Sequence Reconstruction over the Deletion Channel

Ryan Gabrys
Code 53221
Spawar Systems Center, San Diego
San Diego, CA, 92115
ryan.gabrys@navy.mil

Eitan Yaakobi
Computer Science Department
Technion - Israel Institute of Technology
Haifa 32000, Israel
yaakobi@cs.technion.ac.il

***Abstract*—The *sequence-reconstruction problem*, first proposed by Levenshtein, models a setup in which a sequence from some set is transmitted over several independent channels, and the decoder receives the outputs from every channel. The main problem of interest is to determine the minimum number of channels required to reconstruct the transmitted sequence. In the combinatorial context, the problem is equivalent to finding the maximum intersection between two balls of radius t where the distance between their centers is at least d. The setup of this problem was studied before for several error metrics such as the Hamming metric, the Kendall-tau metric, and the Johnson metric.**

In this paper, we extend the study initiated by Levenshtein for reconstructing sequences over the deletion channel. While he solved the case where the transmitted word can be arbitrary, we study the setup where the transmitted word belongs to a single-deletion-correcting code and there are t deletions in every channel. Under this paradigm, we study the minimum number of different channel outputs in order to construct a successful decoder.

I. Introduction

The *sequence reconstruction problem* was first introduced by Levenshtein in [11] and [12]. Under this paradigm he studied the minimum number of different (noisy) channels that is required in order to reconstruct a transmitted sequence. In [11], Levenshtein showed that the number of channels required to recover such a sequence has to be greater than the maximum intersection between the error balls of any two transmitted sequences. This problem was first motivated from the fields of biology and chemistry, however it is also very relevant for applications in wireless sensor networks where a collection of nodes, each with partial information about the operating environment, are trying to form a *common operational picture* (*COP*). In this case each of the received subsequences represents the information available to that node.

Mathematically speaking, let $\mathcal{C}$ be a code over a space V with a distance metric $\rho : V \times V \to \mathbb{N}$. Assume that its minimum distance is d and there are at most t errors in every channel, where $t > (d-1)/2$. Then the problem of calculating the value of

$$\max_{\boldsymbol{x}_1,\boldsymbol{x}_2 \in V \rho(\boldsymbol{x}_1,\boldsymbol{x}_2)\geq d} \{|B_t(\boldsymbol{x}_1) \cap B_t(\boldsymbol{x}_2)|\}, \quad (1)$$

where $B_t(\boldsymbol{x})$ is the ball of radius r surrounding $\boldsymbol{x}$, is referred to as the *reconstruction problem.*

Solving the reconstruction problem from (1) was studied in [11] with respect to several channels such as the Hamming distance, Johnson graphs and other metric distances. In [8], [9], [10], it was analyzed for permutations, and in [13], [14] for other general error graphs. Recently, the problem was studied in [19] for permutations with the Kendall's τ distance and the Grassmann graph, and in [16] for insertions. The connection between the reconstruction problem and associative memories was proposed in [18]. This problem was also studied in [6] for the purpose of asymptotically improving the Gilbert-Varshamov bound.

Solving the reconstruction problem for the deletion channel has received a significant attention in the literature. In fact, it was one of the first models Levenshtein studied in [12] for insertions and deletions along with reconstruction algorithms. The design of such algorithms for the probabilistic model of the reconstruction problem was also studied in [1], [7], [17] and only for deletions in [4], [5].

In this work, we consider the combinatorial reconstruction problem for the deletion channel. While Levenshtein assumed in [11], [12] that the transmitted words are arbitrary, we assume here that their *Levenshtein distance* is at least two and study the minimum number of different channel outputs in order to correct t deletions in every channel. Our main result is showing that for $t \leq \frac{n}{2}$ and $n \geq 10$, this value is given by

$$N(n,t) \triangleq 2D(n-4,t-2)+2D(n-5,t-2)+2D(n-7,t-2) + D(n-6,t-3) + D(n-7,t-3),$$

where $D(m,s)$ is the largest size of a deletion ball for words of length m with s deletions. For example, we get that $N(n,2) = 6$. That is, if the transmitted codeword belongs to a single-deletion-correcting code and two deletions occurred in every channel, then 7 channels are sufficient to construct a successful decoder.

