Upper Bounds on Multiple b-Burst Deletion-Correcting Codes

Chen Wang*, **Xiangliang Kong**†, **Eitan Yaakobi***, and **Tolga M. Duman**†
*Department of Computer Science, Technion – Israel Institute of Technology, Haifa 3200003, Israel
†Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey
Emails: cwang@campus.technion.ac.il, rongxlkong@gmail.com, yaakobi@cs.technion.ac.il, duman@ee.bilkent.edu.tr

***Abstract*—Motivated by their applications in DNA-based storage systems, codes capable of correcting consecutive deletions have attracted significant attention. An important class of such codes consists of those that can correct multiple consecutive deletion errors, which are commonly referred to as *multiple b-burst deletion-correcting codes*. In this paper, we investigate the fundamental limits of multiple b-burst deletion-correcting codes. Specifically, we first characterize several structural properties of the associated deletion balls, and then derive an upper bound on the largest size of multiple b-burst deletion-correcting codes, which improves upon previously known results for general parameter regimes.**

I. Introduction

Codes designed for channel with synchronization errors to correct errors arising from both insertions and deletions, have attracted significant attention due to their applications in wireless communication, disk and DNA-based data storage, racetrack memories, file synchronization, and compression [1–5].

As a key characteristic of DNA-based storage systems, data stored in DNA molecules is often corrupted by bursts of insertions and deletions [6], whereas substitution errors dominate in traditional optical or magnetic storage systems. Motivated by this observation, many works have focused on designing codes capable of correcting consecutive bursts of deletions and on exploring the fundamental limits of the corresponding code parameters, for examples, see [7–17].

In this paper, motivated by the same considerations, we study the maximal size of codes capable of correcting multiple bursts of deletions. Specifically, by exploring the structural properties of the deletion balls corresponding to t b-burst deletions (see Section I-A for the precise definition), we derive an upper bound on the maximum size of a code that can correct t b-burst deletions using the framework of Kulkarni and Kiyavash [18]. Our upper bound recovers existing results for the special case of $t = 1$ and improves the best-known bounds for the general case.

In the remainder of this section, we first briefly review related prior work and then present our results in detail.

A. Previous Work and Relevant Results

For an integer $q \geq 2$, let $\Sigma_q = \{0, 1, \ldots, q-1\}$ denote the q-ary alphabet. For positive integers t, b, and n satisfying $n \geq tb + 1$, and for a sequence $\mathbf{x} \in \Sigma_q^n$, we let $D_t^b(\mathbf{x})$ (respectively, $I_t^b(\mathbf{x})$) denote the set of all subsequences of $\mathbf{x}$ obtained by applying t b-burst deletions (respectively, b-burst insertions). We refer to $D_t^b(\mathbf{x})$ and $I_t^b(\mathbf{x})$ as the (t, b)-*burst-deletion ball* and (t, b)-*burst-insertion ball* centered at $\mathbf{x}$, respectively.

A code $\mathcal{C} \subseteq \Sigma_q^n$ is said to be a (t, b)-*burst-deletion-correcting code* if for any two distinct codewords $\mathbf{c}, \mathbf{c}' \in \mathcal{C}$, $D_t^b(\mathbf{c}) \cap D_t^b(\mathbf{c}') = \varnothing$. We denote by $M_q(n, (t, b))$ the largest size of a (t, b)-burst-deletion-correcting code in Σ_q^n.

The study of bounds on $M_q(n, (t, b))$ was initiated by Levenshtein [19], who determined the size of $D_1^b(\mathbf{x})$ as $|D_1^b(\mathbf{x})| = |U_b(\mathbf{x})|$, where for any $\mathbf{x} \in \Sigma_q^n$, the set $U_b(\mathbf{x})$ is defined as

$$U_b(\mathbf{x}) \triangleq \{i \in [n-b] : x_i \neq x_{i+b}\} \cup \{n-b+1\}.$$

When $b = 1$, the quantity $|U_b(\mathbf{x})|$ coincides with the number of runs in $\mathbf{x}$. Subsequently, based on the above characterization of $\left|D_1^b(\mathbf{x})\right|$, Schoeny *et al.* [8] derived an upper bound on the maximum size of binary $(1, b)$-burst-deletion-correcting codes. This result was later extended to general alphabets by Wang *et al.* [13]; see Table I for the explicit expression of these two bounds.

In a recent work [20], Lan *et al.* studied the sequence reconstruction problem over a channel subject to multiple bursts of insertions and deletions. Specifically, they proved that the size of the (t, b)-burst-insertion ball is independent of its center and provided an exact expression for its size. Based on this result, they further derived a general upper bound on $M_q(n, (t, b))$; see Table I for the explicit expressions.

Beyond the above upper bounds, several constructions for (t, b)-burst deletion codes have been proposed. Levenshtein [21] constructed binary $(1, 2)$-burst-deletion codes with redundancy at most $\log n + 1$. This was subsequently generalized by Cheng *et al.* [7] to $(1, b)$-burst-deletion codes with redundancy $b \log(n/b + 1)$, and further improved by Schoeny *et al.* [8] to $\log n + (b-1) \log \log n + O(1)$. More recently, Sun *et al.* [16] showed that for q-ary alphabets with $q \geq 2$, $(1, b)$-burst-deletion codes can achieve redundancy $\log n + O(1)$. In contrast, for $t \geq 2$, only a limited number of constructions are known. A general framework based on syndrome compression [11] combined with suitable precoding yields (t,b)-burst-deletion codes with redundancy at most $(4t-1)\log n + o(\log n)$ for all $q \geq 2$. For the special case $t=2$, the best known construction achieves a redundancy of $5\log n + o(\log n)$ [17].

B. Our Results

Our contributions in this paper consist of two parts. First, we study structural properties of the (t,b)-burst-deletion ball $D_t^b(\mathbf{x})$ for a general sequence $\mathbf{x} \in \Sigma_q^n$. In particular, we derive upper and lower bounds on $|D_t^b(\mathbf{x})|$ and establish a monotonicity property: for any sequence $\mathbf{x} \in D_1^b(\mathbf{z})$, it holds that $|D_t^b(\mathbf{x})| \leq |D_t^b(\mathbf{z})|$ (see Lemma 1). Then, based on these properties of (t,b)-burst-deletion balls, we adopt the framework introduced by Kulkarni and Kiyavash [18] to prove the following upper bound on $M_q(n,(t,b))$.

