On the Size of Balls and Anticodes of Small Diameter Under the Fixed-Length Levenshtein Metric

Daniella Bar-Lev, *Student Member, IEEE*, Tuvi Etzion, *Life Fellow, IEEE*, and Eitan Yaakobi, *Senior Member, IEEE*

***Abstract*— The rapid development of DNA storage has brought the deletion and insertion channel to the front line of research. When the number of deletions is equal to the number of insertions, the *Fixed Length Levenshtein* (FLL) metric is the right measure for the distance between two words of the same length. Similar to any other metric, the size of a ball is one of the most fundamental parameters. In this work, we consider the minimum, maximum, and average size of a ball with radius one, in the FLL metric. The related minimum and the maximum size of a maximal anticode with diameter one are also considered.**

***Index Terms*— Levenshtein metric, balls, anticodes.**

I. Introduction

CODING for DNA storage has attracted significant attention in the previous decade due to recent experiments and demonstrations of the viability of storing information in macromolecules [2], [4], [9], [12], [14], [15], [27], [34], [37]. Given the trends in cost decreases of DNA synthesis and sequencing, it is estimated that already within this decade DNA storage may become a highly competitive archiving technology. However, DNA molecules induce error patterns that are fundamentally different from their digital counterparts [17], [18], [21], [29]; This distinction results from the specific error behavior in DNA and it is well-known that errors in DNA are typically in the form of substitutions, insertions, and deletions, where most published studies report that deletions are the most prominent ones, depending upon the specific technology for synthesis and sequencing. Hence, due to its high relevance to the error model in DNA storage coding for insertion and deletion errors has received renewed interest recently; see e.g. [5], [6], [7], [8], [10], [13], [16], [25], [26], [28], [32], [33], [35]. This paper takes one more step in advancing this study and its goal is to study the size of balls and anticodes when the number of insertions equals to the number of deletions.

If a word $\boldsymbol{x} \in \mathbb{Z}_q^n$ can be transferred to a word $\boldsymbol{y} \in \mathbb{Z}_q^n$ using t deletions and t insertions (and cannot be transferred using a smaller number of deletions and insertions), then their *Fixed Length Levenshtein* (FLL) *distance* is t, which is denoted by $d_\ell(\boldsymbol{x}, \boldsymbol{y}) = t$. It is relatively easy to verify that the FLL distance defines a metric. Let $G = (V, E)$ be a graph whose set of vertices $V = \mathbb{Z}_q^n$ and two vertices $\boldsymbol{x}, \boldsymbol{y} \in V$ are connected by an edge if $d_\ell(\boldsymbol{x}, \boldsymbol{y}) = 1$. This graph represents the FLL distance. Moreover, the FLL distance defines a *graphic metric*, i.e., it is a metric and for each $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}_q^n$, $d_\ell(\boldsymbol{x}, \boldsymbol{y}) = t$ if and only if the length of the shortest path between $\boldsymbol{x}$ and $\boldsymbol{y}$ in G is t.

One of the most fundamental parameters in any metric is the size of a ball with a given radius t centered at a word $\boldsymbol{x}$. There are many metrics, e.g. the Hamming metric, the Johnson metric, or the Lee metric, where the size of a ball does not depend on the word $\boldsymbol{x}$. This is not the case in the FLL metric. Moreover, the graph G has a complex structure and it makes it much more difficult to find the exact size of any ball and even the size of a ball with minimum size and the size of a ball with maximum size. In [30], a formula for the size of the ball with radius one, centered at a word $\boldsymbol{x}$, in the FLL metric was given. This formula depends on the number of runs in the word and the lengths of its alternating segments (an alternating segment is a substring of consecutive symbols, where no two consecutive symbols are the same). Nevertheless, while it is easy to compute the minimum size of a ball, it is still difficult to determine from this formula what the maximum size of a ball is. In this paper, we find explicit expressions for the minimum and maximum sizes of a ball when the ball is of radius one. We also find the average size of a ball when the radius of the ball is one. Finally, we consider the related basic concept of anticode in the FLL metric, where an anticode with diameter D is a code where the distance between any two elements of the code is at most D. Note, that a ball with radius R has diameter $2R$ and hence it is an anticode with diameter $2R$. We find the maximum size and the minimum size

Manuscript received 15 June 2022; revised 7 October 2022; accepted 23 November 2022. Date of publication 6 December 2022; date of current version 17 March 2023. The work of Daniella Bar-Lev was supported in part by the ISF under Grant 222/19. The work of Tuvi Etzion was supported in part by the Israel Science Foundation (ISF) under Grant 222/19 and in part by the Technion Data Science Initiative. The work of Eitan Yaakobi was supported in part by the Israel Innovation Authority under Grant 75855 and in part by the Technion Data Science Initiative. An earlier version of this paper was presented in part at the 2021 IEEE International Symposium on Information Theory [DOI: 10.1109/ISIT45174.2021.9517922].
(Corresponding author: Daniella Bar-Lev.)

The authors are with the Department of Computer Science, Technion—Israel Institute of Technology, Haifa 3200003, Israel (e-mail: daniellalev@cs.technion.ac.il; etzion@cs.technion.ac.il; yaakobi@cs.technion.ac.il).

Communicated by X. Zhang, Associate Editor for Coding and Decoding.

Digital Object Identifier 10.1109/TIT.2022.3227128

of maximal anticodes with diameter one, where an anticode with diameter one is maximal if any addition of a word to it will increase its diameter.

This paper is the first one which considers a comprehensive discussion and exact computation on the balls with radius one and the anticodes with diameter one in the FLL metric. The rest of this paper is organized as follows. Section II introduces some basic concepts, presents some of the known results on the sizes of balls, presents some results on equivalence of codes correcting deletions and insertions, and finally introduce some observations required for our exposition. The minimum size of a ball of any given radius in the FLL metric over $\mathbb{Z}_q$ is discussed in Section III. Section IV is devoted for the discussion on the maximum size of a ball with radius one in the FLL metric over $\mathbb{Z}_q$. The analysis of non-binary sequences is discussed in Section IV-A. It appears that contrary to many other coding problems the binary case is much more difficult to analyze and it is discussed in Section IV-B. For the binary case, the sequence for which the maximum size is obtained is presented in Theorem 8 and the maximum size is given in Corollary 6. The average size of the FLL ball with radius one over $\mathbb{Z}_q$ is computed in Section V and proved in Theorem 13. In Section VI, we consider binary maximal anticodes with diameter one. The maximum size of such an anticode is discussed in Section VI-A and Section VI-B is devoted to the minimum size of such anticodes. The results can be generalized for the non-binary case, but since they are more complicated and especially messy, they are omitted.

II. Definitions and Previous Results

In this section, we present the definitions and notations as well as several results that will be used throughout the paper.

For an integer $q \geq 2$, let $\mathbb{Z}_q$ denote the set of integers $\{0, 1, \ldots, q-1\}$ and for an integer $n \geq 0$, let $\mathbb{Z}_q^n$ be the set of all sequences (words) of length n over the alphabet $\mathbb{Z}_q$, let $\mathbb{Z}_q^* = \bigcup_{n=0}^{\infty} \mathbb{Z}_q^n$, and let $[n]$ denote the set of integers $\{1, 2, \ldots, n\}$. For two sequences $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}_q^n$, the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, $d(\boldsymbol{x}, \boldsymbol{y})$, can be measured in various ways. When the type of errors is substitution, the *Hamming distance* is the most natural to be considered. The *Hamming weight* of a sequence $\boldsymbol{x} \in \mathbb{Z}_q^*$, denoted by $\text{wt}(\boldsymbol{x})$, is equal to the number of nonzero coordinates in $\boldsymbol{x}$. The Hamming distance between two sequences $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}_q^n$, denoted by $d_H(\boldsymbol{x}, \boldsymbol{y})$, is the number of coordinates in which $\boldsymbol{x}$ and $\boldsymbol{y}$ differ. In other words, $d_H(\boldsymbol{x}, \boldsymbol{y})$ is the number of symbol-substitution operations required to transform $\boldsymbol{x}$ into $\boldsymbol{y}$. The Hamming distance is well known to be a metric on $\mathbb{Z}_q^n$ (also referred as the *Hamming space*), as it satisfies the three conditions of a metric (i.e., coincidence, symmetry and the triangle inequality). Given a distance d on a space V, the *t-ball* centered at $\boldsymbol{x} \in V$ is the set $\{\boldsymbol{y} : d(\boldsymbol{x}, \boldsymbol{y}) \leq t\}$. The *t-sphere* centered at $\boldsymbol{x} \in V$ is the set $\{\boldsymbol{y} : d(\boldsymbol{x}, \boldsymbol{y}) = t\}$. A *code* $\mathcal{C} \subseteq V$ is a subset of words from V. A related concept is an *anticode* with diameter D which is a code in V for which the distance between any two elements is at most D. Clearly, a t-ball is an anticode whose diameter is at most $2t$. The *Hamming t-ball* centered at $\boldsymbol{x} \in \mathbb{Z}_q^n$ will be denoted by $\mathcal{H}_t(\boldsymbol{x})$. For $\boldsymbol{x} \in \mathbb{Z}_q^n$, the number of words in the Hamming t-ball is a function of n, q and t. The number of such words is

$$|\mathcal{H}_t(\boldsymbol{x})| = \sum_{i=0}^{t} \binom{n}{i} (q-1)^i. \tag{1}$$

For an integer t, $0 \leq t \leq n$, a sequence $\boldsymbol{y} \in \mathbb{Z}_q^{n-t}$ is a *t-subsequence* of $\boldsymbol{x} \in \mathbb{Z}_q^n$ if $\boldsymbol{y}$ can be obtained from $\boldsymbol{x}$ by deleting t symbols from $\boldsymbol{x}$. In other words, there exist $n-t$ indices $1 \leq i_1 < i_2 < \cdots < i_{n-t} \leq n$ such that $y_j = x_{i_j}$, for all $1 \leq j \leq n-t$. We say that $\boldsymbol{y}$ is a *subsequence* of $\boldsymbol{x}$ if $\boldsymbol{y}$ is a t-subsequence of $\boldsymbol{x}$ for some t. Similarly, a sequence $\boldsymbol{y} \in \mathbb{Z}_q^{n+t}$ is a *t-supersequence* of $\boldsymbol{x} \in \mathbb{Z}_m^n$ if $\boldsymbol{x}$ is a t-subsequence of $\boldsymbol{y}$ and $\boldsymbol{y}$ is a *supersequence* of $\boldsymbol{x}$ if $\boldsymbol{y}$ is a t-supersequence of $\boldsymbol{x}$ for some t.

Definition 1: The *deletion t-sphere* centered at $\boldsymbol{x} \in \mathbb{Z}_q^n$, $\mathcal{D}_t(\boldsymbol{x}) \subseteq \mathbb{Z}_q^{n-t}$, is the set of all t-subsequences of $\boldsymbol{x}$. The size of the largest deletion t-sphere in $\mathbb{Z}_q^n$ is denoted by $D_q(n, t)$. The *insertion t-sphere* centered at $\boldsymbol{x} \in \mathbb{Z}_q^n$, $\mathcal{I}_t(\boldsymbol{x}) \subseteq \mathbb{Z}_q^{n+t}$, is the set of all t-supersequences of $\boldsymbol{x}$.

Let $\boldsymbol{x} \in \mathbb{Z}_q^n$ be a sequence. The size of the insertion t-sphere $|\mathcal{I}_t(\boldsymbol{x})|$ does not depend on $\boldsymbol{x}$ for any $0 \leq t \leq n$. To be exact, it was shown by Levenshtein [22] that

$$|\mathcal{I}_t(\boldsymbol{x})| = \sum_{i=0}^{t} \binom{n+t}{i} (q-1)^i. \tag{2}$$

On the other hand, calculating the exact size of the deletion sphere is one of the more intriguing problems when studying codes for deletions. Unlike substitution balls and insertions spheres, not all deletion spheres are of the same size. That is, the size of the deletion sphere, $|\mathcal{D}_t(\boldsymbol{x})|$, depends on the choice of the sequence $\boldsymbol{x}$. Let $\{\sigma_1, \ldots, \sigma_q\}$ be the symbols of $\mathbb{Z}_q$ in some order and let $\boldsymbol{c}(n) = (c_1, c_2, \ldots, c_n)$ be a sequence in $\mathbb{Z}_q^n$ such that $c_i = \sigma_i$ for $1 \leq i \leq q$ and $c_i = c_{i-q}$ for $i > q$. It was shown in Hirschberg and Regnier [19] that $\boldsymbol{c}(n)$ has the largest deletion t-sphere and its size is given by

$$D_q(n, t) = |\mathcal{D}_t(\boldsymbol{c}(n))| = \sum_{i=0}^{t} \binom{n-t}{i} D_{q-1}(t, t-i)$$

In particular, $D_2(n, t) = \sum_{i=0}^{t} \binom{n-t}{i}$ and $D_3(n, t) = \sum_{i=0}^{t} \binom{n-t}{i} \sum_{j=0}^{t-i} \binom{i}{j}$. The value $D_2(n, t)$ also satisfies the following recursion

$$D_2(n, t) = D_2(n-1, t) + D_2(n-2, t-1),$$

where the values for the basic cases can be evaluated by $D_2(n, t) = \sum_{i=0}^{t} \binom{n-t}{i}$.

Definition 2: A *run* is a maximal subsequence composed of consecutive identical symbols. For a sequence $\boldsymbol{x} \in \mathbb{Z}_q^n$, the number of runs in $\boldsymbol{x}$ is denoted by $\rho(\boldsymbol{x})$.

Example 1: If $\boldsymbol{x} = 0000000$ then $\rho(\boldsymbol{x}) = 1$ since $\boldsymbol{x}$ has a single run of length 7 and for $\boldsymbol{y} = 1120212$ we have that $\rho(\boldsymbol{y}) = 6$ since $\boldsymbol{y}$ has six runs, the first is of length two and the others are of length one.

There are upper and lower bounds on the size of the deletion ball which depend on the number of runs in the sequence. Namely, it was shown by Levenshtein [22] that

$$\binom{\rho(\boldsymbol{x}) - t + 1}{t} \leq |\mathcal{D}_t(\boldsymbol{x})| \leq \binom{\rho(\boldsymbol{x}) + t - 1}{t}.$$

Later, the lower bound was improved in [19]:

$$\sum_{i=0}^{t}\binom{\rho(\boldsymbol{x})-t}{i} \leq |\mathcal{D}_t(\boldsymbol{x})| \leq \binom{\rho(\boldsymbol{x})+t-1}{t}. \quad (3)$$

Several more results on this value which take into account the number of runs appear in [24].

The *Levenshtein distance* between two words $\boldsymbol{x},\boldsymbol{y} \in \mathbb{Z}_q^*$, denoted by $d_L(\boldsymbol{x},\boldsymbol{y})$, is the minimum number of insertions and deletions required to transform $\boldsymbol{x}$ into $\boldsymbol{y}$. Similarly, for two sequences $\boldsymbol{x},\boldsymbol{y} \in \mathbb{Z}_q^*$, $d_E(\boldsymbol{x},\boldsymbol{y})$ denotes the *edit* distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, which is the minimum number of insertions, deletions, and substitutions required to transform $\boldsymbol{x}$ into $\boldsymbol{y}$.

Definition 3: Let t,n be integers such that $0 \leq t \leq n$. For a sequence $\boldsymbol{x} \in \mathbb{Z}_q^n$, the Levenshtein t-ball centered at $\boldsymbol{x} \in \mathbb{Z}_q^n$, $\widehat{\mathcal{L}}_t(\boldsymbol{x})$, is defined by

$$\widehat{\mathcal{L}}_t(\boldsymbol{x}) \triangleq \{\boldsymbol{y} \in \mathbb{Z}_q^* \ \ d_L(\boldsymbol{x},\boldsymbol{y}) \leq t\}.$$

In case $\boldsymbol{x},\boldsymbol{y} \in \mathbb{Z}_q^n$, for some integer n, the *Fixed Length Levenshtein* (FLL) *distance* between $\boldsymbol{x}$ and $\boldsymbol{y}$, $d_\ell(\boldsymbol{x},\boldsymbol{y})$, is the smallest t for which there exists a t-subsequence $\boldsymbol{z} \in \mathbb{Z}_q^{n-t}$ of both $\boldsymbol{x}$ and $\boldsymbol{y}$, i.e.

$$d_\ell(\boldsymbol{x},\boldsymbol{y}) = \min\{t' : \mathcal{D}_{t'}(\boldsymbol{x}) \cap \mathcal{D}_{t'}(\boldsymbol{y}) \neq \varnothing\} = \frac{d_L(\boldsymbol{x},\boldsymbol{y})}{2}. \quad (4)$$

In other words, t is the smallest integer for which there exists $\boldsymbol{z} \in \mathbb{Z}_q^{n-t}$ such that $\boldsymbol{z} \in \mathcal{D}_t(\boldsymbol{x})$ and $\boldsymbol{y} \in \mathcal{I}_t(\boldsymbol{z})$. Note that if $\boldsymbol{x},\boldsymbol{y} \in \mathbb{Z}_q^n$ and $\boldsymbol{x}$ is obtained from $\boldsymbol{y}$ by t_1 deletions and t_2 insertions, then $t_1 = t_2$.

