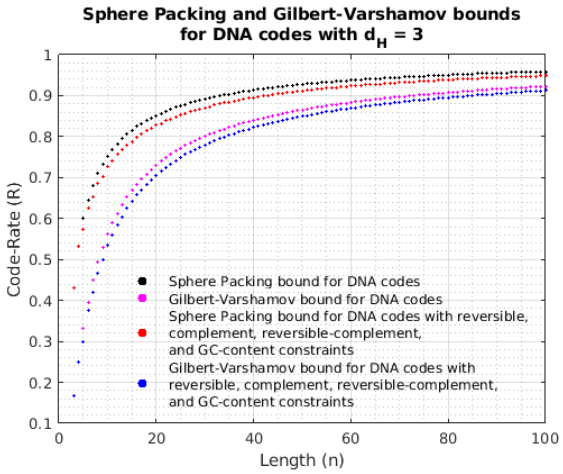

Figure 6. For $n = 3, 4, \ldots, 100$, and $d_H = 3$, Sphere Packing and Gilbert-Varshamov bounds for DNA codes with the minimum distance $d_H = 3$ are shown.

β_2. Then the bounds as given in Theorem 1 and Theorem 2 can be studied for DNA codes. For example, consider $\alpha_1 = G$, $\alpha_2 = C$, $\beta_1 = A$, and $\beta_2 = T$ then any RCCWS code over $\mathcal{A}_4$ is a GC-balanced DNA codes with reversible, complement and reversible-complement constraints. Also, the weight sum of any sequence over $\mathcal{A}_4$ corresponds to the GC-weight of the DNA sequence. Again, any RCCWS code with weight sum w over the alphabet $\mathcal{A}_4$ represents the DNA code that satisfies reversible, complement, reversible-complement constraints, where all the DNA sequences have the same GC-weight w. For $p = q = 1$, from Equation (3), the Hamming sphere size, $V(n, r, w)$, for GC-balanced DNA codes with reversible, complement and reversible-complement constraints is given in Equation (4). Bounds on the maximum size of GC-balanced DNA codes that satisfy reversible and reversible-complement constraints are obtained in Theorem 3 and Theorem 4.

***Theorem** 3:* For any given non-negative integers n, d_H

Table I
LOWER BOUND ON $A_4^{R,C,RC,GC}(n, d_H, \lfloor n/2 \rfloor)$ FROM THEOREM 4

n	d_H									
	1	2	3	4	5	6	7	8	9	10
3	24	6	2	-	-	-	-	-	-	-
4	96	20	4	2	-	-	-	-	-	-
5	320	54	8	3	2	-	-	-	-	-
6	1280	183	23	6	3	2	-	-	-	-
7	4480	560	59	13	4	2	2	-	-	-
8	17920	1992	178	34	9	4	2	2	-	-
9	64512	6452	512	84	19	6	3	2	2	-
10	258048	23459	1655	240	46	13	5	3	2	2

Table II
UPPER BOUND ON $A_4^{R,C,RC,GC}(n, d_H, \lfloor n/2 \rfloor)$ FROM THEOREM 3

n	d_H				
	1, 2	3, 4	5, 6	7, 8	9, 10
3	24	6, -	-	-	-
4	96	19	-	-	-
5	320	53	8, -	-	-
6	1280	182	22	-	-
7	4480	560	58	12, -	-
8	17920	1991	177	33	-
9	64512	6451	512	83	18, -
10	258048	23458	1654	239	45

and w ($\leq n$), the maximum size of reversible, complement, reversible-complement DNA code with length n and fixed GC weight w is

$$A_4^{R,C,RC,GC}(n, d_H, w) \leq \left\lfloor \frac{\binom{n}{w} 2^n}{\sum_{r=0}^{\lfloor \frac{d-1}{2} \rfloor} V(n, r, w)} \right\rfloor.$$

Proof: The proof follows from Theorem 1, and the fact that there exists a DNA code that satisfies reversible, complement, reversible-complement constraints, where all the DNA sequences have the same GC-weight w for any RCCWS code with weight sum w over the alphabet $\mathcal{A}_4$ with $p = 1$ and $q = 1$. ∎

***Theorem** 4:* For any given non-negative integers n, d and w ($\leq n$), the maximum size of reversible, complement, reversible-complement DNA code with length n and fixed GC weight w is

$$A_4^{R,C,RC,GC}(n, d_H, w) \geq \left\lceil \frac{\binom{n}{w} 2^n}{\sum_{r=0}^{d-1} V(n, r, w)} \right\rceil.$$