Theorem 1. *Let q,n,t,b be positive integers satisfying $q \geq 2$, $n \geq 2tb+(t-1)(b+1)$, and $b \mid n$. The maximum size of a (t,b)-burst-deletion-correcting code satisfies*

$$M_q(n,(t,b)) \leq \frac{t!q^{n-tb+t}}{(q-1)^t\left(n-2tb-\frac{(t-1)b}{q}\right)^t}(1+o(1)).$$

Theorem 1 implies that the redundancy of (t,b)-burst-deletion codes is at least $t\log_q n - O(1)$ for fixed q, t, b and $n \to \infty$. In Table I, we compare the bound in Theorem 1 with several known results. As shown in the table, it asymptotically coincides with the previous bounds obtained in [8, 13, 18, 19]. Moreover, our result applies to a strictly broader range of parameters than those considered in the above works. Besides, for fixed t, we have

$$\frac{q^{n+t}}{\sum_{i=0}^{t}\binom{n+t}{i}(q-1)^i} = \frac{q^{n+t}\,t!}{n^t(q-1)^t}\big(1+o(1)\big).$$

Consequently, when $t=1$, our bound matches the bound given in [20], whereas for $t \geq 2$, it is strictly tighter.

TABLE I: Upper bounds of $M_q(n,(t,b))$.

Upper bounds	Parameters	Reference
$\frac{2^t t!}{n^t}$	$q=2, b=1$	[19]
$\frac{t!q^n}{(q-1)^t n^t}$	$b=1$	[18]
$\frac{2^{n-b+1}-2^b}{n-2b+1}$	$q=2, t=1$	[8]
$\frac{q^{n-b+1}-q^b}{(q-1)(n-2b+1)}$	$t=1$	[13]
$\frac{q^{n+t}}{\sum_{i=0}^{t}\binom{n+t}{i}(q-1)^i}$	any q,t,b	[20]
$\frac{t!q^{n-tb+t}}{(q-1)^t\left(n-2tb-\frac{(t-1)b}{q}\right)^t}(1+o(1))$	any q,t,b	Theorem 1

The rest of the paper is organized as follows. In Section II, we introduce some auxiliary notations. In Section III, we present several results on the size of $D_t^b(\mathbf{x})$. Finally, in Section IV, we prove Theorem 1.

II. Notations

For integers m and n, the set $\{m, m+1, \ldots, n\}$ is denoted by $[m:n]$, and $[n]$ is used as shorthand for $[1:n]$. For a sequence $\mathbf{x} = (x_1,\ldots,x_n) \in \Sigma_q^n$ and a subset $R \subseteq [n]$, let $\mathbf{x}_R$ denote the restriction of $\mathbf{x}$ to the coordinates indexed by R. In particular, we define $\mathbf{x}_\emptyset$ to be the null vector, i.e., the vector of length zero. For two sequences $\mathbf{x}$ and $\mathbf{y}$, we denote by $\mathbf{xy}$ their concatenation.

For a positive integer b, we say that a sequence $\mathbf{y} \in \Sigma_q^{n-b}$ is obtained from $\mathbf{x} \in \Sigma_q^n$ by a *b-burst deletion*, if $\mathbf{y} = \mathbf{x}_{[n]\setminus[i:i+b-1]}$ for some $i \in [n-b+1]$. More generally, for integers $i_1, i_2, \ldots, i_t \in [1:n-b+1]$ satisfying $i_{j+1}-i_j \geq b$ for all $j \in [t-1]$, we define

$$\mathrm{Del}_b(\mathbf{x},\{i_1,\ldots,i_t\}) \triangleq \mathbf{x}_{[n]\setminus\bigcup_{j=1}^{t}[i_j:i_j+b-1]} \quad (1)$$

to be the subsequence obtained from $\mathbf{x}$ by applying t disjoint b-burst deletions at positions $i_1,\ldots,i_t$. Then, the (t,b)-burst-deletion ball $D_t^b(\mathbf{x})$ is defined as

$$D_t^b(\mathbf{x}) \triangleq \Big\{\mathrm{Del}_b(\mathbf{x},\{i_1,\ldots,i_t\}) : \ i_j \in [n-b+1], \ i_{j+1}-i_j \geq b, \ \forall j \in [t-1]\Big\}. \quad (2)$$

Given $\mathbf{y} \in D_t^b(\mathbf{x})$, we say that $\mathrm{Del}_b(\mathbf{x},\{i_1,\ldots,i_t\})$ is a *representation* of $\mathbf{y}$ if $\mathbf{y} = \mathrm{Del}_b(\mathbf{x},\{i_1,\ldots,i_t\})$.

Similarly, for a positive integer b, we denote by $I_t^b(\mathbf{x})$ the set of all supersequences of $\mathbf{x}$ obtained after t b-burst insertions, which we refer to as the (t,b)-*burst-insertion ball* centered at $\mathbf{x}$. A code $\mathcal{C} \subseteq \Sigma_q^n$ is said to be a (t,b)-*burst-insertion-correcting code* if for any two distinct codewords $\mathbf{c},\mathbf{c}' \in \mathcal{C}$, $I_t^b(\mathbf{c}) \cap I_t^b(\mathbf{c}') = \varnothing$. With the similar approach [19], it can be verified that a code is a (t,b)-burst-deletion-correcting code if and only if it is a (t,b)-burst-insertion-correcting code. Thereby, we restrict our discussion to (t,b)-burst-deletion-correcting codes in this work.

III. Bounds and Properties of (t,b)-burst-deletion balls

In this section, we first establish several properties of representations of sequences in $D_t^b(\mathbf{x})$. Using these properties, we then derive upper and lower bounds on $|D_t^b(\mathbf{x})|$ for any $\mathbf{x} \in \Sigma_q^n$, and prove a monotonicity property of $|D_t^b(\mathbf{x})|$.

To begin with, we introduce a compact representation for sequences in $D_t^b(\mathbf{x})$, which groups together consecutive b-burst deletions occurring at adjacent positions.