Definition 4: Let n,t be integers such that $0 \leq t \leq n$. For a sequence $\boldsymbol{x} \in \mathbb{Z}_q^n$, the FLL t-ball centered at $\boldsymbol{x} \in \mathbb{Z}_q^n$, $\mathcal{L}_t(\boldsymbol{x}) \subseteq \mathbb{Z}_q^n$, is defined by

$$\mathcal{L}_t(\boldsymbol{x}) \triangleq \{\boldsymbol{y} \in \mathbb{Z}_q^n \ \ d_\ell(\boldsymbol{x},\boldsymbol{y}) \leq t\}.$$

We say that a subsequence $\boldsymbol{x}_{[i,j]} \triangleq x_i x_{i+1} \cdots x_j$ is an *alternating segment* if $\boldsymbol{x}_{[i,j]}$ is a sequence of alternating distinct symbols $\sigma,\sigma' \in \mathbb{Z}_m$. Note that $\boldsymbol{x}_{[i,j]}$ is a *maximal alternating segment* if $\boldsymbol{x}_{[i,j]}$ is an alternating segment and $\boldsymbol{x}_{[i-1,j]}, \boldsymbol{x}_{[i,j+1]}$ are not. The number of maximal alternating segments of a sequence $\boldsymbol{x}$ will be denoted by $A(\boldsymbol{x})$.

Example 2: If $\boldsymbol{x} = 0000000$ then $A(\boldsymbol{x}) = 7$ since $\boldsymbol{x}$ has seven maximal alternating segments, each of length one, and for $\boldsymbol{x} = 1120212$ we have that $A(\boldsymbol{x}) = 4$ and the maximal alternating segments are $1,\ 12,\ 202,\ 212$.

The following formula to compute $|\mathcal{L}_1(\boldsymbol{x})|$ as a function of $\rho(\boldsymbol{x})$ and $A(\boldsymbol{x})$ was given in [30]

$$|\mathcal{L}_1(\boldsymbol{x})| = \rho(\boldsymbol{x}) \cdot (n(q-1)-1) + 2 - \sum_{i=1}^{A(\boldsymbol{x})} \frac{(s_i-1)(s_i-2)}{2}, \quad (5)$$

where s_i for $1 \leq i \leq A(\boldsymbol{x})$ denotes the length of the i-th maximal alternating segment of $\boldsymbol{x}$.

Note that $|\widehat{\mathcal{L}}_1(\boldsymbol{x})|$, $|\widehat{\mathcal{L}}_2(\boldsymbol{x})|$ can be deduced from equations (2), (3), (4), and $|\mathcal{L}_1(\boldsymbol{x})|$, since

$$\widehat{\mathcal{L}}_1(\boldsymbol{x}) = \mathcal{D}_1(\boldsymbol{x}) \cup \mathcal{I}_1(\boldsymbol{x}) \cup \{\boldsymbol{x}\},$$
$$\widehat{\mathcal{L}}_2(\boldsymbol{x}) = \mathcal{L}_1(\boldsymbol{x}) \cup \mathcal{D}_2(\boldsymbol{x}) \cup \mathcal{I}_2(\boldsymbol{x}) \cup \mathcal{D}_1(\boldsymbol{x}) \cup \mathcal{I}_1(\boldsymbol{x}),$$

and the length of the sequences in each ball is different which implies that the sets in these unions are disjoint. However, not much is known about the size of the Levenshtein ball and the FLL ball for arbitrary n,t and $\boldsymbol{x} \in \mathbb{Z}_q^n$.

For $\boldsymbol{x} \in \mathbb{Z}_q^*$, let $|\boldsymbol{x}|$ denote the length of $\boldsymbol{x}$ and for a set of indices $I \subseteq [|\boldsymbol{x}|]$, let $\boldsymbol{x}_I$ denote the *projection* of $\boldsymbol{x}$ on the ordered indices of I, which is the subsequence of $\boldsymbol{x}$ received by the symbols in the entries of I. For a symbol $\sigma \in \mathbb{Z}_m$, σ^n denotes the sequence with n consecutive σ's.

A word $\boldsymbol{x}$ is called a *common supersequence* (*subsequence*) of some sequences $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_\tau$ if $\boldsymbol{x}$ is a supersequence (subsequence) of each one of these t words. The set of all shortest common supersequences of $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_\tau \in \mathbb{Z}_q^*$ is denoted by $\mathcal{SCS}(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_\tau)$ and $\mathsf{SCS}(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_\tau)$ is the *length of the shortest common supersequence* (*SCS*) of $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_\tau$, that is,

$$\mathsf{SCS}(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_\tau) = \min_{\boldsymbol{x} \in \mathcal{SCS}(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_\tau)} \{|\boldsymbol{x}|\}.$$

Similarly, $\mathcal{LCS}(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_\tau)$ is the set of all longest common subsequences of $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_\tau$ and $\mathsf{LCS}(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_\tau)$ is the *length of the longest common subsequence* (*LCS*) of $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_\tau$, that is,

$$\mathsf{LCS}(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_\tau) \triangleq \max_{\boldsymbol{x} \in \mathcal{LCS}(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_\tau)} \{|\boldsymbol{x}|\}.$$

This definition implies the following well known property.

Claim 5: For $\boldsymbol{x}_1,\boldsymbol{x}_2 \in \mathbb{Z}_q^n$, $\mathcal{D}_t(\boldsymbol{x}_1) \cap \mathcal{D}_t(\boldsymbol{x}_2) = \varnothing$ if and only if $\mathsf{LCS}(\boldsymbol{x}_1,\boldsymbol{x}_2) < n-t$.

Combining (4) and Claim 5 implies the following result.

Corollary 1: If $\boldsymbol{x}_1,\boldsymbol{x}_2 \in \mathbb{Z}_q^n$ then

$$\mathsf{LCS}(\boldsymbol{x}_1,\boldsymbol{x}_2) = n - d_\ell(\boldsymbol{x}_1,\boldsymbol{x}_2).$$

For two sequences $\boldsymbol{x} \in \mathbb{Z}_q^n$ and $\boldsymbol{y} \in \mathbb{Z}_q^m$, the value of $\mathsf{LCS}(\boldsymbol{x},\boldsymbol{y})$ is given by the following recursive formula [20]

$$\mathsf{LCS}(\boldsymbol{x},\boldsymbol{y}) = \begin{cases} 0 & n=0 \ \text{ or } m=0 \\ 1+\mathsf{LCS}(\boldsymbol{x}_{[1,n-1]},\boldsymbol{y}_{[1,m-1]}) & x_n = y_m \\ \max\left\{\mathsf{LCS}(\boldsymbol{x}_{[1,n-1]},\boldsymbol{y}),\mathsf{LCS}(\boldsymbol{x},\boldsymbol{y}_{[1,m-1]})\right\} & \text{otherwise.} \end{cases} \quad (6)$$

A subset $\mathcal{C} \subseteq \mathbb{Z}_q^n$ is a *t-deletion-correcting code* (*t-insertion-correcting code*, respectively) if for any two distinct codewords $\boldsymbol{c},\boldsymbol{c}' \in \mathcal{C}$ we have that $\mathcal{D}_t(\boldsymbol{c}) \cap \mathcal{D}_t(\boldsymbol{c}') = \varnothing$ ($\mathcal{I}_t(\boldsymbol{c}) \cap \mathcal{I}_t(\boldsymbol{c}') = \varnothing$, respectively). Similarly, $\mathcal{C}$ is called a *(t_1,t_2)-deletion-insertion-correcting code* if for any two distinct codewords $\boldsymbol{c}$ and $\boldsymbol{c}'$ of the code $\mathcal{C}$, we have that $\mathcal{DI}_{t_1,t_2}(\boldsymbol{c}) \cap \mathcal{DI}_{t_1,t_2}(\boldsymbol{c}') = \varnothing$, where $\mathcal{DI}_{t_1,t_2}(\boldsymbol{x})$ is the set of all words that can be obtained from $\boldsymbol{x}$ by t_1 deletions and t_2 insertions. Levenshtein [22] proved that $\mathcal{C}$ is a t-deletion-correcting code if and only if $\mathcal{C}$ is a t-insertion-correcting code and if and only if $\mathcal{C}$ is a (t_1,t_2)-deletion-insertion-correcting code for every t_1,t_2 such that $t_1+t_2 \leq t$. A straightforward generalization is the following result [11].

Lemma 1: For all $t_1,t_2 \in \mathbb{N}$, if $\mathcal{C} \subseteq \mathbb{Z}_q^n$ is a (t_1,t_2)-deletion-insertion-correcting code, then $\mathcal{C}$ is also a (t_1+t_2)-deletion-correcting code.

Corollary 2: For $\mathcal{C} \subseteq \mathbb{Z}_q^n$, the following statements are equivalent.

1) $\mathcal{C}$ is a (t_1, t_2)-deletion-insertion-correcting code.
2) $\mathcal{C}$ is a $(t_1 + t_2)$-deletion-correcting code.
3) $\mathcal{C}$ is a $(t_1 + t_2)$-insertion-correcting code.
4) $\mathcal{C}$ is a (t'_1, t'_2)-deletion-insertion-correcting code for any t'_1, t'_2 such that $t'_1 + t'_2 = t_1 + t_2$.

We further extend this result in the next lemma.

Lemma 2: A code $\mathcal{C} \in \mathbb{Z}_q^n$ is a $(2t+1)$-deletion-correcting code if and only if the following two conditions are satisfied

- $\mathcal{C}$ is a (t, t)-deletion-insertion-correcting code

and also

- if exactly $t+1$ FLL errors (i.e., $t+1$ insertions and $t+1$ deletions) occurred, then $\mathcal{C}$ can detect these $t+1$ FLL errors.

Proof: If $\mathcal{C}$ is a $(2t+1)$-deletion-correcting code, then by definition for any $\boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathcal{C}$ we have that

$$\mathcal{D}_{2t+1}(\boldsymbol{c}_1) \cap \mathcal{D}_{2t+1}(\boldsymbol{c}_2) = \varnothing.$$

Therefore, by Claim 5 for any two distinct codewords $\boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathcal{C}$ we have that

$$\mathsf{LCS}(\boldsymbol{c}_1, \boldsymbol{c}_2) \leq n - (2t+1).$$

Hence, by Corollary 1, $d_\ell(\boldsymbol{c}_1, \boldsymbol{c}_2) \geq 2(t+1)$. Since the FLL metric is graphic, it follows that $\mathcal{C}$ can correct up to t FLL errors and if exactly $t+1$ FLL errors occurred it can detect them.

For the other direction, assume that $\mathcal{C}$ is a (t,t)-deletion-insertion-correcting code and if exactly $t+1$ FLL errors occurred, then $\mathcal{C}$ can detect them. By Lemma 1, $\mathcal{C}$ is a $(2t)$-deletion-correcting code which implies that $\mathcal{D}_{2t}(\boldsymbol{c}_1) \cap \mathcal{D}_{2t}(\boldsymbol{c}_2) = \varnothing$ for all $\boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathcal{C}$, and hence by (4) we have that

$$\forall \boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathcal{C}: \quad d_\ell(\boldsymbol{c}_1, \boldsymbol{c}_2) > 2t.$$

Let us assume to the contrary that there exist two codewords $\boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathcal{C}$ such that $d_\ell(\boldsymbol{c}_1, \boldsymbol{c}_2) = 2t+1$. Since the FLL metric is a graphic metric, it follows that there exists a word $\boldsymbol{y} \in \mathbb{Z}_q^n$ such that $d_\ell(\boldsymbol{c}_1, \boldsymbol{y}) = t$ and $d_\ell(\boldsymbol{y}, \boldsymbol{c}_2) = t+1$. Hence, if the received word is $\boldsymbol{y}$, then the submitted codeword can be either $\boldsymbol{c}_1$ (t errors) or $\boldsymbol{c}_2$ ($t+1$ errors) which contradicts the fact that in $\mathcal{C}$ up to t FLL errors can be corrected and exactly $t+1$ FLL errors can be detected. Hence,

$$\forall \boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathcal{C}: \quad d_\ell(\boldsymbol{c}_1, \boldsymbol{c}_2) > 2t+1,$$

and by definition, $\mathcal{C}$ can correct $2t+1$ deletions. ■

III. The Minimum Size of an FLL Ball

In this section, the explicit expression for the minimum size of an FLL t-ball of any radius t is derived. Although this result is rather simple and straightforward, it is presented here for the completeness of the problems studied in the paper. Since changing the symbol in the i-th position from σ to σ' in any sequence $\boldsymbol{x}$ can be done by first deleting σ in the i-th position of $\boldsymbol{x}$ and then inserting σ' in the same position of $\boldsymbol{x}$, it follows that

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}_q^n: \quad d_H(\boldsymbol{x}, \boldsymbol{y}) \geq d_\ell(\boldsymbol{x}, \boldsymbol{y}).$$

Since $\boldsymbol{y} \in \mathcal{H}_t(\boldsymbol{x})$ if and only if $d_H(\boldsymbol{x}, \boldsymbol{y}) \leq t$ and $\boldsymbol{y} \in \mathcal{L}_t(\boldsymbol{x})$ if and only if $d_\ell(\boldsymbol{x}, \boldsymbol{y}) \leq t$, the following results are immediately implied.

Lemma 3: If $n \geq t \geq 0$ are integers and $\boldsymbol{x} \in \mathbb{Z}_q^n$, then $\mathcal{H}_t(\boldsymbol{x}) \subseteq \mathcal{L}_t(\boldsymbol{x})$.

Corollary 3: For any two integers $n \geq t \geq 0$ and any sequence $\boldsymbol{x} \in \mathbb{Z}_q^n$, $|\mathcal{H}_t(\boldsymbol{x})| \leq |\mathcal{L}_t(\boldsymbol{x})|$.

Lemma 4: If $n > t \geq 0$ are integers, then $\mathcal{H}_t(\boldsymbol{x}) = \mathcal{L}_t(\boldsymbol{x})$ if and only if $\boldsymbol{x} = \sigma^n$ for $\sigma \in \mathbb{Z}_q$.

Proof: Assume first w.l.o.g. that $\boldsymbol{x} = 0^n$ and let $\boldsymbol{y} \in \mathcal{L}_t(\boldsymbol{x})$ be a sequence obtained from $\boldsymbol{x}$ by at most t insertions and t deletions. Hence, $\text{wt}(\boldsymbol{y}) \leq t$ and $\boldsymbol{y} \in \mathcal{H}_t(\boldsymbol{x})$, which implies that $\mathcal{L}_t(\boldsymbol{x}) \subseteq \mathcal{H}_t(\boldsymbol{x})$. Therefore, Lemma 3 implies that $\mathcal{H}_t(\boldsymbol{x}) = \mathcal{L}_t(\boldsymbol{x})$.

For the other direction, assume that $\mathcal{H}_t(\boldsymbol{x}) = \mathcal{L}_t(\boldsymbol{x})$ and let $\boldsymbol{x} \in \mathbb{Z}_q^n$ where $\boldsymbol{x} \neq \sigma^n$ for all $\sigma \in \mathbb{Z}_q$. Since by Lemma 3, $\mathcal{H}_t(\boldsymbol{x}) \subseteq \mathcal{L}_t(\boldsymbol{x})$, to complete the proof, it is sufficient to show that there exists a sequence $\boldsymbol{y} \in \mathcal{L}_t(\boldsymbol{x}) \backslash \mathcal{H}_t(\boldsymbol{x})$. Denote $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and let i be the smallest index for which $x_i \neq x_{i+1}$. Let $\boldsymbol{y}$ be the sequence defined by

$$\boldsymbol{y} \triangleq (y_1, y_2, \ldots, y_{i-1}, x_{i+1}, x_i, y_{i+2}, \ldots, y_n),$$

where $y_j \neq x_j$ for the first $t-1$ indices (for which $j \notin \{i, i+1\}$) and $y_j = x_j$ otherwise. Clearly, $\boldsymbol{y}$ differs from $\boldsymbol{x}$ in $t+1$ indices and therefore $\boldsymbol{y} \notin \mathcal{H}_t(\boldsymbol{x})$. On the other hand, $\boldsymbol{y}$ can be obtained from $\boldsymbol{x}$ by first deleting x_i and inserting it to the right of x_{i+1} and then applying $t-1$ deletions and $t-1$ insertions whenever $y_j \neq x_j$ (where $j \notin \{i, i+1\}$). Thus, $\boldsymbol{y} \in \mathcal{L}_t(\boldsymbol{x}) \backslash \mathcal{H}_t(\boldsymbol{x})$ which completes the proof. ■

The following simple corollary is a direct result of Corollary 3, Lemma 4 and (1).