The rest of this paper is organized as follows. In Section II, we introduce our notation and establish some preliminary results. In Section III, we show that $N(n,t)$ is a lower bound, that is we find two sequences of Levenshtein distance two such that the intersection size of their deletion balls is $N(n,t)$. In Section IV, we show that $N(n,t)$ is also an upper bound, thereby establishing equality. Due to the lack of space some of the proofs will be omitted.

II. Definitions and Preliminaries

We denote by $\mathbb{Z}_2$ the set $\{0,1\}$. Let $\boldsymbol{x}$ be a length-n binary vector in $\mathbb{Z}_2^n$. A vector $\boldsymbol{y} \in \mathbb{Z}_2^{n-t}$ is the outcome of t deletions from $\boldsymbol{x}$ if $\boldsymbol{y}$ is a subsequence of $\boldsymbol{x}$. The *deletion ball* of radius t centered at $\boldsymbol{x} \in \mathbb{Z}_2^n$ is defined to be

$$\mathcal{D}_t(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{Z}_2^{n-t} \mid \boldsymbol{y} \text{ is a subsequence of } \boldsymbol{x}\}.$$

For two vectors $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{Z}_2^n$, we say that their *Levenshtein distance* is t, and denote $d_L(\boldsymbol{x}_1,\boldsymbol{x}_2) = t$ if $\mathcal{D}_t(\boldsymbol{x}_1) \cap \mathcal{D}_t(\boldsymbol{x}_2) \neq \emptyset$ and $\mathcal{D}_{t-1}(\boldsymbol{x}_1) \cap \mathcal{D}_{t-1}(\boldsymbol{x}_2) = \emptyset$. Equivalently, the Levenshtein distance is one half the minimum number of insertions and deletions required to convert $\boldsymbol{x}$ to $\boldsymbol{y}$. The *minimum Levenshtein distance* of a code $\mathcal{C} \subseteq \mathbb{Z}_2^n$ is the minimum Levenshtein distance between any two different codewords in $\mathcal{C}$. The code is called a *t-deletion-correcting code* if its minimum Levenshtein distance is at least $t+1$. For notational convenience, we denote $\mathcal{D}_t(\boldsymbol{x}) = \emptyset$ when $t < 0$ and thus $|\mathcal{D}_t(\boldsymbol{x})| = 0$. To illustrate these concepts, we include the following example.

Example 1. *Let* $\boldsymbol{x} = (0,1,1,0), \boldsymbol{y} = (1,0,0,0) \in \mathbb{Z}_2^4$ *and* $\mathcal{C} = \{\boldsymbol{x},\boldsymbol{y}\}$. *Notice that* $\mathcal{D}_1(\boldsymbol{x}) = \{(0,1,1),(0,1,0),(1,1,0)\}$,

978-1-5090-1806-2/16/$31.00 ©2016 IEEE

$\mathcal{D}_1(\boldsymbol{y}) = \{(1,0,0),(0,0,0)\}$ *so that* $\mathcal{D}_1(\boldsymbol{x}) \cap \mathcal{D}_1(\boldsymbol{y}) = \emptyset$. *Furthermore, since* $\mathcal{D}_2(\boldsymbol{x}) = \{(0,1),(0,0),(1,1),(1,0)\}$ *and* $\mathcal{D}_2(\boldsymbol{y}) = \{(1,0),(0,0)\}$, $\mathcal{D}_2(\boldsymbol{x}) \cap \mathcal{D}_2(\boldsymbol{y}) = \mathcal{D}_2(\boldsymbol{y})$ *and so* $\mathcal{C}$ *is a single-deletion-correcting code.*

Let $\boldsymbol{a}_n \in \mathbb{Z}_2^n$ be the alternating sequence where its first bit is 1. For $0 < t < n$, we denote by $D(n,t)$ the maximum size of a deletion ball of radius t, i.e.,

$$D(n,t) = \max_{\boldsymbol{x}\in\mathbb{Z}_2^n}\{|\mathcal{D}_t(\boldsymbol{x})|\}.$$