Consider a sequence $\mathrm{Del}_b(\mathbf{x},\{j_1,j_2,\ldots,j_t\}) \in D_t^b(\mathbf{x})$ satisfying $j_{\ell+1}-j_\ell \geq b$ for all $\ell \in [t-1]$. Whenever several deletions occur at positions of the form $i, i+b, \ldots, i+(c-1)b$ for some $c \geq 1$, we group them into a single block and represent them by the pair (i,c). Consequently, the deletion pattern $(j_1,j_2,\ldots,j_t)$ can be represented by an integer $s \in [t]$ with tuples $(i_1,i_2,\ldots,i_s) \in [n-b+1]^s$ and $(c_1,c_2,\ldots,c_s) \in \mathbb{Z}_+^s$ satisfying

$$i_{\ell+1} > i_\ell + c_\ell b, \ \forall 1 \leq \ell \leq s-1, \text{ and } \sum_{\ell=1}^{s} c_\ell = t, \quad (3)$$

as well as $\{j_1,\ldots,j_t\} = \bigcup_{\ell=1}^{s}\{i_\ell, i_\ell+b,\ldots,i_\ell+(c_\ell-1)b\}$. Then, we denote this sequence by $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s))$ and refer to this expression as the *compact representation* of $\mathrm{Del}_b(\mathbf{x},(j_1,j_2,\ldots,j_t))$.

Example 1. *Let $q=3$, $b=2$, and $t=2$. For $\mathbf{x}=(0,2,0,1,1,1,2,0)\in\Sigma_3^8$, we have*

$$\mathrm{Del}_2(\mathbf{x},1,4)=\mathrm{Del}_2(\mathbf{x},2,4)=(0,1,2,0).$$

In the first case, the two 2-burst deletions happen at positions 1 and 4 are separate, which gives that the compact representation is $\mathrm{Del}_2(\mathbf{x},(1,1),(4,1))$. In the second case, the two 2-burst deletions are consecutive, giving the compact representation $\mathrm{Del}_2(\mathbf{x},(2,2))$.

Example 1 shows that a sequence $\mathbf{y}\in D_t^b(\mathbf{x})$ may admit different compact representations. Suppose $\mathbf{y}\in D_t^b(\mathbf{x})$ with two compact representations

$$\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s))=\mathrm{Del}_b(\mathbf{x},(i'_1,c'_1),\ldots,(i'_s,c'_{s'})).$$

If $(i_\ell,c_\ell)=(i'_\ell,c'_\ell)$ for $\ell<a$, and either $i_a<i'_a$ or $i_a=i'_a$, $c_a<c'_a$, we say $\mathrm{Del}_b(\mathbf{x},(i'_1,c'_1),\ldots,(i'_{s'},c'_{s'}))$ *precedes* $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s))$.

Example 2. *Given $\mathbf{x}=(0,2,0,1,1,1,2,0)\in\Sigma_3^8$ be the same as in Example 1. Then, we have $\mathrm{Del}_2(\mathbf{x},(2,2))$ precedes $\mathrm{Del}_2(\mathbf{x},(1,1),(4,1))$.*

In the following, we select a canonical compact representation of $\mathbf{y}$, thereby eliminating ambiguity and ensuring that each sequence admits a unique representation.

Definition 1. *Given $\mathbf{x}\in\Sigma_q^n$ and $\mathbf{y}\in D_t^b(\mathbf{x})$, a compact representation $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s))$ of $\mathbf{y}$ is said to be* ***maximal****, if no other compact representation of $\mathbf{y}$ precedes it.*

It can be verified that the relation *precedes* defines a total order on the set of compact representations of any given sequence in $D_t^b(\mathbf{x})$. As a straightforward consequence, we have the following proposition.

Proposition 1. *For any $\mathbf{x}\in\Sigma_q^n$, let $\mathbf{u}$ and $\mathbf{v}$ be two sequences in $D_t^b(\mathbf{x})$ with maximal compact representations*

$$\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s)),\mathrm{Del}_b(\mathbf{x},(i'_1,c'_1),\ldots,(i'_{s'},c'_{s'})),$$

respectively. Then, $\mathbf{u}=\mathbf{v}$ if and only if $s=s'$ and $(i_\ell,c_\ell)=(i'_\ell,c'_\ell)$ for any $\ell\in[s]$.

Proposition 2. *Let $\mathbf{x}\in\Sigma_q^n$ and $\mathbf{y}\in D_t^b(\mathbf{x})$, and let $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s))$ be a compact representation of $\mathbf{y}$. Then, $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s))$ is maximal if and only if, for every $a\in[s]$,*

$$x_{i_a+c_ab}\neq x_{i_a+rb},\ \forall\ r\in[0:c_a-1]. \tag{4}$$

The condition is vacuous whenever $i_a+c_ab>n$.

Proof: We first prove the necessity by contradiction. Suppose that $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s))$ is not maximal, and let $\mathrm{Del}_b(\mathbf{x},(i'_1,c'_1),\ldots,(i'_s,c'_{s'}))$ be a compact representation of $\mathbf{y}$, which precedes the compact representation $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s))$.

Without loss of generality, let $a\in[s]$ be the smallest index such that either $i_a<i'_a$ or $i_a=i'_a$ and $c_a<c'_a$. Note that for all $\ell<a$, we have $(i_\ell,c_\ell)=(i'_\ell,c'_\ell)$. Let $d\triangleq\sum_{\ell<a}c_\ell b$ denote the total number of symbols deleted before position i_a. We next consider the following two cases.

If $i_a<i'_a$, we have $y_{i_a-d}=x_{i_a}$ by the representation $\mathrm{Del}_b(\mathbf{x},(i'_1,c'_1),\ldots,(i'_s,c'_{s'}))$ and $y_{i_a-d}=x_{i_a+c_ab}\neq x_{i_a}$ by the representation $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s))$, which is a contradiction.

Similarly, if $i_a=i'_a$ and $c_a<c'_a$, we have $y_{i_a-d}=x_{i_a+c'_ab}$ by the representation $\mathrm{Del}_b(\mathbf{x},(i'_1,c'_1),\ldots,(i'_s,c'_s))$ and $y_{i_a-d}=x_{i_a+c_ab}\neq x_{i_a+c'_ab}$ by the representation $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s))$, which is also a contradiction.

Now, we prove the sufficiency by contradiction. Suppose that there exists $a\in[s]$ and some $r\in[0:c_a-1]$ such that $x_{i_a+rb}=x_{i_a+c_ab}$. We next construct a new compact representation of $\mathbf{y}$ which precedes $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s))$.