Corollary 4: If $n > t \geq 0$ are integers, then the size of the minimum FLL t-ball is

$$\min_{\boldsymbol{x} \in \mathbb{Z}_q^n} |\mathcal{L}_t(\boldsymbol{x})| = \sum_{i=0}^{t} \binom{n}{i} (q-1)^i,$$

and the minimum is obtained only by the balls centered at $\boldsymbol{x} = \sigma^n$ for any $\sigma \in \mathbb{Z}_q$.

IV. The Maximum FLL Balls With Radius One

The goal of this section is to compute the size of a ball with maximum size and its centre. For this purpose it is required first to compute the size of a ball. The size of the FLL 1-ball centered at $\boldsymbol{x} \in \mathbb{Z}_q^n$ was proved in [30] and given in (5). In the analysis of the maximum ball we distinguish between the binary case and the non-binary case. Surprisingly, the computation of the non-binary case is not a generalization of the binary case. That is, the binary case is not a special case of the non-binary case. Even more surprising is that the analysis of the non-binary case is much simpler than the analysis of the binary case. Hence, we start with the analysis of the non-binary case which is relatively simple.

A. The Non-Binary Case

By (5), the size of a ball with radius one centered at $\boldsymbol{x}$ depends on $\rho(\boldsymbol{x})$, the number of runs in $\boldsymbol{x}$. For a given number of runs $1 \leq r \leq n$, the size of a ball depends on the lengths of the maximal alternating segments in $\boldsymbol{x}$. The following lemma is an immediate consequence of (5).

Lemma 5: If $n > 0$ and $1 \le r \le n$, then

$$\underset{\substack{\boldsymbol{x}\in\mathbb{Z}_q^n \\ \rho(\boldsymbol{x})=r}}{\arg\max}|\mathcal{L}_1(\boldsymbol{x})| = \underset{\substack{\boldsymbol{x}\in\mathbb{Z}_q^n \\ \rho(\boldsymbol{x})=r}}{\arg\min}\left\{\sum_{i=1}^{A(\boldsymbol{x})}\frac{(s_i-1)(s_i-2)}{2}\right\}.$$

Proof: Let $\boldsymbol{x} \in \mathbb{Z}_q^n$ be a sequence with exactly r runs. Since $r(n(q-1)-1)+2$ is a constant and

$$\sum_{i=1}^{A(\boldsymbol{x})}\frac{(s_i-1)(s_i-2)}{2} \ge 0,$$

the claim follows immediately from (5). ■

Corollary 5: If $n > 0$ and $1 \le r \le n$, then

$$\max_{\substack{\boldsymbol{x}\in\mathbb{Z}_q^n \\ \rho(\boldsymbol{x})=r}} |\mathcal{L}_1(\boldsymbol{x})| = r(n(q-1)-1) + 2 - \min_{\substack{\boldsymbol{x}\in\mathbb{Z}_q^n \\ \rho(\boldsymbol{x})=r}}\left\{\sum_{i=1}^{A(\boldsymbol{x})}\frac{(s_i-1)(s_i-2)}{2}\right\}.$$

Note that

$$\sum_{i=1}^{A(\boldsymbol{x})}\frac{(s_i-1)(s_i-2)}{2} = 0 \iff \forall 1 \le i \le A(\boldsymbol{x}) : s_i \in \{1,2\}. \tag{7}$$

The following claim is a straightforward result from the definitions of a run and an alternating segment.

Lemma 6: Let $n > 0$ and let $\boldsymbol{x} \in \mathbb{Z}_q^n$. For $1 \le i \le \rho(\boldsymbol{x})$, denote by r_i the length of the i-th run and by $\sigma_i \in \mathbb{Z}_q$ the symbol of the i-th run. Then all the maximal alternating segments of $\boldsymbol{x}$ have lengths at most two ($s_i \le 2$ for each i) if and only if for each $1 \le i \le \rho(\boldsymbol{x})-2$, $\sigma_i \ne \sigma_{i+2}$ or $r_{i+1} > 1$.

The maximum value of $|\mathcal{L}_1(\boldsymbol{x})|$ for non-binary alphabet was given in [31] without a proof. For $q = 2$ the value of $|\mathcal{L}_1(\boldsymbol{x})|$ given in [31] without a proof is not accurate and we will give the exact value with a complete proof.

Theorem 6: For $q > 2$, the maximum FLL 1-balls are the balls centered at $\boldsymbol{x} \in \mathbb{Z}_q^n$, such that the number of runs in $\boldsymbol{x}$ is n (i.e., any two consecutive symbols are different) and $x_i \ne x_{i+2}$ for all $1 \le i \le n-2$. In addition, the maximum size of an FLL 1-ball is,

$$\max_{\boldsymbol{x}\in\mathbb{Z}_q^n}|\mathcal{L}_1(\boldsymbol{x})| = n^2(q-1)-n+2.$$

Proof: Corollary 5 implies that

$$\max_{\boldsymbol{x}\in\mathbb{Z}_q^n}|\mathcal{L}_1(\boldsymbol{x})| = \max_{r\in\{1,\dots,n\}}\left\{\max_{\substack{\boldsymbol{x}\in\mathbb{Z}_q^n \\ \rho(\boldsymbol{x})=r}}|\mathcal{L}_1(\boldsymbol{x})|\right\}$$
$$= \max_{r\in\{1,\dots,n\}}\left\{r(n(q-1)-1)+2-\min_{\substack{\boldsymbol{x}\in\mathbb{Z}_q^n \\ \rho(\boldsymbol{x})=r}}\left\{\sum_{i=1}^{A(\boldsymbol{x})}\frac{(s_i-1)(s_i-2)}{2}\right\}\right\}.$$

Clearly, $r(n(q-1)-1)+2$ is maximized for $r = n$ and therefore, using (7), we conclude that $\max_{\boldsymbol{x}\in\mathbb{Z}_q^n}|\mathcal{L}_1(\boldsymbol{x})|$ can be obtained for each $\boldsymbol{x} \in \mathbb{Z}_q^n$ such that $\rho(\boldsymbol{x}) = n$ and $s_i \le 2$ for each i. Note that $\sigma_i = x_i$ since $r = n$. By Lemma 6, it implies that $x_i \ne x_{i+2}$ or $r_{i+1} > 1$ for each $1 \le i \le n-2$. Since $q > 2$, it follows that there exists such an assignment for the symbols of each run such that $x_i \ne x_{i+2}$ for each $1 \le i \le r-2$. It follows that

$$\max_{\boldsymbol{x}\in\mathbb{Z}_q^n}|\mathcal{L}_1(\boldsymbol{x})| = n^2(q-1)-n+2.$$

■

B. The Binary Case

The analysis to find the maximum ball for binary sequences is more difficult, since by definition of a run, there is no sequence $\boldsymbol{x}$ with n runs such that $x_i \ne x_{i+2}$ (see Theorem 6) for some i. Note also that since in the binary case two maximal alternating segments can not overlap it holds that $\sum_{i=1}^{A(\boldsymbol{x})} s_i = n$ for any binary sequence $\boldsymbol{x}$.

For a sequence $\boldsymbol{x} \in \mathbb{Z}_2^n$, the *alternating segments profile* of $\boldsymbol{x}$ is $(s_1, s_2, \dots, s_{A(\boldsymbol{x})})$. Note that each alternating segments profile defines exactly two binary sequences.

Lemma 7: If $\boldsymbol{x} \in \mathbb{Z}_2^n$ then $\rho(\boldsymbol{x}) = n + 1 - A(\boldsymbol{x})$.

Proof: Let $\boldsymbol{x} \in \mathbb{Z}_2^n$ be a sequence and let $\boldsymbol{x}_{[i,j]}$ and $\boldsymbol{x}_{[i',j']}$ be two consecutive maximal alternating segments such that $i < i'$. Since $\boldsymbol{x}$ is a binary sequence, it follows that two maximal alternating segments cannot overlap, and hence $i' = j+1$. Now, let $\alpha = A(\boldsymbol{x})$ and we continue to prove the claim of the lemma by induction on α for any given $n \ge 1$. For $\alpha = 1$, there is one maximal alternating segment whose length is clearly n which consists of alternating symbols, i.e., there are $\rho(\boldsymbol{x}) = n$ runs as required. Assume the claim holds for any α' such that $1 \le \alpha' < \alpha$ and let $\boldsymbol{x} \in \mathbb{Z}_2^n$ be a sequence with exactly α maximal alternating segments. Denote by $\boldsymbol{x}'$ the sequence that is obtained from $\boldsymbol{x}$ by deleting its last maximal alternating segment $\boldsymbol{x}''$. By the induction hypothesis

$$\rho(\boldsymbol{x}') = (n - s_\alpha) + 1 - (\alpha - 1) = n + 2 - s_\alpha - t,$$

where s_α is the length of $\boldsymbol{x}''$. Clearly, the first symbol of $\boldsymbol{x}''$ is equal to the last symbol in $\boldsymbol{x}'$. Thus,

$$\begin{aligned}\rho(\boldsymbol{x}) &= \rho(\boldsymbol{x}'\boldsymbol{x}'') = \rho(\boldsymbol{x}') + s_\alpha - 1 \\ &= n + 2 - s_\alpha - \alpha + s_\alpha - 1 = n + 1 - \alpha.\end{aligned}$$

■

Notice that $\rho(\boldsymbol{x}) = n + 1 - A(\boldsymbol{x})$ does not hold for alphabet size $q > 2$. To clarify, consider the sequences $\boldsymbol{x}_1 = 0120$, $\boldsymbol{x}_2 = 0101$ and $\boldsymbol{x}_3 = 0102$, each of the sequences has four runs even though they differ in the number of maximal alternating segments; $A(\boldsymbol{x}_1) = 3$, $A(\boldsymbol{x}_2) = 1$ and $A(\boldsymbol{x}_3) = 2$.

Definition 7: For a positive integer α, $\boldsymbol{x}^{(\alpha)} \in \mathbb{Z}_2^n$ is an α-***balanced sequence*** if $A(\boldsymbol{x}) = \alpha$ and $s_i \in \{\lceil \frac{n}{\alpha} \rceil, \lceil \frac{n}{\alpha} \rceil - 1\}$ for all $i \in \{1, \dots, \alpha\}$.

Lemma 8: If n is a positive integer and $\alpha \in \{1, \dots, n\}$ then

$$\underset{\substack{\boldsymbol{x}\in\mathbb{Z}_2^n \\ A(\boldsymbol{x})=\alpha}}{\arg\max}|\mathcal{L}_1(\boldsymbol{x})| = \{\boldsymbol{x} \in \mathbb{Z}_2^n : \boldsymbol{x} \text{ is an } \alpha\text{-balanced sequence}\}.$$

Proof: For a sequence $\boldsymbol{x} \in \mathbb{Z}_2^n$ such that $A(\boldsymbol{x}) = \alpha$, Lemma 7 implies that $\rho(\boldsymbol{x}) = n+1-\alpha$. Hence, by Lemma 5,

$$\begin{aligned}\underset{\substack{\boldsymbol{x}\in\mathbb{Z}_2^n\\ A(\boldsymbol{x})=\alpha}}{\arg\max}|\mathcal{L}_1(\boldsymbol{x})| &= \underset{\substack{\boldsymbol{x}\in\mathbb{Z}_2^n\\ A(\boldsymbol{x})=\alpha}}{\arg\min}\sum_{i=1}^{\alpha}\frac{(s_i-1)(s_i-2)}{2}\\ &= \underset{\substack{\boldsymbol{x}\in\mathbb{Z}_2^n\\ A(\boldsymbol{x})=\alpha}}{\arg\min}\sum_{i=1}^{\alpha}(s_i^2-3s_i+2)\\ &= \underset{\substack{\boldsymbol{x}\in\mathbb{Z}_2^n\\ A(\boldsymbol{x})=\alpha}}{\arg\min}\left(\sum_{i=1}^{\alpha}s_i^2-3\sum_{i=1}^{\alpha}s_i+2\alpha\right)\\ &\overset{(a)}{=} \underset{\substack{\boldsymbol{x}\in\mathbb{Z}_2^n\\ A(\boldsymbol{x})=\alpha}}{\arg\min}\left(\sum_{i=1}^{\alpha}s_i^2-3n+2\alpha\right)\\ &= \underset{\substack{\boldsymbol{x}\in\mathbb{Z}_2^n\\ A(\boldsymbol{x})=\alpha}}{\arg\min}\sum_{i=1}^{\alpha}s_i^2,\end{aligned}$$

where (a) holds since alternating segments cannot overlap for binary sequences and therefore $\sum_{i=1}^{\alpha} s_i = n$.

Assume $\boldsymbol{x} \in \mathbb{Z}_2^n$ is a sequence such that $A(\boldsymbol{x}) = \alpha$, $(s_1,\ldots,s_\alpha)$ is the alternating segments profile of $\boldsymbol{x}$ and $\sum_{i=1}^{\alpha} s_i^2$ is minimal among all sequences in $\mathbb{Z}_2^n$. Assume to the contrary that $\boldsymbol{x}$ is not an α-balanced sequence. Then there exist indices $i \neq j$ such that $s_i \le \lceil\frac{n}{\alpha}\rceil - 1$ and $s_j > \lceil\frac{n}{\alpha}\rceil$ or there exist indices $i \neq j$ such that $s_i < \lceil\frac{n}{\alpha}\rceil - 1$ and $s_j \ge \lceil\frac{n}{\alpha}\rceil$. Consider a sequence $\boldsymbol{x}'$ with the alternating segments profile $(\nu_1,\ldots,\nu_\alpha)$ where

$$\nu_k = \begin{cases} s_i+1 & \text{if } k=i\\ s_j-1 & \text{if } k=j\\ s_k & \text{otherwise.}\end{cases}$$

Therefore,

$$\begin{aligned}\sum_{k=1}^{\alpha}\nu_k^2-\sum_{k=1}^{\alpha}s_k^2 &= \sum_{k=1}^{\alpha}\left(\nu_k^2-s_k^2\right) = (\nu_i^2-s_i^2)+(\nu_j^2-s_j^2)\\ &= \left((s_i+1)^2-s_i^2\right)+\left((s_j-1)^2-s_j^2\right)\\ &= \left(s_i^2+2s_i+1-s_i^2\right)+\left(s_j^2-2s_j+1-s_j^2\right)\\ &= 2(s_i-s_j+1)\\ &< 2\left(\left\lceil\frac{n}{\alpha}\right\rceil-1-\left\lceil\frac{n}{\alpha}\right\rceil+1\right)=0,\end{aligned}$$

and hence $\sum_{k=1}^{\alpha}\nu_k^2 < \sum_{k=1}^{\alpha}s_k^2$. This implies that if $\boldsymbol{x}$ is not an α-balanced sequence, then $\sum_{k=1}^{\alpha}s_k^2$ is not minimal, a contradiction. Thus,

$$\begin{aligned}\underset{\substack{\boldsymbol{x}\in\mathbb{Z}_2^n\\ A(\boldsymbol{x})=\alpha}}{\arg\max}|\mathcal{L}_1(\boldsymbol{x})| &= \underset{\substack{\boldsymbol{x}\in\mathbb{Z}_2^n\\ A(\boldsymbol{x})=\alpha}}{\arg\min}\sum_{i=1}^{\alpha}s_i^2\\ &= \left\{\boldsymbol{x}\in\mathbb{Z}_2^n : \boldsymbol{x} \text{ is an } \alpha\text{-balanced sequence}\right\}.\end{aligned}$$

■

Lemma 9: Let $\boldsymbol{x}^{(\alpha)}$ be an α-balanced sequence of length n. Then,

$$\begin{aligned}\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right| &= (n+1-\alpha)(n-1)+2\\ &\quad -\frac{k}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-1\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\\ &\quad -\frac{\alpha-k}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-3\right),\end{aligned}$$

where $k \equiv n \pmod{\alpha}$ and $1 \le k \le \alpha$.

Proof: By (5) we have that

$$\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right| = \rho\left(\boldsymbol{x}^{(\alpha)}\right)\cdot(n-1)+2-\sum_{i=1}^{\alpha}\frac{(s_i-1)(s_i-2)}{2}, \tag{8}$$

and Lemma 7 implies that $\rho\left(\boldsymbol{x}^{(\alpha)}\right) = n+1-\alpha$. Let k be the number of entries in the alternating segments profile of $\boldsymbol{x}^{(\alpha)}$ such that $s_i = \lceil\frac{n}{\alpha}\rceil$. Note further that $\sum_{i=1}^{\alpha} s_i = n$ and $s_i \in \{\lceil\frac{n}{\alpha}\rceil, \lceil\frac{n}{\alpha}\rceil - 1\}$ for $1 \le i \le \alpha$. Hence,

$$k\left\lceil\frac{n}{\alpha}\right\rceil+(\alpha-k)\left(\left\lceil\frac{n}{\alpha}\right\rceil-1\right)=n,$$

which is equivalent to

$$k = n-\alpha\left(\left\lceil\frac{n}{\alpha}\right\rceil-1\right).$$

Therefore, k is the value between 1 to α such that $k \equiv n \pmod{\alpha}$. Thus, by (8) we have that

$$\begin{aligned}\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right| &= (n+1-\alpha)(n-1)+2\\ &\quad -\frac{k}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-1\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\\ &\quad -\frac{\alpha-k}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-3\right).\end{aligned}$$

■

By Lemma 8 we have that

$$\begin{aligned}\max_{x\in\mathbb{Z}_2^n}|\mathcal{L}_1(\boldsymbol{x})| &= \max_{1\le\alpha\le n}\left\{\max_{\substack{\boldsymbol{x}\in\mathbb{Z}_2^n\\ A(\boldsymbol{x})=\alpha}}|\mathcal{L}_1(\boldsymbol{x})|\right\}\\ &= \max_{1\le\alpha\le n}\left\{\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right|\right\},\end{aligned}$$

and the size $\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right|$ for $1 \le \alpha \le n$ is given in Lemma 9. Hence, our goal is to find the set

$$\mathsf{A} \triangleq \underset{1\le\alpha\le n}{\arg\max}\left\{\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right|\right\},$$

i.e., for which values of α the maximum of $|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)|$ is obtained. The answer for this question is given in the following lemma whose proof can be found in the Appendix.