It is known from [2] that

$$D(n,t) = |\mathcal{D}_t(\boldsymbol{a}_n)| = \sum_{i=0}^{t}\binom{n-t}{i}, \tag{2}$$

and from [2] the following recursion holds

$$D(n,t) = D(n-1,t) + D(n-2,t-1). \tag{3}$$

We assume here and afterwards that $D(n,t) = 0$ if $t > n$, $n < 0$, or $t < 0$. The main goal of this work is to study the combinatorial value

$$N(n,t_1,t_2) = \max_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{Z}_2^n, d_L(\boldsymbol{x},\boldsymbol{y})\geq t_1}\{|\mathcal{D}_{t_2}(\boldsymbol{x}) \cap \mathcal{D}_{t_2}(\boldsymbol{y})|\}, \tag{4}$$

where $0 < t_1 < t_2 < n$. In [12], Levenshtein studied the case $t_1 = 1$ (i.e. $\boldsymbol{x}$ and $\boldsymbol{y}$ only need to be different from each other) and showed that for $1 \leq t_2 \leq n-1$

$$N(n,1,t) = 2D(n-2,t-1). \tag{5}$$

This value is achieved for example as the intersection of the balls centered at the two sequences

$$\boldsymbol{x} = (0,1,\boldsymbol{a}_{n-2}), \boldsymbol{y} = (1,0,\boldsymbol{a}_{n-2}).$$

In this paper we will focus on the combinatorial problem stated in (4) for $t_1 = 2$. Our main result in the paper is showing that for $t_1 = 2$ and $n \geq 7$,

$$N(n,2,t) = N(n,t),$$

where $N(n,t)$ is given by

$$\begin{aligned} N(n,t) =& 2D(n-4,t-2)+2D(n-5,t-2)+2D(n-7,t-2) \\ &+D(n-6,t-3)+D(n-7,t-3), \end{aligned} \tag{6}$$

when $t \leq n/2$ and $N(n,t) = 2^{n-t}$ when $t > n/2$. In fact, it is not hard to verify that if $t > n/2$, and $n \geq 8$, then $N(n,t) = 2^{n-t}$ since one can consider the intersection of the sequences $1001\boldsymbol{a}_{n-4}$ and $0110\boldsymbol{a}_{n-4}$, which are of Levenshtein distance two from each other. Thus, unless stated otherwise, we assume in this paper that $t \leq \frac{n}{2}$ and n is not too small (for example $n \geq 10$). In the rest of this section we present some tools and observations that will help us in our solutions.

Let $\mathcal{X} \subseteq \mathbb{Z}_2^n$ be a set and $\boldsymbol{v}$ a vector of length at most n. We denote by $\mathcal{X}^{\boldsymbol{v}}$ the set of all vectors in $\mathcal{X}$ that start with the vector $\boldsymbol{v}$, that is,

$$\mathcal{X}^{\boldsymbol{v}} = \{\boldsymbol{x} \in \mathcal{X} \mid \boldsymbol{v} \text{ is a prefix of } \boldsymbol{x}\}.$$

Similarly, $\mathcal{X}_{\boldsymbol{v}}$ is the set of all vectors in $\mathcal{X}$ that end with the vector $\boldsymbol{v}$. We can apply these notations simultaneously so for two vectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ each of length at most n, $\mathcal{X}_{\boldsymbol{v}_2}^{\boldsymbol{v}_1}$ is the set of all vectors starting with $\boldsymbol{v}_1$ and ending with $\boldsymbol{v}_2$. For a vector $\boldsymbol{v} \in \mathbb{Z}_2^m$ and a set $\mathcal{X} \subseteq \mathbb{Z}_2^n$, the set $\boldsymbol{v} \circ \mathcal{X}$ is the result of prepending the vector $\boldsymbol{v}$ after every vector in $\mathcal{X}$,

$$\boldsymbol{v} \circ \mathcal{X} = \{(v_1,\ldots,v_m,x_1,\ldots,x_n) \mid (x_1,\ldots,x_n) \in \mathcal{X}\}.$$

Similarly, the set $\mathcal{X} \circ \boldsymbol{v}$ consists of the concatenation of the vector $\boldsymbol{v}$ at the end of every vector in $\mathcal{X}$.

We finish this section with the following two lemmas which follow from the same ideas as in [3] and [15]. These lemmas will be used repeatedly in the next two sections. The first one claims that finding the cardinality of a set can be done by splitting it into mutually disjoint sets according to the prefixes and suffixes of its sequences.