For simplicity, we assume that $i_{a+1}>i_a+c_ab+1$, the case $i_{a+1}=i_a+c_ab+1$ follows similarly.

When $r\geq 1$, let $s'=s+1$ and $\{(i'_\ell,c'_\ell)\}_{\ell\in[s']}$ be as follows:

$$(i'_\ell,c'_\ell)=\begin{cases}(i_\ell,c_\ell), & 1\leq\ell\leq a-1,\\ (i_a,r), & \ell=a,\\ (i_a+rb+1,\ c_a-r), & \ell=a+1,\\ (i_{\ell-1},c_{\ell-1}), & a+2\leq\ell\leq s'.\end{cases} \tag{5}$$

Clearly, pairs $\{(i'_\ell,c'_\ell)\}_{\ell\in[s']}$ satisfy (3), which implies that $\mathrm{Del}_b(\mathbf{x},(i'_1,c'_1),\ldots,(i'_{s'},c'_{s'}))$ is a compact representation of some $\mathbf{y}'\in D_t^b(\mathbf{x})$. Moreover, the deleted positions of $\mathbf{y}'$ differ from those of $\mathbf{y}$ only on two intervals:

$$[i_a:i_a+rb-1]\text{ and }[i_a+rb+1:i_a+c_ab].$$

Recall that in $\mathbf{y}$, the a-th deleted interval is $[i_a:i_a+c_ab-1]$. Hence, $\mathbf{y}$ and $\mathbf{y}'$ can be written as

$$\begin{aligned}\mathbf{y}&=(\ldots,x_{i_a-1},\,x_{i_a+c_ab},\,x_{i_a+c_ab+1},\ldots),\\ \mathbf{y}'&=(\ldots,x_{i_a-1},\,x_{i_a+rb},\,x_{i_a+c_ab+1},\ldots).\end{aligned} \tag{6}$$

Since $x_{i_a+rb}=x_{i_a+c_ab}$, (6) implies that $\mathbf{y}=\mathbf{y}'$.

When $r=0$, let $s'=s$ and $\{(i'_\ell,c'_\ell)\}_{\ell\in[s']}$ be as follows:

$$(i'_\ell,c'_\ell)=\begin{cases}(i_\ell,c_\ell), & 1\leq\ell\leq a-1,\\ (i_a+1,c_a), & \ell=a,\\ (i_\ell,c_\ell), & a+1\leq\ell\leq s'.\end{cases}$$

Similarly, we can verify that $\mathrm{Del}_b(\mathbf{x},(i'_1,c'_1),\ldots,(i'_{s'},c'_{s'}))$ is a compact representation of $\mathbf{y}$.

In both cases, there is a new compact representation of $\mathbf{y}$ that precedes $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s))$, contradicting the assumption that $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),\ldots,(i_s,c_s))$ is maximal. This contradicts our assumption, completing the proof. ■

Example 3. *Given* $\mathbf{x}=(0,2,0,1,1,1,2,0)\in\Sigma_3^8$ *be the same as in Example 1. We have* $\mathbf{y}\triangleq\mathrm{Del}_2(\mathbf{x},(2,2))=(0,1,2,0)$.

Since $x_4=x_6$, $\mathrm{Del}_2(\mathbf{x},(2,2))$ *is not maximal. By shifting the deleted intervals, the maximal compact representation of* $\mathbf{y}$ *is* $\mathrm{Del}_2(\mathbf{x},(2,1),(5,1))$*, which precedes* $\mathrm{Del}_2(\mathbf{x},(2,2))$.

Next, based on the above characterization of maximal compact representations of sequences in $D_t^b(\mathbf{x})$, we prove the following bounds on $|D_t^b(\mathbf{x})|$ in the spirit of [19].

Theorem 2. *For any* $\mathbf{x}\in\Sigma_q^n$*, it holds that*

$$\binom{|U_b(\mathbf{x})|-(t-1)b}{t}\le|D_t^b(\mathbf{x})|\le\binom{|U_b(\mathbf{x})|+t-1}{t}.$$

Proof: By Propositions 2 and 1, it suffices to show that, for any $\mathbf{x}\in\Sigma_q^n$, the number of sets of pairs $\{(i_j,c_j)\}_{j\in[s]}$ satisfying (3) and (4) satisfies the stated bounds.

Note that, by the definition of $U_b(\mathbf{x})$, the following collection of pair sets

$$\left\{\{(i_j,1)\}_{j\in[t]}:i_j\in U_b(\mathbf{x}),i_{j+1}>i_j+b,\forall\ j\in[t-1]\right\}\quad(7)$$

satisfies the conditions (3) and (4). Hence, the lower bound follows directly, since there are at least $\binom{|U_b(\mathbf{x})|-(t-1)b}{t}$ different choices of $(i_1,i_2,\ldots,i_t)$ satisfying (7).

Next, we proceed to prove the upper bound. Observe that each set of pairs $\{(i_j,c_j)\}_{j\in[s]}$ satisfying (3) and (4) is uniquely associated with a multiset of indices of the form

$$\{\underbrace{i_1,\ldots,i_1}_{c_1},\underbrace{i_2,\ldots,i_2}_{c_2},\ldots,\underbrace{i_s,\ldots,i_s}_{c_s}\},\qquad(8)$$

where $i_j\in U_b(\mathbf{x})$, $i_{j+1}>i_j+c_jb$, and $\sum_{j\in[s]}c_j=t$. Hence, the number of sets of pairs $\{(i_j,c_j)\}_{j\in[s]}$ satisfying (3) and (4) is upper bounded by the number of multisets of the form (8), with $(i_1,\ldots,i_s)\in U_b(\mathbf{x})^s$ and $(c_1,\ldots,c_s)\in\mathbb{Z}_+^s$ satisfying

$$i_j\ne i_{j'},\ \text{for all } j\ne j',\quad\text{and}\quad\sum_{j\in[s]}c_j=t.$$

The number of such multisets is given by $\binom{|U_b(\mathbf{x})|+t-1}{t}$, which concludes the proof of the upper bound. ■

Next, we establish the following property of $|D_t^b(\mathbf{x})|$.

Lemma 1. *For any sequences* $\mathbf{z}\in\Sigma_q^{n+b}$ *and* $\mathbf{x}\in D_1^b(\mathbf{z})\subseteq\Sigma_q^n$*, it holds that* $|D_t^b(\mathbf{x})|\le|D_t^b(\mathbf{z})|$.