Lemma 10: Let $\boldsymbol{x}^{(\alpha)}$ be an α-balanced sequence of length $n > 1$. Then,

$$\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right| > \left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha-1)}\right)\right|$$

if and only if $n > 2(\alpha-1)\alpha$.

Theorem 8: If n is an integer, then

$$\mathsf{A}=\arg\min_{\alpha\in\mathbb{N}}\left\{\left|\alpha-\frac{1}{2}\sqrt{1+2n}\right|\right\},$$

and the maximum FLL 1-balls are the balls centered at the α-balanced sequences of length n, for $\alpha\in\mathsf{A}$. In addition, the size of the maximum FLL 1-balls is given by

$$\begin{aligned}\max_{\boldsymbol{x}\in\mathbb{Z}_2^n}\{|\mathcal{L}_1(\boldsymbol{x})|\}&=n^2-n\alpha+\alpha+1\\&-\frac{k}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-1\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\\&-\frac{\alpha-k}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-3\right),\end{aligned}$$

where $k\equiv n \pmod{\alpha}$ and $1\le k\le\alpha$.

Proof: Let n be a positive integer. By Lemma 8 we have that

$$\begin{aligned}\max_{\boldsymbol{x}\in\mathbb{Z}_2^n}|\mathcal{L}_1(\boldsymbol{x})|&=\max_{1\le\alpha\le n}\left\{\max_{\substack{\boldsymbol{x}\in\mathbb{Z}_2^n\\A(\boldsymbol{x})=\alpha}}|\mathcal{L}_1(\boldsymbol{x})|\right\}\\&=\max_{1\le\alpha\le n}\left\{\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right|\right\}.\end{aligned}$$

If there exists an integer α, $1\le\alpha\le n$ such that $n=2(\alpha-1)\alpha$, then by Lemma 9,

$$\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right|=\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha-1)}\right)\right|.$$

Additionally, by Lemma 10, for $n>2(\alpha-1)\alpha$ we have that

$$\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right|>\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha-1)}\right)\right|,$$

which implies that $\left|\mathcal{L}_1(\boldsymbol{x}^{(\alpha)})\right|$ is maximized for $\alpha\in\{1,\dots,n\}$ such that

$$2\alpha(\alpha+1)\ge n\ge 2(\alpha-1)\alpha. \tag{9}$$

To find α we have to solve the two quadratic equations from (9). The solution for α must satisfies both equations and hence $-\frac{1}{2}+\frac{\sqrt{1+2n}}{2}\le\alpha\le\frac{1}{2}+\frac{\sqrt{1+2n}}{2}$. Namely, for $\alpha\in\mathsf{A}$,

$$\max_{\boldsymbol{x}\in\mathbb{Z}_2^n}\{|\mathcal{L}_1(\boldsymbol{x})|\}=\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right|.$$

The size of $\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)$ was derived in Lemma 9, which completes the proof. ■

Corollary 6: If n is an integer, then

$$\max_{\boldsymbol{x}\in\mathbb{Z}_2^n}\{|\mathcal{L}_1(\boldsymbol{x})|\}=n^2-\sqrt{2}n^{\frac{3}{2}}+O(n).$$

Proof: By Theorem 8 we have that $\max_{\boldsymbol{x}\in\mathbb{Z}_2^n}\{|\mathcal{L}_1(\boldsymbol{x})|\}=\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right|$ for $\alpha=\left[\frac{1}{2}\sqrt{1+2n}\right]$. By Lemma 9 we have that

$$\begin{aligned}\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right|&=(n+1-\alpha)(n-1)+2\\&-\frac{k}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-1\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\\&-\frac{\alpha-k}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-3\right).\end{aligned}$$

Notice that

$$\frac{1}{2}\left(\sqrt{1+2n}-2\right)\le\alpha\le\frac{1}{2}\left(\sqrt{1+2n}+2\right)$$

and hence, $\alpha=\frac{\sqrt{1+2n}}{2}+\epsilon_1$, where $|\epsilon_1|\le 1$. Similarly,

$$\begin{aligned}\frac{2n}{\sqrt{1+2n}+2}&\le\left\lceil\frac{2n}{\sqrt{1+2n}+2}\right\rceil\le\left\lceil\frac{n}{\alpha}\right\rceil\\&\le\left\lceil\frac{2n}{\sqrt{1+2n}-2}\right\rceil\le\frac{2n}{\sqrt{1+2n}-2}+1.\end{aligned}$$

which implies that

$$\left\lceil\frac{n}{\alpha}\right\rceil=\frac{2n}{\sqrt{1+2n}}+\epsilon_2,$$

where by simple calculation we can find that $|\epsilon_2|\le 3$. Thus,

$$\begin{aligned}\max_{\boldsymbol{x}\in\mathbb{Z}_2^n}|\mathcal{L}_1(\boldsymbol{x})|&=(n+1-\alpha)(n-1)+2-\frac{k}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-1\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\\&-\frac{\alpha-k}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-3\right)\\&=(n+1-\alpha)(n-1)+2-\frac{\alpha}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-3\right)\\&-\frac{k}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-1-\left\lceil\frac{n}{\alpha}\right\rceil+3\right)\\&=(n+1-\alpha)(n-1)+2-k\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\\&-\frac{\alpha}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-3\right)\\&=(n+1-\frac{\sqrt{1+2n}}{2}-\epsilon_1)(n-1)+2\\&-k\left(\frac{2n}{\sqrt{1+2n}}+\epsilon_2-2\right)\\&-\frac{\sqrt{1+2n}+2\epsilon_1}{4}\left(\frac{2n}{\sqrt{1+2n}}+\epsilon_2-2\right)\left(\frac{2n}{\sqrt{1+2n}}+\epsilon_2-3\right)\\&=n^2+1-\left(\frac{\sqrt{1+2n}}{2}+\epsilon_1\right)(n-1)\\&-\left(\frac{2n}{\sqrt{1+2n}}+\epsilon_2-2\right)\left(k+\frac{\sqrt{1+2n}+2\epsilon_1}{4}\left(\frac{2n}{\sqrt{1+2n}}+\epsilon_2-3\right)\right).\end{aligned}$$

Note that $1\le k\le\alpha\le\frac{1}{2}\left(\sqrt{1+2n}+2\right)$, which implies that

$$\begin{aligned}\max_{\boldsymbol{x}\in\mathbb{Z}_2^n}|\mathcal{L}_1(\boldsymbol{x})|&=n^2-\frac{n\sqrt{1+2n}}{2}-\frac{n^2}{\sqrt{1+2n}}+O(n)\\&=n^2-\sqrt{2}n^{\frac{3}{2}}+O(n).\end{aligned}$$

■

V. The Expected Size of an FLL 1-Ball

Let n and $q>1$ be integers and let $\boldsymbol{x}\in\mathbb{Z}_q^n$. By (5), for every $\boldsymbol{x}\in\mathbb{Z}_q^n$, we have

$$\begin{aligned}|\mathcal{L}_1(\boldsymbol{x})|&=\rho(\boldsymbol{x})(n(q-1)-1)+2-\sum_{i=1}^{A(\boldsymbol{x})}\frac{(s_i-1)(s_i-2)}{2}\\&=\rho(\boldsymbol{x})(nq-n-1)+2-\frac{1}{2}\sum_{i=1}^{A(\boldsymbol{x})}s_i^2+\frac{3}{2}\sum_{i=1}^{A(\boldsymbol{x})}s_i-A(\boldsymbol{x}).\end{aligned}$$

Thus, the average size of an FLL 1-ball, $\mathbb{E}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[|\mathcal{L}_1(\boldsymbol{x})|\right]$, is

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[\rho(\boldsymbol{x})(n(q-1)-1)+2-\frac{1}{2}\sum_{i=1}^{A(\boldsymbol{x})}s_i^2+\frac{3}{2}\sum_{i=1}^{A(\boldsymbol{x})}s_i-A(\boldsymbol{x})\right]. \tag{10}$$

Based on (10) and the linearity of the expectation, in order to derive the expected size of an FLL 1-ball, in the following lemmas and claims we analyze the expected values of $\rho(\boldsymbol{x}), A(\boldsymbol{x}), \sum_{i=1}^{A(\boldsymbol{x})} s_i$, and $\sum_{i=1}^{A(\boldsymbol{x})} s_i^2$.

Lemma 11: For any two integers $n, q > 1$,

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[\sum_{i=1}^{A(\boldsymbol{x})} s_i\right] = n + (n-2)\cdot\frac{(q-1)(q-2)}{q^2}.$$

Proof: If $\boldsymbol{x} \in \mathbb{Z}_q^n$, then by the definition of an alternating segment, we have that for each $1 \le i \le n$, x_i is contained in at least one maximal alternating segment and not more than two maximal alternating segments. Hence,

$$\sum_{i=1}^{A(\boldsymbol{x})} s_i = n + \zeta(\boldsymbol{x}), \tag{11}$$

where $\zeta(\boldsymbol{x})$ denotes the number of entries in $\boldsymbol{x}$ which are contained in exactly two alternating segments. Define, for each $1 \le i \le n$

$$\zeta_i(\boldsymbol{x}) \triangleq \begin{cases} 1 & x_i \text{ is contained in two maximal alternating segments} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

Thus,

$$\begin{aligned}\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[\sum_{i=1}^{A(\boldsymbol{x})} s_i\right] &= n + \mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}[\zeta(\boldsymbol{x})] = n + \frac{1}{q^n}\sum_{\boldsymbol{x}\in\mathbb{Z}_q^n}\zeta(\boldsymbol{x}) \\ &= n + \frac{1}{q^n}\sum_{\boldsymbol{x}\in\mathbb{Z}_q^n}\sum_{i=1}^{n}\zeta_i(\boldsymbol{x}) \\ &= n + \frac{1}{q^n}\sum_{i=1}^{n}\sum_{x\in\mathbb{Z}_q^n}\zeta_i(\boldsymbol{x}).\end{aligned}$$

Clearly, if $i \in \{1, n\}$ then $\zeta_i(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathbb{Z}_q^n$. Otherwise, $\zeta_i(\boldsymbol{x}) = 1$ if and only if x_{i-1}, x_i and x_{i+1} are all different. Therefore, for $2 \le i \le n-1$, there are $\binom{q}{3}\cdot 3!$ distinct ways to select values for x_{i-1}, x_i, and x_{i+1} and q^{n-3} distinct ways to select values for the other entries of $\boldsymbol{x}$. That is,

$$\begin{aligned}\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[\sum_{i=1}^{A(\boldsymbol{x})} s_i\right] &= n + \frac{1}{q^n}\sum_{i=1}^{n}\sum_{\boldsymbol{x}\in\mathbb{Z}_q^n}\zeta_i(\boldsymbol{x}) \\ &= n + \frac{1}{q^n}\sum_{i=2}^{n-1}\binom{q}{3}3!q^{n-3} \\ &= n + (n-2)\cdot\frac{(q-1)(q-2)}{q^2}.\end{aligned}$$

■

Corollary 7: For $q = 2$, we have that

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_2^n}\left[\sum_{i=1}^{A(\boldsymbol{x})} s_i\right] = n.$$

Before we continue with the calculation of the other expected values, we define the difference vector, which will be used in the analysis to follow.

Definition 9: For a sequence $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{Z}_q^n$, denote by $\boldsymbol{x}' \in \mathbb{Z}_q^{n-1}$ the *difference vector* of $\boldsymbol{x}$, which is defined by

$$\boldsymbol{x}' \triangleq (x_2 - x_1, x_3 - x_2, \ldots, x_n - x_{n-1}).$$

Next, we study the expected size of $A(\boldsymbol{x})$. We start by presenting a relation between $A(\boldsymbol{x})$, $\sum_{i=1}^{A(\boldsymbol{x})} s_i$, and the difference vector of $\boldsymbol{x}$ in the next claim.

Claim 10: For integers n and $q > 1$ and a sequence $\boldsymbol{x} \in \mathbb{Z}_q^n$,

$$\sum_{i=1}^{A(\boldsymbol{x})} s_i = n + A(\boldsymbol{x}) - 1 - \mathsf{Zeros}(\boldsymbol{x}'),$$

where $\mathsf{Zeros}(\boldsymbol{y})$ denotes the number of zeros in $\boldsymbol{y}$.

Proof: By (11) we have that

$$\sum_{i=1}^{A(\boldsymbol{x})} s_i = n + \zeta(\boldsymbol{x}).$$

Since there are $A(\boldsymbol{x})$ alternating segments, it follows that there are $A(\boldsymbol{x})$ entries that start with a maximal alternating segment. Denote this set of entries by $\mathsf{Ind}(\boldsymbol{x})$ and let $\mathsf{Ind}_1(\boldsymbol{x}) \subseteq \mathsf{Ind}(\boldsymbol{x})$ be the set of entries $i \in \mathsf{Ind}(\boldsymbol{x})$ that are contained in exactly one maximal alternating segment. This implies that

$$\sum_{i=1}^{A(\boldsymbol{x})} s_i = n + |\mathsf{Ind}(\boldsymbol{x})| - |\mathsf{Ind}_1(\boldsymbol{x})|.$$

Clearly, $1 \in \mathsf{Ind}_1(\boldsymbol{x})$. For any other index $i \in \mathsf{Ind}(\boldsymbol{x})$, x_i is contained in exactly one maximal alternating segment if and only if $x_i = x_{i-1}$, i.e., $x'_{i-1} = 0$. Thus,

$$\sum_{i=1}^{A(\boldsymbol{x})} s_i = n + A(\boldsymbol{x}) - 1 - \mathsf{Zeros}(\boldsymbol{x}').$$

■

Using the latter relation, Lemma 11 and the linearity of expectation, the expected value of $A(\boldsymbol{x})$ can be derived from the expected value of $\mathsf{Zeros}(\boldsymbol{x}')$, which is given in the next claim.

Claim 11: Given two integers n and $q > 1$, we have that

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}[\mathsf{Zeros}(\boldsymbol{x}')] = \frac{n-1}{q}.$$

Proof: By the definition of the difference vector, given $\boldsymbol{y} \in \mathbb{Z}_q^{n-1}$, the sequence $\boldsymbol{x} \in \Sigma_q^n$ such that $\boldsymbol{x}' = \boldsymbol{y}$ is defined uniquely by the selection of the first entry of $\boldsymbol{x}$ from $\mathbb{Z}_q$. Hence, we have that for each $\boldsymbol{y} \in \mathbb{Z}_q^{n-1}$ there are exactly q sequences $\boldsymbol{x} \in \mathbb{Z}_q^n$ such that $\boldsymbol{x}' = \boldsymbol{y}$. In other words, the function $f(\boldsymbol{x}) = \boldsymbol{x}'$ is a q to 1 function. Define,

$$\mathsf{zero}_i(\boldsymbol{y}) \triangleq \begin{cases} 1 & \text{if } y_i = 0 \\ 0 & \text{otherwise.} \end{cases}$$

It follows that,

$$\begin{aligned}\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}[\mathsf{Zeros}(\boldsymbol{x}')] &= \mathop{\mathbb{E}}_{\boldsymbol{y}\in\mathbb{Z}_q^{n-1}}[\mathsf{Zeros}(\boldsymbol{y})]\\ &= \frac{1}{q^{n-1}}\sum_{\boldsymbol{y}\in\mathbb{Z}_q^{n-1}}\mathsf{Zeros}(\boldsymbol{y})\\ &= \frac{1}{q^{n-1}}\sum_{\boldsymbol{y}\in\mathbb{Z}_q^{n-1}}\sum_{i=1}^{n-1}\mathsf{zero}_i(\boldsymbol{y})\\ &= \frac{1}{q^{n-1}}\sum_{i=1}^{n-1}\sum_{\boldsymbol{y}\in\mathbb{Z}_q^{n-1}}\mathsf{zero}_i(\boldsymbol{y}).\end{aligned}$$

For each i, the set $\{\boldsymbol{y}\in\mathbb{Z}_q^{n-1} : y_i = 0\}$ is of size $\frac{q^{n-1}}{q} = q^{n-2}$. Thus,

$$\begin{aligned}\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}[\mathsf{Zeros}(\boldsymbol{x}')] &= \frac{1}{q^{n-1}}\sum_{i=1}^{n-1}\sum_{\boldsymbol{y}\in\mathbb{Z}_q^{n-1}}\mathsf{zero}_i(\boldsymbol{y})\\ &= \frac{1}{q^{n-1}}\cdot\sum_{i=1}^{n-1}q^{n-2} = \frac{n-1}{q}.\end{aligned}$$

■

By combining the results from Lemma 11 and Claims 10 and 11 we infer the following result.