Proof: The proof follows from Theorem 2, and the fact that there exists a DNA code that satisfies reversible, complement, reversible-complement constraints, where all the DNA sequences have the same GC-weight w for any RCCWS code with weight sum w over the alphabet $\mathcal{A}_4$ with $p = 1$ and $q = 1$. ∎

Now, in Table I and Table II, we compute bounds on the maximum size $A_4^{R,C,RC,GC}(n, d_H, \lfloor n/2 \rfloor)$ of DNA code with length n ($= 3, 4, \ldots, 10$) and the minimum Hamming distance d_H ($= 1, 2, \ldots, n$), where the DNA code satisfies reversible, complement, reversible-complement, and GC-content constraints. Further, in Figure 4, Figure 5, and Figure 6, bounds on various DNA codes are studied. In these figures, the black, magenta, red and blue curves represent $\frac{1}{n}\log_4\left(4^n / \sum_{r=0}^{\lfloor (d_H-1)/2 \rfloor} \binom{n}{r} 3^r\right)$ (Sphere Packing bound on DNA codes from Equation (1)), $\frac{1}{n}\log_4\left(4^n / \sum_{r=0}^{d_H-1} \binom{n}{r} 3^r\right)$ (Gilbert-Varshamov bound on DNA codes from Equation (2)), $\frac{1}{n}\log_4\left(\left\lfloor \binom{n}{w} 2^n / \sum_{r=0}^{\lfloor (d-1)/2 \rfloor} V(n, r, w) \right\rfloor\right)$ (Sphere Packing bound on DNA codes from Theorem 3), and $\frac{1}{n}\log_4\left(\left\lfloor \frac{\binom{n}{w} 2^n}{\sum_{r=0}^{d-1} V(n,r,w)} \right\rfloor\right)$ (Gilbert-Varshamov bound on DNA codes from Theorem 4), respectively.

REFERENCES

[1] J. H. Van Lint, "A survey of perfect codes," *Rocky Mountain Journal of Mathematics*, vol. 5, no. 2, pp. 199–224, 1975.

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

[3] 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.

[4] ——, "Constant-weight code bounds from spherical code bounds," in *IEEE International Symposium on Information Theory (Cat. No.00CH37060)*, 2000, pp. 391–391.

[5] A. Ashikhmin, A. Barg, and S. Litsyn, "New upper bounds on generalized weights," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1258–1263, 1999.

[6] I. Krasikov and S. Litsyn, "On upper bounds for the distance of codes of small size," in *Proceedings of IEEE International Symposium on Information Theory*, 1997, pp. 84–84.

[7] G. M. Church, Y. Gao, and S. Kosuri, "Next-generation digital information storage in DNA," *Science*, vol. 337, no. 6102, pp. 1628–1628, Sept 2012.

[8] 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, pp. 77–80, Jan 2013.

[9] D. Limbachiya, B. Rao, and M. K. Gupta, "The Art of DNA Strings: Sixteen Years of DNA Coding Theory," *arXiv e-prints*, p. arXiv:1607.00266, Jul 2016.

[10] A. Marathe, A. E. Condon, and R. M. Corn, "On combinatorial DNA word design," *Journal of Computational Biology*, vol. 8, no. 3, pp. 201–219, 2001.

[11] D. Limbachiya, K. G. Benerjee, B. Rao, and M. K. Gupta, "On DNA codes using the ring $\mathbb{Z}_4 + w\mathbb{Z}_4$," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 2401–2405.

[12] V. V. Rykov, A. J. Macula, D. C. Torney, and P. S. White, "DNA sequences and quaternary cyclic codes," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, 2001, pp. 248–248.

[13] K. G. Benerjee and A. Banerjee, "On DNA codes with multiple constraints," *IEEE Communications Letters*, vol. 25, no. 2, pp. 365–368, 2021.

[14] K. G. Benerjee and A. Banerjee, "On homopolymers and secondary structures avoiding, reversible, reversible-complement and GC-balanced DNA codes," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 204–209.

[15] Y. M. Chee and S. Ling, "Improved lower bounds for constant GC-content DNA codes," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 391–394, 2008.

[16] O. D. King, "Bounds for DNA codes with constant GC-content," *The Electronic Journal of Combinatorics*, vol. 10, article no. R33, 2003.

[17] Q. Zhang and C. Xu, "Lower bounds of DNA codes with reverse constraint," *Journal of Computational and Theoretical Nanoscience*, vol. 7, no. 10, pp. 2072–2076, 2010.

[18] D. Limbachiya, M. K. Gupta, and V. Aggarwal, "Family of constrained codes for archival DNA data storage," *IEEE Communications Letters*, vol. 22, no. 10, pp. 1972–1975, 2018.