Proof: Let $\mathbf{x}=\mathbf{u}\mathbf{v}$ and $\mathbf{z}=\mathbf{u}\boldsymbol{\sigma}\mathbf{v}$ for some $\mathbf{u},\mathbf{v}\in\Sigma_q^*$ and $\boldsymbol{\sigma}\in\Sigma_q^b$. We aim to construct an injective map ϕ from $D_t^b(\mathbf{x})$ to $D_t^b(\mathbf{z})$. The result then follows directly.

By Propositions 2 and 1, every sequence $\mathbf{y}\in D_t^b(\mathbf{x})$ has a unique maximal compact representation $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),(i_2,c_2),\ldots,(i_s,c_s))$, and the same property holds for every sequence in $D_t^b(\mathbf{z})$. Next, we construct a map ϕ from $D_t^b(\mathbf{x})$ to $D_t^b(\mathbf{z})$ that preserves the maximality of compact representations. The injectivity of ϕ then follows directly from the uniqueness of the maximal compact representation. The construction of ϕ proceeds in two separate cases.

Case 1: For all $1\le\ell\le s$, either $i_\ell+c_\ell b\le|\mathbf{u}|$ or $i_\ell>|\mathbf{u}|$.

In Case 1, we define $\phi(\mathbf{y})\triangleq\mathrm{Del}_b(\mathbf{z},(i'_1,c_1),(i'_2,c_2),\ldots,(i'_s,c_s))$, where

$$i'_\ell=\begin{cases}i_\ell, & \text{if } i_\ell\le|\boldsymbol{u}|-c_\ell b,\\ i_\ell+b, & \text{if } i_\ell>|\boldsymbol{u}|.\end{cases}\qquad(9)$$

Clearly, pairs $\{(i'_\ell,c'_\ell)\}_{\ell\in[s]}$ satisfy (3) and for each $\ell\in[s]$, it also holds that either $i'_\ell+c'_\ell b\le|\boldsymbol{u}|$ or $i'_\ell>|\boldsymbol{u}|$. Moreover, since $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),(i_2,c_2),\ldots,(i_s,c_s))$ is maximal, it holds that $x_{i_\ell+c_\ell b}\ne x_{i_\ell+rb}$, for all $r\in[0:c_\ell-1]$ and $\ell\in[s]$. Thus, by writing $\mathbf{z}=\mathbf{u}\boldsymbol{\sigma}\mathbf{v}$, we have

$$z_i=\begin{cases}x_i, & 1\le i\le|\mathbf{u}|,\\ \sigma_{i-|\mathbf{u}|}, & |\mathbf{u}|+1\le i\le|\mathbf{u}|+b,\\ x_{i-b}, & |\mathbf{u}|+b+1\le i\le n+b,\end{cases}\qquad(10)$$

which, combined with (9), implies that

$$z_{i'_\ell+c_\ell b}\ne z_{i'_\ell+rb},\ \forall\,r\in[0:c_\ell-1],\ \forall\,\ell\in[s].$$

Case 2: There exists some $\ell_0\in[s]$ such that $i_{\ell_0}\le|\mathbf{u}|<i_{\ell_0}+c_{\ell_0}b$.

Let $r_0\in[c_{\ell_0}]$ be the minimum integer such that $|\mathbf{u}|<i_{\ell_0}+r_0b\le|\mathbf{u}|+b$. Let $\lambda\in[b]$ be such that $|\mathbf{u}|+\lambda=i_{\ell_0}+r_0b$. Next, we consider the following two subcases.

Case 2.1: $x_{i_{\ell_0}+c_{\ell_0}b}\ne\sigma_\lambda$.

In Case 2.1, we define $\phi(\mathbf{y})\triangleq\mathrm{Del}_b(\mathbf{z},(i'_1,c_1),(i'_2,c_2),\ldots,(i'_s,c_s))$, where

$$i'_\ell=\begin{cases}i_\ell, & \text{if } i_\ell\le|\mathbf{u}| \text{ and } \ell<\ell_0,\\ i_\ell+b, & \text{if }\ \ell=\ell_0 \text{ or } i_\ell>|\mathbf{u}|.\end{cases}\qquad(11)$$

Clearly, pairs $\{(i'_\ell,c'_\ell)\}_{\ell\in[s]}$ satisfy (3). Then, by (10) and the maximality of $\mathrm{Del}_b(\mathbf{x},(i_1,c_1),(i_2,c_2),\ldots,(i_s,c_s))$, we have $z_{i'_\ell+c_\ell b}\ne z_{i'_\ell+rb}$, for all $r\in[0:c_\ell-1]$, $\ell\in[s]\setminus\{\ell_0\}$. Moreover, by the definitions of r_0 and λ, (10) implies that

$$z_{i'_{\ell_0}+rb}=z_{i_{\ell_0}+rb+b}=\begin{cases}x_{i_{\ell_0}+(r+1)b}, & \text{if } 0\le r<r_0-1,\\ \sigma_\lambda, & \text{if } r=r_0-1,\\ x_{i_{\ell_0}+rb}, & \text{if } r_0\le r\le c_{\ell_0}.\end{cases}$$

Thus, since $x_{i_{\ell_0}+c_{\ell_0}b}\ne\sigma_\lambda$, we obtain $z_{i'_{\ell_0}+c_{\ell_0}b}\ne z_{i'_{\ell_0}+rb}$, for all $r\in[0:c_{\ell_0}-1]$.

Case 2.2: $x_{i_{\ell_0}+c_{\ell_0}b}=\sigma_\lambda$.

In Case 2.2, we define

$$\begin{aligned}\phi(\mathbf{y})\triangleq&\mathrm{Del}_b\big(\mathbf{z},(i'_1,c_1),\ldots,(i'_{\ell_0-1},c_{\ell_0-1}),(i'_{\ell_0},c'_{\ell_0}),\\&(i''_{\ell_0},c''_{\ell_0}),(i'_{\ell_0+1},c_{\ell_0+1}),\ldots,(i'_s,c_s)\big)\end{aligned}$$

where $(i'_{\ell_0}, c'_{\ell_0}) = (i_{\ell_0}, r_0)$ and $(i''_{\ell_0}, c''_{\ell_0}) = (i_{\ell_0} + (r_0 + 1)b, c_{\ell_0} - r_0)$ and

$$i'_\ell = \begin{cases} i_\ell, & \text{if } i_\ell \leq |\mathbf{u}| \text{ and } \ell < \ell_0, \\ i_\ell + b, & \text{if } i_\ell > |\mathbf{u}|. \end{cases}$$

Clearly, pairs $\{(i'_\ell, c_\ell)\}_{\ell \in [s] \setminus \{\ell_0\}} \cup \{(i'_{\ell_0}, c'_{\ell_0}), (i''_{\ell_0}, c''_{\ell_0})\}$ satisfies (3).