Corollary 8: For two integers n and $q > 1$, the average number of alternating segments of a sequence $\boldsymbol{x}\in\mathbb{Z}_q^n$ is

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}[A(\boldsymbol{x})] = 1 + \frac{(n-2)(q-1)(q-2)}{q^2} + \frac{n-1}{q},$$

and in particular for $q = 2$

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_2^n}[A(\boldsymbol{x})] = \frac{n+1}{2}.$$

Proof: For each $q > 1$ we have that

$$\begin{aligned}\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}[A(\boldsymbol{x})] &= \mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[\sum_{i=1}^{A(\boldsymbol{x})}s_i\right] + \mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}[\mathsf{Zeros}(\boldsymbol{x}')] - n + 1\\ &\qquad\text{(by Claim 10)}\\ &= n + \frac{(n-2)(q-1)(q-2)}{q^2} + \frac{n-1}{q} - n + 1\\ &\qquad\text{(by Lemma 11 and Claim 11)}\\ &= 1 + \frac{(n-2)(q-1)(q-2)}{q^2} + \frac{n-1}{q}.\end{aligned}$$

When $q = 2$ the latter implies the claim. ■

The expected size of $\rho(\boldsymbol{x})$ is given in the next lemma.

Lemma 12: For any two integers n and $q > 1$, the average number of runs in a sequence $\boldsymbol{x}\in\mathbb{Z}_q^n$ is

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}[\rho(\boldsymbol{x})] = n - \frac{n-1}{q}.$$

Proof: For a sequence $\boldsymbol{x}\in\mathbb{Z}_q^n$, the number of runs in $\boldsymbol{x}$ is equal to the number of entries which begin a run in $\boldsymbol{x}$. Clearly, x_1 is the beginning of the first run and by the definition of the difference vector, we have that for each i, $2\le i\le n$, x_i starts a run if and only if $x'_{i-1}\neq 0$. Thus,

$$\rho(\boldsymbol{x}) = n - \mathsf{Zeros}(\boldsymbol{x}'),$$

and, by Claim 11,

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}[\rho(\boldsymbol{x})] = n - \mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}[\mathsf{Zeros}(\boldsymbol{x}')] = n - \frac{n-1}{q}.$$

■

Considering (10), our current goal is to evaluate $\mathbb{E}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[\sum_{i=1}^{A(\boldsymbol{x})}s_i^2\right]$. Denote by $\chi(s)$ the number of maximal alternating segments of length s over all the sequences $\boldsymbol{x}\in\mathbb{Z}_q^n$, i.e.,

$$\chi(s) = \sum_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left|\{1\le i\le A(\boldsymbol{x})\ \ s_i = s\}\right|.$$

It holds that

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[\sum_{i=1}^{A(\boldsymbol{x})}s_i^2\right] = \frac{1}{q^n}\sum_{\boldsymbol{x}\in\mathbb{Z}_2^n}\sum_{i=1}^{A(\boldsymbol{x})}s_i^2 = \frac{1}{q^n}\sum_{s=1}^{n}s^2\chi(s),$$

and the values of $\chi(s)$ for $1\le s\le n$ are given in the following lemmas.

Lemma 13: If n and $q > 1$ are two positive integers then

$$\chi(1) = 2q^{n-1} + (n-2)q^{n-2}.$$

Proof: Let us count the number of maximal alternating segments of length one over all the sequences $\boldsymbol{x}\in\mathbb{Z}_q^n$. Consider the following two cases:

Case 1 - If the alternating segment is at x_1, we can choose the symbols of x_1 in q different ways. Since the alternating segment's length is one, i.e., $x_1 = x_2$, it follows that the value of x_2 is determined. The symbols at $x_3,\ldots,x_n$ can be selected in q^{n-2} different ways. Therefore, there are q^{n-1} distinct sequences with such an alternating segment. The same arguments hold for an alternating segment at x_n.

Case 2 - If the alternating segment is at index i, where $2\le i\le n-1$, it must be that $x_{i-1} = x_i = x_{i+1}$. The symbol at x_i can be selected in q different ways and the symbols of x_{i-1}, x_{i+1} are fixed. In addition. we can set the symbols of $\boldsymbol{x}$ at indices $j\notin\{i-1,i,i+1\}$ in q^{n-3} different ways. Therefore, there are q^{n-2} distinct sequences with such an alternating segment.

Thus,

$$\chi(1) = 2q^{n-1} + (n-2)q^{n-2}.$$

■

Lemma 14: For any two integers n and $q > 1$,

$$\chi(n) = q(q-1).$$

Proof: Any alternating segment of length n is defined by the first two symbols which must be distinct (the rest of the symbols are determined by the first two symbols). There are $q(q-1)$ different ways to select the first two symbols and hence the claim follows. ■

For $2\le s\le n-1$ we need to consider whether the alternating segment overlaps with the preceding or the succeeding segment, or not. To this end, we distinguish between the maximal alternating segments of length s as follows

$\chi_1(s)$ - The number of alternating segments that do not overlap with the preceding segment and the succeeding segment.

$\chi_2(s)$ - The number of alternating segments that overlap with the preceding segment and the succeeding segment.

$\chi_3(s)$ - The number of alternating segments that overlap only with the succeeding segment.

$\chi_4(s)$ - The number of alternating segments that overlap only with the preceding segment.

Claim 12: If $n, q > 1$ are integers and $2 \leq s \leq n-1$ then,

1) $\chi_1(s) = 2(q-1)q^{n-s} + (n-s-1)(q-1)q^{n-s-1}$.
2) $\chi_2(s) = (n-s-1)(q-1)(q-2)^2 q^{n-s-1}$.
3) $\chi_3(s)=(q-1)(q-2)q^{n-s}+(q-1)(q-2)(n-s-1)q^{n-s-1}$.
4) $\chi_4(s)=(q-1)(q-2)q^{n-s}+(q-1)(q-2)(n-s-1)q^{n-s-1}$.

Proof:

1) To count the number of maximal alternating segments of length s that do not overlap with the preceding segment and the succeeding segment we distinguish two distinct cases.
Case 1 - If the alternating segment is at the beginning of the sequence, then there are $q(q-1)$ distinct ways to select the symbols of the segment. The symbol after the segment is determined (and is equal to the last symbol of the discussed alternating segment) in order to prevent an overlap and the other symbols can be chosen in q^{n-s-1} different ways. Hence, the number of different sequences with such segments is $(q-1)q^{n-s}$. The same arguments hold for an alternating segment at the end of the sequence.
Case 2 - If the alternating segment is not at the edges of the sequence, then there are $n-s-1$ possible positions to start the alternating segment, and $q(q-1)$ ways to choose the two symbols of the alternating segment. The symbol preceding and the symbol succeeding the alternating segment are determined. The other symbols can be chosen in q^{n-s-2} distinct ways and hence the number of different alternating segments is $(n-s-1)(q-1)q^{n-s-1}$. Thus,
$$\chi_1(s) = 2(q-1)q^{n-s} + (n-s-1)(q-1)q^{n-s-1}.$$
2) A maximal alternating segment that overlaps with the preceding segment and the succeeding segment can not be at the sequence edges. Hence, there are $n-s-1$ possible positions to start the alternating segment and the symbols of the segment can be chosen in $q(q-1)$ different ways. In order to overlap with the preceding (succeeding, respectively) segment, the symbol before (after, respectively) the segment must be different from the two symbols of the segment. Therefore, there are $(q-2)^2$ options to choose the symbol before and the symbol after the segment. In addition, the rest of the sequence can be chosen in q^{n-s-2} different ways and hence
$$\chi_2(s) = (n-s-1)(q-1)(q-2)^2 q^{n-s-1}.$$
3) Since the alternating segment must intersect with the succeeding segment, it can not be the last alternating segment, that is, the segment ends at index $j < n$. To count the number of maximal alternating segments of length s that overlap only with the succeeding segment we consider two distinct cases.
Case 1 - If the alternating segment is at the beginning of the sequence then there are $q(q-1)$ different ways to choose the symbols for it and the symbol after the segment must be different from the two symbols of the alternating segment so there are $(q-2)$ options to select it. The other symbols can be chosen in q^{n-s-1} different ways. Hence, the number of different segments is $(q-1)(q-2)q^{n-s}$.
Case 2 - If the alternating segment does not start at the beginning of the sequence, since the segment ends at index $j < n$, it follows that there are $(n-s-1)$ possible locations to start the segment. There are $q(q-1)$ different ways to select the symbols for the alternating segment. The symbol before the alternating segment is determined in order to prevent an overlap with the previous segment and the symbol after the segment must be different from the two symbols of the alternating segment and hence there are $(q-2)$ ways to choose it. The other symbols can be chosen in q^{n-s-2} different ways and hence the number of different segments is $q^{n-s-1}(q-1)(q-2)(n-s-1)$.
Thus,
$$\chi_3(s)=(q-1)(q-2)q^{n-s}+(q-1)(q-2)(n-s-1)q^{n-s-1}.$$
4) Clearly, the number of maximal alternating segments of length s that overlap only with the succeeding segment is equal to the number alternating segments of length s that overlap only with the preceding segment.

■

Lemma 15: In $n, q > 1$ are integers and $2 \leq s \leq n-1$ then
$$\chi(s) = 2(q-1)^2 q^{n-s} + (n-s-1)(q-1)^3 q^{n-s-1}.$$

Proof: By Claim 12,
$$\begin{aligned}\chi(s) &= \chi_1(s) + \chi_2(s) + \chi_3(s) + \chi_4(s)\\ &= 2(q-1)q^{n-s} + (n-s-1)(q-1)q^{n-s-1}\\ &\quad + (n-s-1)(q-1)(q-2)^2 q^{n-s-1}\\ &\quad + 2(q-1)(q-2)q^{n-s}\\ &\quad + 2(n-s-1)(q-1)(q-2)q^{n-s-1}\\ &= 2(q-1)^2 q^{n-s}\\ &\quad + (n-s-1)(q-1)q^{n-s-1}\left(1+(q-2)^2+2(q-2)\right)\\ &= 2(q-1)^2 q^{n-s}\\ &\quad + (n-s-1)(q-1)q^{n-s-1}\left(q^2-2q+1\right)\\ &= 2(q-1)^2 q^{n-s} + (n-s-1)(q-1)^3 q^{n-s-1}.\end{aligned}$$

■

Lemma 16: If $n, q > 1$ are integers then,

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[\sum_{i=1}^{A(\boldsymbol{x})} s_i^2\right] = \frac{n(4q^2-3q+2)}{q^2} + \frac{6q-4}{q^2} - 4 - \frac{2}{q-1}\left(1-\frac{1}{q^n}\right).$$

Proof: We have that

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[\sum_{i=1}^{A(\boldsymbol{x})} s_i^2\right] = \frac{1}{q^n}\sum_{\boldsymbol{x}\in\mathbb{Z}_q^n}\sum_{i=1}^{A(\boldsymbol{x})} s_i^2 = \frac{1}{q^n}\sum_{s=1}^{n} s^2\chi(s) = \frac{\chi(1)}{q^n} + \frac{n^2\chi(n)}{q^n} + \frac{1}{q^n}\sum_{s=2}^{n-1} s^2\chi(s).$$

Let us first calculate $\sum_{s=2}^{n-1} s^2\chi(s)$. By Lemma 15,

$$\begin{aligned}\sum_{s=2}^{n-1} s^2\chi(s) &= \sum_{s=2}^{n-1} s^2\left(2(q-1)^2q^{n-s} + (n-s-1)(q-1)^3q^{n-s-1}\right)\\ &= 2(q-1)^2\sum_{s=2}^{n-1} s^2q^{n-s} + (q-1)^3\sum_{s=2}^{n-1}(n-s-1)s^2q^{n-s-1}.\end{aligned}$$

It can be verified that

$$\sum_{s=2}^{n-1} s^2\chi(s) = \frac{2q^3 - q^3n^2(q-1)^2 + q^n(2-2q(3+q(2q-3)))}{(q-1)q^2} + \frac{q^n n(q-1)(1+q(4q-3))}{(q-1)q^2}$$

and after rearranging the latter, we obtain that

$$\sum_{s=2}^{n-1} s^2\chi(s) = nq^{n-2}(4q^2-3q+1) - n^2q(q-1) - 2q^{n-2}\cdot\frac{(2q-1)(q^2-q+1)}{(q-1)} + \frac{2}{q-1}.$$

Hence,

$$\begin{aligned}\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[\sum_{i=1}^{A(\boldsymbol{x})} s_i^2\right] &= \frac{\chi(1)}{q^n} + \frac{n^2\chi(n)}{q^n} + \frac{1}{q^n}\sum_{s=2}^{n-1} s^2\chi(s)\\ &= \frac{2q^{n-1}+(n-2)q^{n-2}}{q^n} + \frac{n^2q(q-1)}{q^n} + \frac{nq^{n-2}(4q^2-3q+1)}{q^n} - \frac{n^2q(q-1)}{q^n}\\ &\quad - 2q^{n-2}\cdot\frac{(2q-1)(q^2-q+1)}{q^n(q-1)} + \frac{2}{q^n(q-1)}\\ &= \frac{n(4q^2-3q+2)}{q^2} + \frac{2}{q} - \frac{2}{q^2} - \frac{2(2q-1)(q^2-q+1)}{q^2(q-1)} + \frac{2}{q^n(q-1)}\\ &= \frac{n(4q^2-3q+2)}{q^2} + \frac{6q-4}{q^2} - 4 - \frac{2}{q-1}\left(1-\frac{1}{q^n}\right).\end{aligned}$$

■

Theorem 13: If $n, q > 1$ are integers, then

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[|\mathcal{L}_1(\boldsymbol{x})|\right] = n^2\left(q+\frac{1}{q}-2\right) - \frac{n}{q} - \frac{(q-1)(q-2)}{q^2} + 3 - \frac{3}{q} + \frac{2}{q^2} + \frac{q^n-1}{q^n(q-1)}.$$

Proof: By (10) we have that

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[|\mathcal{L}_1(\boldsymbol{x})|\right] = (nq-n-1)\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[\rho(\boldsymbol{x})\right] + 2 - \frac{1}{2}\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[\sum_{i=1}^{A(\boldsymbol{x})} s_i^2\right] + \frac{3}{2}\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[\sum_{i=1}^{A(\boldsymbol{x})} s_i\right] - \mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[A(\boldsymbol{x})\right].$$

Using Corollary 8 and Lemmas 11, 12, and 16 we infer that

$$\begin{aligned}\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{Z}_q^n}\left[|\mathcal{L}_1(\boldsymbol{x})|\right] &= (nq-n-1)\left(n-\frac{n-1}{q}\right) + 2\\ &\quad - \frac{1}{2}\left(\frac{n(4q^2-3q+2)}{q^2} + \frac{6q-4}{q^2} - 4 - \frac{2}{q-1}\left(1-\frac{1}{q^n}\right)\right)\\ &\quad + \frac{3}{2}\left(n + (n-2)\cdot\frac{(q-1)(q-2)}{q^2}\right)\\ &\quad - 1 - \frac{(n-2)(q-1)(q-2)}{q^2} - \frac{n-1}{q}\\ &= n^2\left(q+\frac{1}{q}-2\right) - \frac{n}{q} - \frac{(q-1)(q-2)}{q^2} + 3 - \frac{3}{q} + \frac{2}{q^2} + \frac{q^n-1}{q^n(q-1)}.\end{aligned}$$

■

VI. Binary Anticodes With Diameter One

Before presenting the analysis of the anticodes under the FLL metric, we state the following lemma, which was proven in [23, Sections 3 and 5] and will be used in some of the proofs in this section.

Lemma 17: If $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}_2^n$ are distinct words, then

$$|\mathcal{D}_1(\boldsymbol{x}) \cap \mathcal{D}_1(\boldsymbol{y})| \leq 2 \quad \text{and} \quad |\mathcal{I}_1(\boldsymbol{x}) \cap \mathcal{I}_1(\boldsymbol{y})| \leq 2.$$

Definition 14: An *anticode of diameter* t in $\mathbb{Z}_q^n$ is a subset $\mathcal{A} \subseteq \mathbb{Z}_q^n$ such that for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{A}$, $d_\ell(\boldsymbol{x}, \boldsymbol{x}') \leq t$. We say that $\mathcal{A}$ is a *maximal anticode* if there is no other anticode of diameter t in $\mathbb{Z}_q^n$ which contains $\mathcal{A}$.