By (10) and the maximality of $\mathrm{Del}_b(\mathbf{x}, (i_1, c_1), (i_2, c_2), \ldots, (i_s, c_s))$, we have $z_{i'_\ell + c_\ell b} \neq z_{i'_\ell + rb}$, for all $r \in [0 : c_\ell - 1]$, $\ell \in [s] \setminus \{\ell_0\}$. Moreover, by the definitions of r_0 and λ, (10) implies that

$$z_{i'_{\ell_0} + rb} = \begin{cases} x_{i_{\ell_0} + rb}, & \text{if } 0 \leq r < r_0, \\ \sigma_\lambda, & \text{if } r = r_0. \end{cases} \tag{12}$$

and $z_{i''_{\ell_0} + rb} = x_{i_{\ell_0} + (r + r_0)b}$ for all $r \in [0 : c_{\ell_0} - r_0]$. Since $x_{i_{\ell_0} + c_{\ell_0} b} = \sigma_\lambda$, we have $z_{i'_{\ell_0} + c'_{\ell_0} b} = x_{i_{\ell_0} + c_{\ell_0} b} \neq z_{i'_{\ell_0} + rb}$ for all $r \in [0 : c'_{\ell_0} - 1]$ and $z_{i''_{\ell_0} + c''_{\ell_0} b} = x_{i_{\ell_0} + c_{\ell_0} b} \neq z_{i''_{\ell_0} + rb}$ for all $r \in [0 : c''_{\ell_0} - 1]$ by $x_{i_{\ell_0} + c_{\ell_0} b} \neq x_{i_{\ell_0} + rb}$ for all $r \in [0 : c_{\ell_0} - 1]$.

To sum up, for all two cases, we show that the defined map ϕ preserves the maximality of the compact representation of the corresponding sequence. This completes the proof. ∎

Example 4. *Given $\mathbf{x} = (0, 2, 0, 1, 1, 1, 2, 0) \in \Sigma_3^8$ and $\mathbf{z} = (0, 2, 0, 1, 1, 1, 0, 2, 2, 0) \in \Sigma_3^{10}$, we have $\mathbf{x} \in D_1^b(\mathbf{z})$. Let $\mathbf{u} = (0, 2, 0, 1, 1, 1)$, $\sigma = (0, 2)$, and $\mathbf{v} = (2, 0)$, we have $\mathbf{x} = \mathbf{u}\mathbf{v}$ and $\mathbf{z} = \mathbf{u}\boldsymbol{\sigma}\mathbf{v}$.*

Let $t = 2$, $\mathbf{y}_1 = \mathrm{Del}_2(\mathbf{x}, (2, 1), (7, 1)) = (0, 1, 1, 1)$, $\mathbf{y}_2 = \mathrm{Del}_2(\mathbf{x}, (4, 2)) = (0, 2, 0, 0)$, and $\mathbf{y}_3 = \mathrm{Del}_2(\mathbf{x}, (3, 2)) = (0, 2, 2, 0)$. Denote ϕ as the map in the proof of Lemma 1, the following holds:

- *As the deletion interval is disjoint with $[|\mathbf{u}| : |\mathbf{u}| + b] = [6 : 8]$, we have $\phi(\mathbf{y}_1) = \mathrm{Del}_2(\mathbf{z}, (2, 1), (9, 1)) = (0, 1, 1, 1)$ (Case 1);*
- *As $x_4, x_6 \neq \sigma_2$, we have $\phi(\mathbf{y}_2) = \mathrm{Del}_2(\mathbf{z}, (4, 2)) = (0, 2, 0, 2, 2, 0)$ (Case 2.1);*
- *As $x_3 = \sigma_1$, we have $\phi(\mathbf{y}_3) = \mathrm{Del}_2(\mathbf{z}, (5, 2)) = (0, 2, 0, 1, 2, 0)$ (Case 2.2).*

IV. Upper bounds on $M_q(n, (t, b))$

In this section, we present the proof of Theorem 1. As in [18], we model the problem of finding the largest (t, b)-burst-deletion-correcting code as a matching problem on a hypergraph. We then prove Theorem 1 by constructing a feasible solution to the dual linear program corresponding to this hypergraph matching formulation.

Consider the hypergraph $\mathcal{H}_{q,n,t}^b$ with vertex set Σ_q^{n-tb} and edge set $\{D_t^b(\mathbf{x}) : \mathbf{x} \in \Sigma_q^n\}$. That is, each vertex of $\mathcal{H}_{q,n,t}^b$ is a sequence of length $n - tb$ over Σ_q, and a collection of vertices forms a hyperedge if and only if it coincides with $D_t^b(\mathbf{x})$ for some $\mathbf{x} \in \Sigma_q^n$. Then, a (t, b)-burst-deletion correcting code in Σ^n corresponds to a matching in $\mathcal{H}_{q,n,t}^b$ and hence we have $M_q(n, (t, b)) = \nu(\mathcal{H}_{q,n,t}^b)$, where $\nu(\mathcal{H}_{q,n,t}^b)$ denote the matching number of $\mathcal{H}_{q,n,t}^b$. Thus, we have

$$\begin{aligned} M_q(n, (t, b)) = \text{maximize} \quad & \sum_{\mathbf{x} \in \Sigma_q^n} z(\mathbf{x}) \\ \text{subject to} \quad & \sum_{\mathbf{x} \in I_t^b(\mathbf{y})} z(\mathbf{x}) \leq 1, \ \forall \ \mathbf{y} \in \Sigma_q^{n-tb}; \\ & z(\mathbf{x}) \in \mathbb{Z}^+, \ \forall \ \mathbf{x} \in \Sigma_q^n. \end{aligned} \tag{13}$$

Since the feasible regions of the integer programs are strictly contained in the feasible regions of their of the linear programming relaxations, by the Duality Theorem of linear programming (see [22, Corollary 7.1g]), $M_q(n, (t, b))$ is upper bounded by

$$\begin{aligned} \text{minimize} \quad & \sum_{\mathbf{y} \in \Sigma_q^{n-tb}} w(\mathbf{y}) \\ \text{subject to} \quad & \sum_{\mathbf{y} \in D_t^b(\mathbf{x})} w(\mathbf{y}) \geq 1, \ \forall \ \mathbf{x} \in \Sigma_q^n; \\ & w(\mathbf{y}) \geq 0, \ \forall \ \mathbf{y} \in \Sigma_q^{n-tb}. \end{aligned} \tag{14}$$

Theorem 3. *Let q, n, t, b be positive integers satisfying $q \geq 2$ and $n \geq tb + 1$. Then, it holds that*

$$M_q(n, (t, b)) \leq \sum_{\mathbf{y} \in \Sigma_q^{n-tb}} |D_t^b(\mathbf{y})|^{-1}.$$

Proof: It suffices to show that $w(\mathbf{y}) = |D_t^b(\mathbf{y})|^{-1}$, for any $\mathbf{y} \in \Sigma_q^{n-tb}$, is a feasible solution for the dual LP problem (14).

Clearly, $w(\mathbf{y}) \geq 0$. Moreover, for any $\mathbf{x} \in \Sigma_q^n$,

$$\begin{aligned} \sum_{\mathbf{y} \in D_t^b(\mathbf{x})} w(\mathbf{y}) &= \sum_{\mathbf{y} \in D_t^b(\mathbf{x})} |D_t^b(\mathbf{y})|^{-1} \\ &\geq \sum_{\mathbf{y} \in D_t^b(\mathbf{x})} |D_t^b(\mathbf{x})|^{-1} = 1, \end{aligned}$$

where the second inequality follows since $|D_t^b(\mathbf{y})| \leq |D_t^b(\mathbf{x})|$ by Lemma 1. ∎

Then, by substituting the lower bound on $|D_t^b(x)|$ in Theorem 2, we can obtain the upper bound on $M_q(n, (t, b))$ in Theorem 1. The detailed calculation is left in the appendix.

V. Acknowledgement

This research was partially funded by the European Union (DiDAX, 101115134) and the ERC Advanced Grant 101054904: TRANCIDS. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. This research was also partially supported in part by the Israel Science Foundation (ISF) Grant 2462/24. The research of C. Wang was supported in part at the Technion by a fellowship from the Lady Davis Foundation.

References

[1] H. Mercier, V. K. Bhargava, and V. Tarokh, "A survey of error-correcting codes for channels with symbol synchronization errors," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 1, pp. 87–96, 2010.

[2] S. M. H. T. Yazdi, H. M. Kiah, E. Garcia-Ruiz, J. Ma, H. Zhao, and O. Milenkovic, "DNA-based storage: Trends and methods," *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, vol. 1, no. 3, pp. 230–248, 2015.

[3] J. Bornholt, R. Lopez, D. M. Carmean, L. Ceze, G. Seelig, and K. Strauss, "A DNA-based archival storage system," in *Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems*, 2016, pp. 637–649.

[4] S. H. T. Yazdi, R. Gabrys, and O. Milenkovic, "Portable and error-free DNA-based data storage," *Scientific reports*, vol. 7, no. 1, p. 5011, 2017.

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

[6] H. H. Lee, R. Kalhor, N. Goela, J. Bolot, and G. M. Church, "Terminator-free template-independent enzymatic DNA synthesis for digital information storage," *Nature communications*, vol. 10, no. 1, p. 2383, 2019.

[7] L. Cheng, T. G. Swart, H. C. Ferreira, and K. A. Abdel-Ghaffar, "Codes for correcting three or more adjacent deletions or insertions," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 1246–1250.

[8] C. Schoeny, A. Wachter-Zeh, R. Gabrys, and E. Yaakobi, "Codes correcting a burst of deletions or insertions," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 1971–1985, 2017.

[9] C. Schoeny, F. Sala, and L. Dolecek, "Novel combinatorial coding results for DNA sequencing and data storage," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2017, pp. 511–515.

[10] T. Saeki and T. Nozaki, "An improvement of non-binary code correcting single b-burst of insertions or deletions," in *2018 International Symposium on Information Theory and Its Applications (ISITA)*. IEEE, 2018, pp. 6–10.

[11] J. Sima, R. Gabrys, and J. Bruck, "Syndrome compression for optimal redundancy codes," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 751–756.

[12] A. Lenz and N. Polyanskii, "Optimal codes correcting a burst of deletions of variable length," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 757–762.

[13] S. Wang, Y. Tang, J. Sima, R. Gabrys, and F. Farnoud, "Non-binary codes for correcting a burst of at most t deletions," *IEEE Transactions on Information Theory*, vol. 70, no. 2, pp. 964–979, 2024.

[14] T. T. Nguyen, K. Cai, and P. H. Siegel, "A new version of q-ary varshamov-tenengolts codes with more efficient encoders: the differential vt codes and the differential shifted vt codes," *IEEE Transactions on Information Theory*, vol. 70, no. 10, pp. 6989–7004, 2024.

[15] Y. Sun and G. Ge, "Codes for correcting a burst of edits using weighted-summation vt sketch," *IEEE Transactions on Information Theory*, 2025.

[16] Y. Sun, Z. Lu, Y. Zhang, and G. Ge, "Asymptotically optimal codes for (t, s)-burst error," *IEEE Transactions on Information Theory*, 2025.

[17] Z. Ye, Y. Sun, W. Yu, G. Ge, and O. Elishco, "Codes correcting two bursts of exactly b deletions," *IEEE Transactions on Information Theory*, 2025.

[18] A. A. Kulkarni and N. Kiyavash, "Nonasymptotic upper bounds for deletion correcting codes," *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 5115–5130, 2013.

[19] V. I. Levenshtein *et al.*, "Binary codes capable of correcting deletions, insertions, and reversals," in *Soviet physics doklady*, vol. 10, no. 8. Soviet Union, 1966, pp. 707–710.

[20] Z. Lan, Y. Sun, W. Yu, and G. Ge, "Sequence reconstruction under channels with multiple bursts of insertions or deletions," *IEEE Transactions on Information Theory*, vol. 72, no. 1, pp. 315–330, 2026.