Next, we present tight lower and upper bounds on the size of maximal binary anticodes of diameter one in the FLL metric. To prove these bounds we need some useful properties of anticodes with diameter one in the FLL metric.

Lemma 18: If an anticode $\mathcal{A}$ of diameter one contains three distinct words with the suffix 00 then there is at most one word in $\mathcal{A}$ with the suffix 01.

Proof: Let $\boldsymbol{a}, \boldsymbol{a}', \boldsymbol{a}'' \in \mathcal{A}$ be three words with the suffix 00 and assume to the contrary that there exist two distinct words $\boldsymbol{b}, \boldsymbol{b}' \in \mathcal{A}$ with the suffix 01. Let $\boldsymbol{y} \in \mathcal{LCS}(\boldsymbol{a}, \boldsymbol{b})$; by Corollary 1 the length of $\boldsymbol{y}$ is $n-1$ and since $\boldsymbol{a}$ ends with 00, $\boldsymbol{y}$ must end with 0 which implies that $\boldsymbol{y} = \boldsymbol{b}_{[1,n-1]}$. By the same arguments $\boldsymbol{y} \in \mathcal{LCS}(\boldsymbol{b}, \boldsymbol{a}')$ and $\boldsymbol{y} \in \mathcal{LCS}(\boldsymbol{b}, \boldsymbol{a}'')$. Similarly,

$$\boldsymbol{y}' = \boldsymbol{b}'_{[1,n-1]} \in \mathcal{LCS}(\boldsymbol{b}', \boldsymbol{a}, \boldsymbol{a}', \boldsymbol{a}'').$$

Hence, $\boldsymbol{a}, \boldsymbol{a}', \boldsymbol{a}'' \in \mathcal{I}_1(\boldsymbol{y}) \cap \mathcal{I}_1(\boldsymbol{y}')$ which is a contradiction to Lemma 17. Thus, $\mathcal{A}$ contains at most one word with the suffix 01. ■

Lemma 19: If an anticode $\mathcal{A}$ of diameter one contains three distinct words with the suffix 01, then there is at most one word in $\mathcal{A}$ with the suffix 00.

Proof: Let $\boldsymbol{a}, \boldsymbol{a}', \boldsymbol{a}'' \in \mathcal{A}$ be three words with the suffix 01 and assume to the contrary that there exist two distinct words $\boldsymbol{b}, \boldsymbol{b}' \in \mathcal{A}$ with the suffix 00. For $\boldsymbol{y} \in \mathcal{LCS}(\boldsymbol{a}, \boldsymbol{b})$, by Corollary 1 the length of $\boldsymbol{y}$ is $n-1$ and since $\boldsymbol{b}$ ends with 00, $\boldsymbol{y}$ must end with 0 which implies that $\boldsymbol{y} = \boldsymbol{a}_{[1,n-1]}$. By the same arguments $\boldsymbol{y} \in \mathcal{LCS}(\boldsymbol{a}, \boldsymbol{b}')$. Similarly,

$$\boldsymbol{y}' = \boldsymbol{a}'_{[1,n-1]} \in \mathcal{LCS}(\boldsymbol{a}', \boldsymbol{b}, \boldsymbol{b}')$$
$$\boldsymbol{y}'' = \boldsymbol{a}''_{[1,n-1]} \in \mathcal{LCS}(\boldsymbol{a}'', \boldsymbol{b}, \boldsymbol{b}').$$

Hence, $\boldsymbol{y}, \boldsymbol{y}', \boldsymbol{y}'' \in \mathcal{D}_1(\boldsymbol{b}) \cap \mathcal{D}_1(\boldsymbol{b}')$ which is a contradiction to Lemma 17. Thus, $\mathcal{A}$ contains at most one word with the suffix 00. ■

Lemma 20: Let $\mathcal{A}$ be an anticode of diameter one. If $\boldsymbol{a}, \boldsymbol{a}' \in \mathcal{A}$ are two distinct words that end with 00 and $\boldsymbol{b}, \boldsymbol{b}' \in \mathcal{A}$ are two distinct words that end with 01, then $\boldsymbol{a}_{[1,n-1]} \neq \boldsymbol{b}_{[1,n-1]}$ or $\boldsymbol{a}'_{[1,n-1]} \neq \boldsymbol{b}'_{[1,n-1]}$.

Proof: Assume to the contrary that there exist $\boldsymbol{a}, \boldsymbol{a}', \boldsymbol{b}, \boldsymbol{b}' \in \mathcal{A}$ such that $\boldsymbol{a}_{[1,n-1]} = \boldsymbol{b}_{[1,n-1]} = \boldsymbol{y}0$ and $\boldsymbol{a}'_{[1,n-1]} = \boldsymbol{b}'_{[1,n-1]} = \boldsymbol{y}'0$, $\boldsymbol{a}, \boldsymbol{a}'$ end with 00 and $\boldsymbol{b}, \boldsymbol{b}'$ end with 01. Let

$$\begin{aligned}
\boldsymbol{a} &= a_1\ a_2 \ldots a_{n-2}\ 0\ 0 = \boldsymbol{y}\ \ 0\ 0\\
\boldsymbol{a}' &= a'_1\ a'_2 \ldots a'_{n-2}\ 0\ 0 = \boldsymbol{y}'\ 0\ 0\\
\boldsymbol{b} &= a_1\ a_2 \ldots a_{n-2}\ 0\ 1 = \boldsymbol{y}\ \ 0\ 1\\
\boldsymbol{b}' &= a'_1\ a'_2 \ldots a'_{n-2}\ 0\ 1 = \boldsymbol{y}'\ 0\ 1.
\end{aligned}$$

Notice that since the FLL distance between any two words in $\mathcal{A}$ is one, it follows that the Hamming weight of any two words can differ by at most one, which implies that $\text{wt}(\boldsymbol{y}) = \text{wt}(\boldsymbol{y}')$ (by considering the pairs $\boldsymbol{a}, \boldsymbol{b}'$ and $\boldsymbol{a}', \boldsymbol{b}$). Clearly, $\boldsymbol{y}0 \in \mathcal{LCS}(\boldsymbol{a}', \boldsymbol{b})$ which implies that $\boldsymbol{a}'$ can be obtained from $\boldsymbol{b}$ by deleting the last 1 of $\boldsymbol{b}$ and then inserting 0 into the LCS. Hence, there exists an index $0 \le j \le n-2$ such that

$$a_1 a_2 \ldots a_j 0 a_{j+1} \ldots a_{n-2} 0 = a'_1 a'_2 \ldots a'_j a'_{j+1} \ldots a'_{n-2} 00. \tag{13}$$

Similarly, $\boldsymbol{a}$ can be obtained from $\boldsymbol{b}'$, i.e., there exists an index $0 \le i \le n-2$ such that

$$a'_1 a'_2 \ldots a'_i 0 a'_{i+1} \ldots a'_{n-2} 0 = a_1 a_2 \ldots a_i a_{i+1} \ldots a_{n-2} 00. \tag{14}$$

Assume w.l.o.g. that $i \le j$. Equation (13) implies that $a_r = a_{r'}$ for $1 \le r \le j$. In addition, $a_{n-2} = 0$ by (13) and $a'_{n-2} = 0$ by (14). By assigning $a_{n-2} = a'_{n-2} = 0$ into (13) and (14) we obtain that $a_{n-3} = a'_{n-3} = 0$. Repeating this process implies that $a_r = a_{r'} = 0$ for $j+1 \le r \le n-2$. Thus, we have that $\boldsymbol{y} = \boldsymbol{y}'$ which is a contradiction. ■

Definition 15: For an anticode $\mathcal{A} \subseteq \mathbb{Z}_2^n$, the *puncturing of $\mathcal{A}$ in the n-th coordinate*, $\mathcal{A}'$, is defined by

$$\mathcal{A}' \triangleq \left\{\boldsymbol{a}_{[1,n-1]}\ \ \boldsymbol{a} \in \mathcal{A}\right\}.$$

Lemma 21: Let $\mathcal{A} \subseteq \mathbb{Z}_2^n$ be an anticode of diameter one. If the last symbol in all the words in $\mathcal{A}$ is the same symbol $\sigma \in \mathbb{Z}_2$, then $\mathcal{A}'$ is an anticode of diameter one and $|\mathcal{A}'| = |\mathcal{A}|$.

Proof: Let $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{A}$ be two different words and let $\boldsymbol{y} \in \mathcal{LCS}(\boldsymbol{a}_{[1,n-1]}, \boldsymbol{b}_{[1,n-1]})$. By (6), $\mathsf{LCS}(\boldsymbol{a}, \boldsymbol{b}) \le |\boldsymbol{y}| + 1$ and since $d_\ell(\boldsymbol{a}, \boldsymbol{b}) = 1$, Corollary 1 implies that $|\boldsymbol{y}| \ge n-2$ and that

$$d_\ell(\boldsymbol{a}_{[1,n-1]}, \boldsymbol{b}_{[1,n-1]}) \le 1.$$

Hence, $\mathcal{A}$ is an anticode of diameter one. Since any two distinct words $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{A}$ end with the symbol σ, it follows that $\boldsymbol{a}_{[1,n-1]} \neq \boldsymbol{b}_{[1,n-1]}$ and thus $|\mathcal{A}| = |\mathcal{A}'|$. ■

Lemma 22: Let $\mathcal{A}$ be an anticode of diameter one. If the suffix of each word in $\mathcal{A}$ is either 01 or 10, then $\mathcal{A}'$ is an anticode of diameter one and $|\mathcal{A}'| = |\mathcal{A}|$.

Proof: Let $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{A}$ be two different words and let $\boldsymbol{y} \in \mathcal{LCS}(\boldsymbol{a}_{[1,n-1]}, \boldsymbol{b}_{[1,n-1]})$. By (6), $\mathsf{LCS}(\boldsymbol{a}, \boldsymbol{b}) \le |\boldsymbol{y}| + 1$ and since $d_\ell(\boldsymbol{a}, \boldsymbol{b}) = 1$, it follows that $|\boldsymbol{y}| \ge n-2$ and that

$$d_\ell(\boldsymbol{a}_{[1,n-1]}, \boldsymbol{b}_{[1,n-1]}) \le 1.$$

Hence, $\mathcal{A}'$ is an anticode of diameter one. If $\boldsymbol{a}$ and $\boldsymbol{b}$ end with the same symbol $\sigma \in \{0,1\}$, then $\boldsymbol{a}_{[1,n-1]} \neq \boldsymbol{b}_{[1,n-1]}$. Otherwise, one of the words has the suffix 01 and the other has the suffix 10. That is, $a_{n-1} \neq b_{n-1}$ and therefore $\boldsymbol{a}_{[1,n-1]} \neq \boldsymbol{b}_{[1,n-1]}$ and thus, $|\mathcal{A}'| = |\mathcal{A}|$. ■

A. Upper Bound

Theorem 16: Let $n > 1$ be an integer and let $\mathcal{A} \subseteq \mathbb{Z}_2^n$ be a maximal anticode of diameter one. Then, $|\mathcal{A}| \le n+1$, and there exists a maximal anticode with exactly $n+1$ codewords.

Proof: Since two words $\boldsymbol{x}, \boldsymbol{y}$ such that $\boldsymbol{x}$ ends with 00 and $\boldsymbol{y}$ ends with 11 are at FLL distance at least 2, w.l.o.g. assume that $\mathcal{A}$ does not contain codewords that end with 11. It is easy to verify that the theorem holds for $n \in \{2,3,4\}$. Assume that the theorem does not hold and let $n^* > 4$ be the smallest integer such that there exists an anticode $\mathcal{A} \subseteq \mathbb{Z}_2^{n^*}$ such that $|\mathcal{A}| = n^* + 2$. Since there are only three possible options for the last two symbols of codewords in $\mathcal{A}$ (00, 01, or 10) and $|\mathcal{A}| \ge 7$, it follows that there exist three different codewords in $\mathcal{A}$ with the same suffix of two symbols.

Case 1 - Assume $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{A}$ are three different words with the suffix 00. By Lemma 18, there exists at most one codeword in $\mathcal{A}$ with the suffix 01 and since $\mathcal{A}$ does not contain codewords with the suffix 11, there exists at most one codeword in $\mathcal{A}$ that ends with the symbol 1. That is, there exist at least n^*+1 codewords with 0 as the last symbol. Denote such a set with n^*+1 codewords by $\mathcal{A}_1$. As a subset of the anticode $\mathcal{A}$, $\mathcal{A}_1$ is also an anticode and hence by Lemma 21, $\mathcal{A}'_1$ is an anticode of length n^*-1 and size n^*+1 which is a contradiction to the minimality of n^*.

Case 2 - Assume $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{A}$ are three different words with the suffix 01. By Lemma 19, there exists at most one codeword in $\mathcal{A}$ with the suffix 00 and since $\mathcal{A}$ does not contain codewords with the suffix 11 there exist n^*+1 codewords that end with either 01 or 10. Denote this set of n^*+1 codewords as $\mathcal{A}_1$. As a subset of the anticode $\mathcal{A}$, $\mathcal{A}_1$ is also an anticode and hence by Lemma 22, $\mathcal{A}'_1$ is an anticode of length n^*-

1 and size n^*+1 which is a contradiction to the minimality of n^*.

Case 3 - Assume $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\in\mathcal{A}$ are three different words with the suffix 10. By the previous two cases, there exist at most two codewords in $\mathcal{A}$ with the suffix 00 and at most two codewords with the suffix 01. Since there are no codewords with the suffix 11, it follows that the number of words that end with 1 is at most two. If there exist at most one codeword in $\mathcal{A}$ that ends with 1, then there are n^*+1 codewords in $\mathcal{A}$ that end with 0 and as in the first case, this leads to a contradiction. Otherwise there are exactly two codewords in $\mathcal{A}$ with the suffix 01. If there are less than two codewords with the suffix 00, then, the number of codewords with suffixes 01 and 10 is at least n^*+1 and similarly to Case 2, this is a contradiction to the minimality of n^*. Hence, there exist exactly two codewords in $\mathcal{A}$ with the suffix 00. There are exactly n^*-2 codewords in $\mathcal{A}$ with the suffix 10 and two more codewords with the suffix 01. By Lemma 22 the words in $\mathcal{A}'$ that were obtained from these n^* codewords are all different and have FLL distance one from each other. In addition, by Lemma 20, the prefix of length n^*-1 of at least one of the codewords that end with 00 is different from the prefixes of length n^*-1 of the codewords that end with 01. This prefix also differs from the prefixes of the codewords that end with 10. Therefore, $\mathcal{A}'$ is an anticode with n^*+1 different codewords which is a contradiction to the minimality of n^*.

Note that the set $\mathcal{A}=\{a\in\mathbb{Z}_2^n\ \ \mathrm{wt}(a)\le 1\}$ is an anticode of diameter one with exactly $n+1$ codewords. Thus, the maximum size of an anticode of diameter one is $n+1$. ■

B. Lower Bound

Theorem 17: Let $n>2$ be a positive integer and let $\mathcal{A}\subseteq\mathbb{Z}_2^n$ be a maximal anticode of diameter one, then $|\mathcal{A}|\ge 4$ and there exists a maximal anticode with exactly 4 codewords.

Proof: For $n=3$ the maximal anticodes are

$$\mathcal{A}_1=\{000,\ 001,\ 010,\ 100\}\ \mathcal{A}_2=\{001,\ 010,\ 100,\ 101\}$$
$$\mathcal{A}_3=\{001,\ 010,\ 011,\ 101\}\ \mathcal{A}_4=\{010,\ 011,\ 101,\ 110\}$$
$$\mathcal{A}_5=\{011,\ 101,\ 110,\ 111\}\ \mathcal{A}_6=\{010,\ 100,\ 101,\ 110\}$$

and all of them have size $4=n+1$. Assume that the theorem does not hold and let $n^*>3$ be the smallest integer such that there exists a maximal anticode $\mathcal{A}\subseteq\mathbb{Z}_2^{n^*}$ with less than four codewords. For each $\boldsymbol{x}\in\mathbb{Z}_2^{n^*}$ there exists a sequence $\boldsymbol{y}\in\mathbb{Z}_2^{n^*}$ such that $d_\ell(\boldsymbol{x},\boldsymbol{y})=1$ and hence $|\mathcal{A}|>1$. If $\mathcal{A}=\{\boldsymbol{x},\boldsymbol{y}\}\subseteq\mathbb{Z}_2^{n^*}$ by the definition of an anticode $d_\ell(\boldsymbol{x},\boldsymbol{y})=1$ and $\mathsf{LCS}(\boldsymbol{x},\boldsymbol{y})=n-1$. For $\boldsymbol{z}\in\mathcal{LCS}(\boldsymbol{x},\boldsymbol{y})$, by (2), the insertion ball of radius one centered at $\boldsymbol{z}$ contains $n^*-1>2$ codewords in addition to $\boldsymbol{x}$ and $\boldsymbol{y}$ and each of them can be added into $\mathcal{A}$. Hence, $\mathcal{A}$ is an anticode of diameter one with three codewords. We will prove that there exists a word that can be added into $\mathcal{A}$ which is a contradiction to the maximality of $\mathcal{A}$. Consider the following cases:

Case 1 - If all the codewords in $\mathcal{A}$ have the same last symbol $\sigma\in\mathbb{Z}_2$, then by Lemma 21, $\mathcal{A}'\subseteq\mathbb{Z}_2^{n^*-1}$, is an anticode of diameter one that contains three codewords. Since n^* is the smallest integer for which there exists a maximal anticode with less than four codewords, $\mathcal{A}'$ is not maximal. That is, there exists a word $\boldsymbol{x}'\in\mathbb{Z}_2^{n^*-1}$ such that $\mathcal{A}'\cup\{\boldsymbol{x}'\}$ is an anticode of diameter one. It can be readily verified that $\boldsymbol{x}'\sigma\notin\mathcal{A}$ and that $\mathcal{A}\cup\{\boldsymbol{x}'\sigma\}$ is an anticode of diameter one which is a contradiction to the maximality of $\mathcal{A}$.