[21] V. Levenshtein, "Asymptotically optimum binary code with correction for losses of one or two adjacent bits," *Problemy Kibernetiki*, vol. 19, pp. 293–298, 1967.

[22] A. Schrijver, *Theory of linear and integer programming*. John Wiley & Sons, 1998.

APPENDIX.A

We present detailed estimates of

$$\sum_{\mathbf{y}\in\Sigma_q^{n-tb}} |D_t^b(\mathbf{y})|^{-1},$$

which were omitted in Section IV, and which complete the proof of the upper bound on $M_q(n,(t,b))$.

Note that

$$\sum_{\mathbf{y}\in\Sigma_q^{n-tb}} |D_t^b(\mathbf{y})|^{-1} = \sum_{r=1}^{n-(t+1)b+1} \sum_{\substack{\mathbf{y}\in\Sigma_q^{n-tb}\\ |U_b(\mathbf{y})|=r}} |D_t^b(\mathbf{y})|^{-1}.$$

Thus, by the lower bound on $|D_t^b(\mathbf{y})|$ in Theorem 2, we have

$$\begin{aligned}
&\sum_{\mathbf{y}\in\Sigma_q^{n-tb}} |D_t^b(\mathbf{y})|^{-1}\\
\le &\sum_{r=r'+1}^{n-(t+1)b+1} \sum_{\substack{\mathbf{y}\in\Sigma_q^{n-tb}\\ |U_b(\mathbf{y})|=r}} \binom{r-(t-1)b}{t}^{-1}\\
&+\sum_{r=1}^{r'} \sum_{\substack{\mathbf{y}\in\Sigma_q^{n-tb}\\ |U_b(\mathbf{y})|=r}} 1\\
= &\sum_{r=r'+1}^{n-(t+1)b+1} \frac{\left|\{\mathbf{y}\in\Sigma_q^{n-tb} : |U_b(\mathbf{y})|=r\}\right|}{\binom{r-(t-1)b}{t}}\\
&+\sum_{r=1}^{r'} \left|\{\mathbf{y}\in\Sigma_q^{n-tb} : |U_b(\mathbf{y})|=r\}\right|\\
=&q^b \sum_{r=r'+1}^{n-(t+1)b+1} (q-1)^{r-1} \frac{\binom{n-tb-b}{r-1}}{\binom{r-(t-1)b}{t}} \qquad (15)\\
&+q^b \sum_{r=1}^{r'} (q-1)^{r-1}\binom{n-tb-b}{r-1}, \qquad (16)
\end{aligned}$$

where $r' \ge (t-1)(b+1)$ is a parameter to be determined later, and the last equality follows from

$$\left|\{\mathbf{y}\in\Sigma_q^{n-tb} : |U_b(\mathbf{y})|=r\}\right| = \binom{n-tb-b}{r-1} q^b (q-1)^{r-1},$$

see Claim 4 in Appendix A of [13] for a detailed proof.

To further estimate (15) and (16), we consider the following binomial random variable $X \sim \text{Binomial}(N,\theta)$ with parameters $N = n-(t+1)b$ and $\theta = \frac{q-1}{q}$. The expectation of X is $\mu \triangleq \mathbb{E}(X) = N\theta$ and the variance is $\text{Var}(X) = N\theta(1-\theta)$. Moreover, by the Chernoff bound, it holds that

$$\Pr(X \ge (1-\epsilon)\mu) \le e^{\left(-\frac{\epsilon^2\mu}{2}\right)}.$$

The term in (16) can be simplified as

$$\begin{aligned}
&q^b \sum_{r=0}^{r'-1} (q-1)^r \binom{N}{r}\\
&= q^{N+b} \sum_{r=0}^{r'-1} \binom{N}{r}\left(1-\frac{1}{q}\right)^r \left(\frac{1}{q}\right)^{N-r}\\
&= q^{N+b} \Pr(X < r'). \qquad (17)
\end{aligned}$$

Moreover, since $\binom{r-(t-1)b}{t}$ is monotone increasing in r, thus the term in (15) can be simplified as

$$\begin{aligned}
&q^b \sum_{r=r'+1}^{n-(t+1)b+1} (q-1)^{r-1} \frac{\binom{n-tb-b}{r-1}}{\binom{r-(t-1)b}{t}}\\
\le &\frac{q^b}{\binom{r'+1-(t-1)b}{t}} \sum_{r=r'+1}^{n-(t+1)b+1} (q-1)^{r-1}\binom{n-tb-b}{r-1}\\
= &\frac{q^b}{\binom{r'+1-(t-1)b}{t}} \sum_{r=r'}^{N} (q-1)^r \binom{N}{r}\\
\le &\frac{q^{N+b}}{\binom{r'+1-(t-1)b}{t}}, \qquad (18)
\end{aligned}$$

where the last inequality follows by $\sum_{r=r'}^{N}(q-1)^r\binom{N}{r} \le q^N$.

Setting $r' = \mu - \sqrt{2tN\ln N}$ in (18) and (17), then by the Chernoff bound, the RHS of (17) is at most

$$\begin{aligned}
q^{N+b}\cdot e^{-\frac{tN\ln N}{\mu}} &\le \frac{q^{N+b}}{N^{\frac{tq}{q-1}}}\\
&= o\left(\frac{q^{n-tb}}{(n-(t+1)b)^t}\right),
\end{aligned}$$

as $n\to\infty$. Meanwhile, by

$$\begin{aligned}
\binom{r'+1-(t-1)b}{t} &\ge \frac{(\mu-(t-1)b)^t}{t!}(1-o(1))\\
&= \left(\frac{q-1}{q}\right)^t \left(n-2tb-\frac{(t-1)b}{q}\right)^t \frac{1}{t!}(1-o(1)) \qquad (19)
\end{aligned}$$

as $n\to\infty$, the RHS of (18) is at most

$$\frac{q^{N+b+t}t!}{(q-1)^t\left(n-2tb-\frac{(t-1)b}{q}\right)^t}(1+o(1)).$$

In total, this leads to

$$\begin{aligned}
&\sum_{\mathbf{y}\in\Sigma_q^{n-tb}} |D_t^b(\mathbf{y})|^{-1}\\
\le &\frac{t!q^{n-tb+t}}{(q-1)^t\left(n-2tb-\frac{(t-1)b}{q}\right)^t}(1+o(1))
\end{aligned}$$

and confirms the upper bound in Theorem 1.