Case 2 - If all the codewords in $\mathcal{A}$ have the same first symbol $\sigma\in\mathbb{Z}_2$ then a contradiction is obtained by symmetrical arguments to those presented in Case 1.

Case 3 - Assume all the words in $\mathcal{A}$ neither have the same first symbol nor the same last symbol. Let $|\mathcal{A}|=\{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\}$ and assume w.l.o.g. that $\boldsymbol{x}$ and $\boldsymbol{y}$ are codewords that end with 0 and that $\boldsymbol{z}$ ends with 1. If $|\mathcal{A}'|=3$, then $\boldsymbol{z}_{[1,n^*-1]}\ne\boldsymbol{x}_{[1,n^*-1]}$ and $\boldsymbol{z}_{[1,n^*-1]}\ne\boldsymbol{y}_{[1,n^*-1]}$. Hence the word $\boldsymbol{z}_{[1,n^*-1]}0$ is not in $\mathcal{A}$ and it is easy to verify that it has distance one from each codeword in $\mathcal{A}$, which is a contradiction. Otherwise, since $\boldsymbol{x}_{[1,n^*-1]}\ne\boldsymbol{y}_{[1,n^*-1]}$, it must hold that $\boldsymbol{z}_{[1,n^*-1]}$ is equal either to $\boldsymbol{x}_{[1,n^*-1]}$ or to $\boldsymbol{y}_{[1,n^*-1]}$. Assume w.l.o.g. that $\boldsymbol{z}_{[1,n^*-1]}=\boldsymbol{x}_{[1,n^*-1]}$, then $\boldsymbol{x}$ and $\boldsymbol{z}$ have the same first symbol σ and hence $\boldsymbol{y}$ must begin with $\overline{\sigma}=1-\sigma$. The three codewords can be described as follows:

$$\boldsymbol{x}=\sigma x_2x_3\ldots x_{n^*-1}0$$
$$\boldsymbol{y}=\overline{\sigma}y_2y_3\ldots y_{n^*-1}0$$
$$\boldsymbol{z}=\sigma x_2x_3\ldots x_{n^*-1}1.$$

Since $\boldsymbol{y}$ and $\boldsymbol{z}$ have different first and last symbols, their LCS must be equal to the suffix of length n^*-1 of one word and to the prefix of length n^*-1 of the other word. If

$$\boldsymbol{z}_{[1,n^*-1]}=\sigma a_2a_3\ldots a_{n^*-1}\in\mathcal{LCS}(\boldsymbol{y},\boldsymbol{z}),$$

then $\boldsymbol{z}_{[1,n^*-1]}$ is a common LCS of the three codewords $\boldsymbol{x},\boldsymbol{y}$, and $\boldsymbol{z}$ and hence any word from $\mathcal{I}_1(\boldsymbol{z}_{[1,n^*-1]})$ has distance one from all the words in $\mathcal{A}$. Since, by (2),

$$|\mathcal{I}_1(\boldsymbol{z}_{[1,n^*-1]})|=n^*+1\ge 4,$$

there is a word different from $\boldsymbol{x},\boldsymbol{y}$ and $\boldsymbol{z}$ that can be added into $\mathcal{A}$. In the other case,

$$\begin{aligned}\overline{\sigma}y_2y_3\ldots y_{n^*-1}=\boldsymbol{y}_{[1,n-1]}&=\boldsymbol{z}_{[2,n]}\\ &=x_2x_3\ldots x_{n^*-1}1\in\mathcal{LCS}(\boldsymbol{y},\boldsymbol{z})\end{aligned}$$

and hence the codewords $\boldsymbol{x}$ and $\boldsymbol{z}$ can be written as

$$\boldsymbol{x}=\sigma\overline{\sigma}y_2y_3\ldots y_{n^*-2}0$$
$$\boldsymbol{z}=\sigma\overline{\sigma}y_2y_3\ldots y_{n^*-2}y_{n^*-1}$$

and the word

$$\boldsymbol{w}=\sigma\overline{\sigma}y_2y_3\ldots y_{n^*-1}0$$

is a common SCS of $\boldsymbol{x},\boldsymbol{y}$, and $\boldsymbol{z}$. If $\rho(\boldsymbol{w})>3$ then there is a word in $\mathcal{D}_1(\boldsymbol{w})$ that is different from $\boldsymbol{x},\boldsymbol{y}$, and $\boldsymbol{z}$ that can be added into $\mathcal{A}$, which is again a contradiction. Otherwise, since the first two symbols of $\boldsymbol{w}$ are different and the last two symbols are also different, it holds that $\rho(\boldsymbol{w})=3$. It is easy to verify that

$$\mathcal{A}=\{0\underbrace{11\ldots1}_{n^*-2\text{ times}}0,0\underbrace{11\ldots1}_{n^*-1\text{ times}},\underbrace{11\ldots1}_{n^*-1\text{ times}}0\}$$

and that $\underbrace{11\ldots1}_{n^*-2\text{ times}}01$ can be added into $\mathcal{A}$, which is a contradiction to the minimality of $\mathcal{A}$. To see that the given bound is tight, one can simply consider the set of codewords that consist from the binary representation of length n^* of the numbers $2,3,5,6$ that is, the set

$$\mathcal{A}=\{\underbrace{0\ldots0}_{n^*-3}010,\ \underbrace{0\ldots0}_{n^*-3}011,\ \underbrace{0\ldots0}_{n^*-3}101,\ \underbrace{0\ldots0}_{n^*-3}110\}$$

and verify that it is indeed a maximal anticode of diameter one. ■

VII. Conclusion

In this paper we studied the size of balls with radius one and the anticodes of diameter one under the FLL metric. In particular we give explicit expressions for the maximum size of a ball with radius one and the minimum size of a ball of any given radius in the FLL metric over $\mathbb{Z}_q$. We also found the average size of a 1-ball in the FLL metric. Finally, we considered the related concept of anticode in the FLL distance and we found that the maximum and minimum size of a binary maximal anticode of diameter one are $n+1$ and 4, respectively. The latter can be extended to a non-binary alphabet and while the minimum size of a maximal anticode with diameter one is 4 for any alphabet size q, the maximum size of a maximal anticode with diameter one is $n(q-1)+1$. Recently, based on these results, G. Wang and Q. Wang [36] extended the analysis of 1-FLL balls by proving that the size of the 1-FLL balls is highly concentrated around its mean using Azuma's inequality [1]. A future direction is to study the maximum size of FLL balls for larger radii, and in particular for radius two. Based on a computer search, it appears that the maximum balls are again centered at α-balanced sequences, however, understanding the value of α for this case is more challenging. For example,

- for $n=16$, the maximum ball is centered at

$$\boldsymbol{x}=0101101001011010,$$

which is a 4-balanced sequence for which $|\mathcal{L}_2(\boldsymbol{x})|=4,513$.

- for $n=20$, the maximum ball is centered at

$$\boldsymbol{x}=01010010100101001010,$$

which is a 4-balanced sequence for which $|\mathcal{L}_2(\boldsymbol{x})|=12,759$.

- for $n=25$, the maximum ball is centered at

$$\boldsymbol{x}=0101001010010100101001010,$$

which is a 5-balanced sequence for which $|\mathcal{L}_2(\boldsymbol{x})|=35,893$.

Appendix

Proof of Lemma 10

Let $\alpha>1$ be some integer and define $\mathsf{diff}\triangleq\left|L_1\left(\boldsymbol{x}^{(\alpha)}\right)\right|-\left|L_1\left(\boldsymbol{x}^{(\alpha-1)}\right)\right|$. We will prove that $\mathsf{diff}>0$ if and only if $n>2\alpha(\alpha-1)$ by proving that $\mathsf{diff}>0$ for any $n>2\alpha(\alpha-1)$ and that $\mathsf{diff}<0$ for any $\alpha<n<2\alpha(\alpha-1)$. Before we analyze each case we present different expression for $\left|L_1\left(\boldsymbol{x}^{(\alpha)}\right)\right|$ that will be in use within the proof. By Lemma 9, if n is divisible by α, then

$$\begin{aligned}\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right| &= (n+1-\alpha)(n-1)+2-\frac{\alpha}{2}\left(\frac{n}{\alpha}-1\right)\left(\frac{n}{\alpha}-2\right)\\ &=(n+1-\alpha)(n-1)+2-\frac{n^2}{2\alpha}+\frac{3n}{2}-\alpha.\end{aligned}$$

Otherwise when n is not divisible by α, we have that $\left\lceil\frac{n}{\alpha}\right\rceil=\frac{n-k_\alpha}{\alpha}+1$ and hence, by Lemma 9,

$$\begin{aligned}\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right| &= (n+1-\alpha)(n-1)+2-\frac{k_\alpha}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-1\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\\ &\quad-\frac{\alpha-k_\alpha}{2}\left(\left\lceil\frac{n}{\alpha}\right\rceil-2\right)\left(\left\lceil\frac{n}{\alpha}\right\rceil-3\right)\\ &=(n+1-\alpha)(n-1)+2-\frac{k_\alpha}{2}\left(\frac{n-k_\alpha}{\alpha}\right)\left(\frac{n-k_\alpha}{\alpha}-1\right)\\ &\quad-\frac{\alpha-k_\alpha}{2}\left(\frac{n-k_\alpha}{\alpha}-1\right)\left(\frac{n-k_\alpha}{\alpha}-2\right)\\ &=(n+1-\alpha)(n-1)+2\\ &\quad-\frac{k_\alpha}{2}\left(\frac{n-k_\alpha}{\alpha}-1\right)\left(\frac{n-k_\alpha}{\alpha}-\frac{n-k_\alpha}{\alpha}+2\right)\\ &\quad-\frac{\alpha}{2}\left(\frac{n-k_\alpha}{\alpha}-1\right)\left(\frac{n-k_\alpha}{\alpha}-2\right)\\ &=(n+1-\alpha)(n-1)+2-k_\alpha\left(\frac{n-k_\alpha}{\alpha}-1\right)\\ &\quad-\frac{\alpha}{2}\left(\frac{n-k_\alpha}{\alpha}-1\right)\left(\frac{n-k_\alpha}{\alpha}-2\right)\\ &=(n+1-\alpha)(n-1)+2-\left(\frac{n-k_\alpha}{\alpha}-1\right)\left(k_\alpha+\frac{n-k_\alpha}{2}-\alpha\right)\\ &=(n+1-\alpha)(n-1)+2-\frac{(n-k_\alpha-\alpha)(n+k_\alpha-2\alpha)}{2\alpha}\\ &=(n+1-\alpha)(n-1)+2-\frac{n^2}{2\alpha}+\frac{3n}{2}+\frac{k_\alpha^2}{2\alpha}-\frac{k_\alpha}{2}-\alpha.\end{aligned}$$

Hence, by abuse of notation, if we let $0\le k_\alpha\le\alpha-1$ we have that

$$\left|\mathcal{L}_1\left(\boldsymbol{x}^{(\alpha)}\right)\right|=(n+1-\alpha)(n-1)+2-\frac{n^2}{2\alpha}+\frac{3n}{2}+\frac{k_\alpha^2}{2\alpha}-\frac{k_\alpha}{2}-\alpha,$$

which implies that

$$\begin{aligned}\mathsf{diff} &= (n+1-\alpha)(n-1)+2-\frac{n^2}{2\alpha}+\frac{3n}{2}+\frac{k_\alpha^2}{2\alpha}-\frac{k_\alpha}{2}-\alpha\\ &\quad-(n+1-(\alpha-1))(n-1)-2+\frac{n^2}{2(\alpha-1)}-\frac{3n}{2}\\ &\quad-\frac{k_{\alpha-1}^2}{2(\alpha-1)}+\frac{k_{\alpha-1}}{2}+\alpha-1\\ &=-n+n^2\left(\frac{1}{2(\alpha-1)}-\frac{1}{2\alpha}\right)+\left(\frac{k_\alpha^2}{2\alpha}-\frac{k_\alpha}{2}\right)\\ &\quad+\left(\frac{k_{\alpha-1}}{2}-\frac{k_{\alpha-1}^2}{2(\alpha-1)}\right)\\ &=\frac{n^2}{2\alpha(\alpha-1)}-n+\left(\frac{k_\alpha^2}{2\alpha}-\frac{k_\alpha}{2}\right)+\left(\frac{k_{\alpha-1}}{2}-\frac{k_{\alpha-1}^2}{2(\alpha-1)}\right).\end{aligned}$$

Let us consider the following distinct cases for the value of n.

Case 1 - If $n=2\alpha(\alpha-1)$ then $k_\alpha=k_{\alpha-1}=0$ and

$$\mathsf{diff}=\frac{(2\alpha(\alpha-1))^2}{2(\alpha-1)\alpha}-2\alpha(\alpha-1)=0.$$

Case 2 - If $n = 2\alpha(\alpha-1)+k$ for some integer $1 \leq k \leq \alpha - 2$ then we have that $k_\alpha = k_{\alpha-1} = k$ and

$$\begin{aligned}\mathsf{diff} &= \frac{n^2}{2\alpha(\alpha-1)} - n + \left(\frac{k^2}{2\alpha} - \frac{k}{2}\right) + \left(\frac{k}{2} - \frac{k^2}{2(\alpha-1)}\right)\\ &= \frac{n^2}{2\alpha(\alpha-1)} - n + \frac{k^2}{2\alpha} - \frac{k^2}{2(\alpha-1)}\\ &= \frac{n^2}{2\alpha(\alpha-1)} - n - \frac{k^2}{2\alpha(\alpha-1)}\\ &= \frac{(2\alpha(\alpha-1)+k)^2}{2\alpha(\alpha-1)} - 2\alpha(\alpha-1) - k - \frac{k^2}{2\alpha(\alpha-1)}\\ &= 2\alpha(\alpha-1)+2k+\frac{k^2}{2\alpha(\alpha-1)} - 2\alpha(\alpha-1) - k - \frac{k^2}{2\alpha(\alpha-1)}\\ &= k > 0.\end{aligned}$$

Case 3 - If $n \geq 2\alpha(\alpha-1)+\alpha-1$, then we first note that for any $0 \leq k_{\alpha-1} \leq \alpha-2$ we have that $\left(\frac{k_{\alpha-1}}{2} - \frac{k^2_{\alpha-1}}{2(\alpha-1)}\right) \geq 0$ and hence

$$\begin{aligned}\mathsf{diff} &= \frac{n^2}{2\alpha(\alpha-1)} - n + \left(\frac{k^2_\alpha}{2\alpha} - \frac{k_\alpha}{2}\right) + \left(\frac{k_{\alpha-1}}{2} - \frac{k^2_{\alpha-1}}{2(\alpha-1)}\right)\\ &\geq \frac{n^2}{2\alpha(\alpha-1)} - n + \left(\frac{k^2_\alpha}{2\alpha} - \frac{k_\alpha}{2}\right).\end{aligned}$$

Define $f : [0, \alpha-1] \to \mathbb{R}$ by $f(x) \triangleq \frac{x^2}{2\alpha} - \frac{x}{2}$. It is easy to verify that f has a single minimum point at $x = \frac{\alpha}{2}$. Hence,

$$\frac{k^2_\alpha}{2\alpha} - \frac{k_\alpha}{2} = f(k_\alpha) \geq f\left(\frac{\alpha}{2}\right) = \frac{\alpha^2}{4 \cdot 2\alpha} - \frac{\alpha}{2} = -\frac{\alpha}{8}$$

and

$$\mathsf{diff} \geq \frac{n^2}{2\alpha(\alpha-1)} - n - \frac{\alpha}{8}.$$

It holds that

$$\frac{n^2}{2\alpha(\alpha-1)} - n - \frac{\alpha}{8} \geq 0$$

if and only if

$$\begin{aligned}n &\geq \alpha(\alpha-1) + \frac{1}{2}\sqrt{4\alpha^4 - 7\alpha^3 + 3\alpha^2}\\ &= \alpha(\alpha-1) + \sqrt{\alpha^4 - \frac{7}{4}\alpha^3 + \frac{3}{4}\alpha^2}\\ &= \alpha(\alpha-1) + \alpha\sqrt{\alpha^2 - \frac{7}{4}\alpha + \frac{3}{4}}\\ &= \alpha(\alpha-1) + \alpha\sqrt{\left(\alpha - \frac{3}{4}\right)(\alpha-1)}.\end{aligned}$$

Note that

$$\alpha\sqrt{\left(\alpha - \frac{3}{4}\right)(\alpha-1)} < \alpha\left(\alpha - \frac{3}{4}\right),$$

and additionally, it can be verified that for any $\alpha > 1$,

$$2\alpha(\alpha-1) + (\alpha-1) \geq \alpha(\alpha-1) + \alpha\left(\alpha - \frac{3}{4}\right),$$

and thus,

$$\begin{aligned}n &\geq 2\alpha(\alpha-1) + (\alpha-1)\\ &> \alpha(\alpha-1) + \alpha\sqrt{\left(\alpha - \frac{3}{4}\right)(\alpha-1)},\end{aligned}$$

which implies that $\mathsf{diff} > 0$.

Case 4 - If $n = 2\alpha(\alpha-1) - k$ for some integer $1 \leq k \leq \alpha - 2$ then we have that $k_\alpha = \alpha - k,\ k_{\alpha-1} = \alpha - 1 - k$, and thus

$$\begin{aligned}\mathsf{diff} &= \frac{n^2}{2\alpha(\alpha-1)} - n + \left(\frac{(\alpha-k)^2}{2\alpha} - \frac{\alpha-k}{2}\right)\\ &\quad + \left(\frac{\alpha-1-k}{2} - \frac{(\alpha-1-k)^2}{2(\alpha-1)}\right)\\ &= \frac{n^2}{2\alpha(\alpha-1)} - n + \frac{\alpha^2}{2\alpha} - \frac{2\alpha k}{2\alpha} + \frac{k^2}{2\alpha}\\ &\quad + \frac{\alpha-1-k-\alpha+k}{2} - \frac{(\alpha-1)^2}{2(\alpha-1)}\\ &\quad + \frac{2(\alpha-1)k}{2(\alpha-1)} - \frac{k^2}{2(\alpha-1)}\\ &= \frac{n^2}{2\alpha(\alpha-1)} - n + \frac{\alpha}{2} - k + \frac{k^2}{2\alpha} - \frac{1}{2}\\ &\quad - \frac{\alpha-1}{2} + k - \frac{k^2}{2(\alpha-1)}\\ &= \frac{n^2}{2\alpha(\alpha-1)} - n + \frac{k^2}{2\alpha} - \frac{k^2}{2(\alpha-1)}\\ &= \frac{(2\alpha(\alpha-1)-k)^2}{2\alpha(\alpha-1)} - 2\alpha(\alpha-1) + k + \frac{k^2}{2\alpha} - \frac{k^2}{2(\alpha-1)}\\ &= 2\alpha(\alpha-1) - 2k + \frac{k^2}{2\alpha(\alpha-1)} - 2\alpha(\alpha-1)\\ &\quad + k - \frac{k^2}{2\alpha(\alpha-1)}\\ &= -k < 0.\end{aligned}$$

Case 5 - If $\alpha \leq n \leq 2\alpha(\alpha-1) - (\alpha-1)$ then we first note that for any $0 \leq k_\alpha \leq \alpha-1$ we have that $\left(\frac{k^2_\alpha}{2\alpha} - \frac{k_\alpha}{2}\right) \leq 0$ and hence

$$\begin{aligned}\mathsf{diff} &= \frac{n^2}{2\alpha(\alpha-1)} - n + \left(\frac{k^2_\alpha}{2\alpha} - \frac{k_\alpha}{2}\right) + \left(\frac{k_{\alpha-1}}{2} - \frac{k^2_{\alpha-1}}{2(\alpha-1)}\right)\\ &\leq \frac{n^2}{2\alpha(\alpha-1)} - n + \left(\frac{k_{\alpha-1}}{2} - \frac{k^2_{\alpha-1}}{2(\alpha-1)}\right).\end{aligned}$$

Define $f : [0, \alpha-2] \to \mathbb{R}$ by $f(x) \triangleq \frac{x}{2} - \frac{x^2}{2(\alpha-1)}$. It can be verified that f has a single maximum point at $x = \frac{\alpha-1}{2}$ and hence

$$\begin{aligned}\frac{k_{\alpha-1}}{2} - \frac{k^2_{\alpha-1}}{2(\alpha-1)} &= f(k_{\alpha-1}) \leq f\left(\frac{\alpha-1}{2}\right)\\ &= \frac{\alpha-1}{4} - \frac{(\alpha-1)^2}{8(\alpha-1)} = \frac{\alpha-1}{8}\end{aligned}$$

and

$$\mathsf{diff} \leq \frac{n^2}{2\alpha(\alpha-1)} - n + \frac{\alpha-1}{8}.$$

It holds that $\frac{n^2}{2\alpha(\alpha-1)} - n + \frac{\alpha-1}{8} \leq 0$ if and only if

$$\alpha(\alpha-1) - \sqrt{\frac{(\alpha-1)^2\alpha(4\alpha-1)}{4}} \leq n \leq \alpha(\alpha-1) + \sqrt{\frac{(\alpha-1)^2\alpha(4\alpha-1)}{4}}.$$

Note that

$$\begin{aligned}\alpha(\alpha-1) - \sqrt{\frac{(\alpha-1)^2\alpha(4\alpha-1)}{4}} &= (\alpha-1)\left(\alpha - \sqrt{\frac{\alpha\,(4\alpha-1)}{4}}\right)\\ &= (\alpha-1)\left(\alpha - \sqrt{\alpha\left(\alpha-\frac{1}{4}\right)}\right)\\ &\leq (\alpha-1)\left(\alpha - \left(\alpha-\frac{1}{4}\right)\right)\\ &\leq \frac{\alpha-1}{4},\end{aligned}$$

and

$$\begin{aligned}\alpha(\alpha-1) + \sqrt{\frac{(\alpha-1)^2\alpha(4\alpha-1)}{4}} &= (\alpha-1)\left(\alpha + \sqrt{\frac{\alpha\,(4\alpha-1)}{4}}\right)\\ &= (\alpha-1)\left(\alpha + \sqrt{\alpha\left(\alpha-\frac{1}{4}\right)}\right)\\ &\geq (\alpha-1)\left(\alpha + \left(\alpha-\frac{1}{4}\right)\right)\\ &= 2\alpha(\alpha-1) - \frac{\alpha-1}{4},\end{aligned}$$

and since $\alpha \leq n \leq 2\alpha(\alpha-1)-(\alpha-1)$, we have that $\mathsf{diff} < 0$ as required.

Since α is the number of alternating segments in a sequence of length n, it holds that $n \geq \alpha$. In addition, Case 1 states that for $n = 2\alpha(\alpha-1)$ we have that $\mathsf{diff} = 0$. Furthermore, by combining the results from Case 2 and Case 3 we have that for any $n > 2\alpha(\alpha-1)$ the value of diff is a positive number. Similarly Case 4 and Case 5 prove that for any $\alpha \leq n < 2\alpha(\alpha-1)$ the value of diff is negative. Thus,

$$\mathsf{diff} = 0 \iff n \geq 2\alpha(\alpha-1).$$

ACKNOWLEDGMENT

The authors would like to thank the associate editor Xiande Zhang and the two anonymous reviewers for their very helpful comments.

REFERENCES

[1] N. Alon and J. H. Spencer, *The Probabilistic Method* (Series in Discrete Mathematics and Optimization), 4th ed. Hoboken, NJ, USA: Wiley, 2016.

[2] L. Anavy, I. Vaknin, O. Atar, R. Amit, and Z. Yakhini, "Data storage in DNA with fewer synthesis cycles using composite DNA letters," *Nature Biotechnol.*, vol. 37, no. 10, pp. 1229–1236, Oct. 2019.

[3] D. Bar-Lev, T. Etzion, and E. Yaakobi, "On Levenshtein balls with radius one," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Melbourne, VIC, Australia, Jul. 2021, pp. 1979–1984.

[4] D. Bar-Lev, I. Orr, O. Sabary, T. Etzion, and E. Yaakobi, "Deep DNA storage: Scalable and robust DNA storage via coding theory and deep learning," 2021, *arXiv:2109.00031*.

[5] J. Brakensiek, V. Guruswami, and S. Zbarsky, "Efficient low-redundancy codes for correcting multiple deletions," in *Proc. 27th Annu. ACM-SIAM Symp. Discrete Algorithms*, Philadelphia, PA, USA, Jan. 2016, pp. 1884–1892.

[6] B. Bukh, V. Guruswami, and J. Håstad, "An improved bound on the fraction of correctable deletions," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 93–103, Jan. 2017.

[7] J. Castiglione and A. Kavčić, "Trellis based lower bounds on capacities of channels with synchronization errors," in *Proc. IEEE Inf. Theory Workshop-Fall (ITW)*, Oct. 2015, pp. 24–28.

[8] M. Cheraghchi, "Capacity upper bounds for deletion-type channels," *J. ACM*, vol. 66, no. 2, pp. 1–79, Apr. 2019.

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

[10] R. Con and A. Shpilka, "Explicit and efficient constructions of coding schemes for the binary deletion channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Los Angeles, CA, USA, Jun. 2020, pp. 84–89.

[11] D. Cullina and N. Kiyavash, "An improvement to Levenshtein's upper bound on the cardinality of deletion correcting codes," in *Proc. IEEE Int. Symp. Inf. Theory, (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 699–703.

[12] Y. Erlich and D. Zielinski, "DNA Fountain enables a robust and efficient storage architecture," *Science*, vol. 355, no. 6328, pp. 950–954, 2017.

[13] R. Gabrys and F. Sala, "Codes correcting two deletions," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 965–974, Feb. 2019.

[14] N. Goldman et al., "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA," *Nature*, vol. 494, no. 7435, pp. 77–80, 2013.

[15] 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," *Angew. Chem. Int. Ed.*, vol. 54, no. 8, pp. 2552–2555, Feb. 2015.

[16] V. Guruswami and C. Wang, "Deletion codes in the high-noise and high-rate regimes," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 1961–1970, Apr. 2017.

[17] R. Heckel, G. Mikutis, and R. N. Grass, "A characterization of the DNA data storage channel," *Sci. Rep.*, vol. 9, no. 1, p. 9663, Dec. 2019.

[18] R. Heckel, I. Shomorony, K. Ramchandran, and D. N. C. Tse, "Fundamental limits of DNA storage systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 3130–3134.

[19] D. S. Hirschberg and M. Regnier, "Tight bounds on the number of string subsequences," *J. Disc. Algorithms*, vol. 1, no. 1, pp. 123–132, 2000.

[20] S. Y. Itoga, "The string merging problem," *BIT Numer. Math.*, vol. 21, no. 1, pp. 20–30, Mar. 1981.

[21] A. Lenz, P. H. Siegel, A. Wachter-Zeh, and E. Yaakobi, "On the capacity of DNA-based data storage under substitution errors," in *Proc. Int. Conf. Vis. Commun. Image Process. (VCIP)*, Dec. 2021, pp. 1–5.

[22] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Sov. Phys.-Dokl.*, vol. 10, no. 8, pp. 707–710, 1966.

[23] V. I. Levenshtein, "Efficient reconstruction of sequences from their subsequences or supersequences," *J. Combinat. Theory A*, vol. 93, no. 2, pp. 310–332, 2001.

[24] Y. Liron and M. Langberg, "A characterization of the number of subsequences obtained via the deletion channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2300–2312, May 2015.

[25] M. Mitzenmacher, "A survey of results for deletion channels and related synchronization channels," *Probab. Surv.*, no. 6, pp. 1–33, Jan. 2009.

[26] M. Mitzenmacher and E. Drinea, "A simple lower bound for the capacity of the deletion channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4657–4660, Oct. 2006.

[27] L. Organick et al., "Random access in large-scale DNA data storage," *Nature Biotechnol.*, vol. 36, no. 3, pp. 242–248, Feb. 2018.

[28] M. Rahmati and T. M. Duman, "Upper bounds on the capacity of deletion channels using channel fragmentation," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 146–156, Jan. 2015.

[29] O. Sabary, Y. Orlev, R. L. S. Anavy, E. Yaakobi, and Z. Yakhini, "SOLQC: Synthetic oligo library quality control tool," *Bioinformatics*, vol. 37, pp. 720–722, Mar. 2021.

[30] F. Sala and L. Dolecek, "Counting sequences obtained from the synchronization channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 2925–2929.

[31] F. Sala, R. Gabrys, and L. Dolecek, "Gilbert–Varshamov-like lower bounds for deletion-correcting codes," *Proc. IEEE Inf. Theory Workshop*, Hobart, TAS, pp. 147–151, Nov. 2014.

[32] J. Sima and J. Bruck, "Optimal k-deletion correcting codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Paris, France, Jul. 2019, pp. 847–851.
[33] J. Sima, N. Raviv, and J. Bruck, "On coding over sliced information," *IEEE Trans. Inf. Theory*, vol. 67, no. 5, pp. 2793–2807, May 2021.
[34] S. K. Tabatabaei et al., "DNA punch cards for storing data on native DNA sequences via enzymatic nicking," *Nature Commun.*, vol. 11, no. 1, pp. 1–10, Apr. 2020.
[35] I. Tal, H. D. Pfister, A. Fazeli, and A. Vardy, "Polar codes for the deletion channel: Weak and strong polarization," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 1362–1366.
[36] G. Wang and Q. Wang, "On the size distribution of Levenshtein balls with radius one," 2022, *arXiv:2204.02201*.
[37] S. M. H. T. Yazdi, R. Gabrys, and O. Milenkovic, "Portable and error-free DNA-based data storage," *Sci. Rep.*, vol. 7, no. 1, p. 5011, Jul. 2017.

Daniella Bar-Lev (Student Member, IEEE) received the B.Sc. degree in computer science and mathematics and the M.Sc. degree in computer science from the Technion—Israel Institute of Technology, Haifa, Israel, in 2019 and 2021, respectively, where she is currently pursuing the Ph.D. degree with the Computer Science Department. Her research interests include algorithms, discrete mathematics, coding theory, and DNA storage.

Tuvi Etzion (Life Fellow, IEEE) was born in Tel Aviv, Israel, in 1956. He received the B.A., M.Sc., and D.Sc. degrees from the Technion—Israel Institute of Technology, Haifa, Israel, in 1980, 1982, and 1984, respectively.

In 1984, he held a position at the Department of Computer Science, Technion—Israel Institute of Technology, where he is currently the Bernard Elkin Chair in Computer Science. From 1985 to 1987, he was a Visiting Research Professor with the Department of Electrical Engineering–Systems, University of Southern California, Los Angeles, CA, USA. During the summers of 1990 and 1991, he was visiting Bellcore, Morristown, NJ, USA. From 1994 to 1996, he was a Visiting Research Fellow with the Computer Science Department, Royal Holloway University of London, Egham, England. He also had several visits to the Coordinated Science Laboratory, University of Illinois Urbana–Champaign from 1995 to 1998; two visits to HP Bristol during the summers of 1996 and 2000; a few visits to the Department of Electrical Engineering, University of California at San Diego, from 2000 to 2022; several visits to the Mathematics Department, Royal Holloway University of London, from 2007 to 2022; a few visits to the School of Physical and Mathematical Science (SPMS), Nanyang Technological University, and to the Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore, from 2016 to 2019, and 2022; and a few visits to Beijing Jiaotong University, Beijing, from 2017 to 2019. His research interests include applications of discrete mathematics to problems in computer science and information theory, coding theory, coding for storage, and combinatorial designs.

Dr. Etzion was an Associate Editor for Coding Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2006 to 2009. From 2004 to 2009, he was an Editor of the *Journal of Combinatorial Designs*. Since 2011, he has been an Editor of *Designs, Codes and Cryptography*, and an Editor of *Advances in Mathematics of Communications* since 2013. He has been the Editor-in-Chief of *Journal of Combinatorial Theory, Series A*, since 2021.

Eitan Yaakobi (Senior Member, IEEE) received the B.A. degree in computer science and mathematics and the M.Sc. degree in computer science from the Technion—Israel Institute of Technology, Haifa, Israel, in 2005 and 2007, respectively, and the Ph.D. degree in electrical engineering from the University of California at San Diego, San Diego, CA, USA, in 2011. From 2011 to 2013, he was a Post-Doctoral Researcher with the Department of Electrical Engineering, California Institute of Technology, and at the Center for Memory and Recording Research, University of California at San Diego. He is currently an Associate Professor at the Computer Science Department, Technion—Israel Institute of Technology. His research interests include information and coding theory with applications to non-volatile memories, associative memories, DNA storage, data storage and retrieval, and private information retrieval. He received the Marconi Society Young Scholar in 2009 and the Intel Ph.D. Fellowship in 2010–2011.