Lemma 1. *Let* n, m_1, m_2, *be positive integers such that* $m_1 + m_2 \leq n$, *and* $\mathcal{X} \subseteq \mathbb{Z}_2^n$. *Then,*

$$|\mathcal{X}| = \sum_{\boldsymbol{x}_1\in\mathbb{Z}_2^{m_1}}\sum_{\boldsymbol{x}_2\in\mathbb{Z}_2^{m_2}}\left|\mathcal{X}_{\boldsymbol{x}_2}^{\boldsymbol{x}_1}\right|.$$

As an immediate result of Lemma 1, we conclude that for for every $\boldsymbol{x} \in \mathbb{Z}_2^n$ and two positive integers m_1, m_2 such that $m_1 + m_2 \leq n$, the following equality holds

$$|\mathcal{D}_t(\boldsymbol{x})| = \sum_{\boldsymbol{x}_1\in\mathbb{Z}_2^{m_1}}\sum_{\boldsymbol{x}_2\in\mathbb{Z}_2^{m_2}}\left|\mathcal{D}_t(\boldsymbol{x})_{\boldsymbol{x}_2}^{\boldsymbol{x}_1}\right|. \tag{7}$$

The second lemma claims that finding all sequences in a deletion ball $\mathcal{D}_t(\boldsymbol{x})$ which start with $\boldsymbol{x}_1$ and end $\boldsymbol{x}_2$ can be done by first finding the smallest prefix, suffix that contains $\boldsymbol{x}_1, \boldsymbol{x}_2$ as a subsequence, respectively, and then calculating the deletion ball in the remainder of the sequence $\boldsymbol{x}$.

Lemma 2. *Let* n, m_1, m_2, t *be positive integers, and* $\boldsymbol{x} \in \mathbb{Z}_2^n, \boldsymbol{x}_1 \in \mathbb{Z}_2^{m_1}, \boldsymbol{x}_2 \in \mathbb{Z}_2^{m_2}$. *Assume that* k_1 *is the smallest integer such that* $\boldsymbol{x}_1$ *is a subsequence of* $(x_1,\ldots,x_{k_1})$ *and* k_2 *is the largest integer where* $\boldsymbol{x}_2$ *is a subsequence of* $(x_{k_2},\ldots,x_n)$. *If* $k_1 < k_2$ *then*

$$\mathcal{D}_t(\boldsymbol{x})_{\boldsymbol{x}_2}^{\boldsymbol{x}_1} = \boldsymbol{x}_1 \circ \mathcal{D}_{t^*}(x_{k_1+1},\ldots,x_{k_2-1}) \circ \boldsymbol{x}_2,$$

where $t^* = t - (k_1 - m_1) - (n - k_2 + 1 - m_2)$. *In particular,*

$$\left|\mathcal{D}_t(\boldsymbol{x})_{\boldsymbol{x}_2}^{\boldsymbol{x}_1}\right| = |\mathcal{D}_{t^*}(x_{k_1+1},\ldots,x_{k_2-1})|.$$

Lastly, we state the following claim, whose proof is straightforward. For a binary vector $\boldsymbol{x} = (x_1,x_2,\ldots,x_{n-1},x_n)$, let $R(\boldsymbol{x}) = (x_n,x_{n-1},\ldots,x_2,x_1)$, and $\bar{\boldsymbol{x}} = (1-x_1,\ldots,1-x_n)$.

Claim 1. *Let* $\boldsymbol{x},\boldsymbol{y} \in \mathbb{Z}_2^n$. *Then,*
$|\mathcal{D}_t(\boldsymbol{x})\cap\mathcal{D}_t(\boldsymbol{y})| = |\mathcal{D}_t(R(\boldsymbol{x}))\cap\mathcal{D}_t(R(\boldsymbol{y}))| = |\mathcal{D}_t(\bar{\boldsymbol{x}})\cap\mathcal{D}_t(\bar{\boldsymbol{y}})|$.

The following example illustrates our notations.

Example 2. *Let* $\mathcal{X} \subseteq \mathbb{Z}_2^4$ *be the following set*

$$\begin{aligned}\mathcal{X} = \{&(0,0,0,1),(0,0,1,1),(0,1,0,0),(0,1,0,1),\\ &(0,1,1,0),(1,0,0,0),(1,0,1,1),(1,1,0,1)\}.\end{aligned}$$

Then,

$$\begin{aligned}\mathcal{X}_0^0 &= \{(0,1,0,0),(0,1,1,0)\},\\ \mathcal{X}_1^0 &= \{(0,0,0,1),(0,0,1,1),(0,1,0,1)\},\\ \mathcal{X}_0^1 &= \{(1,0,0,0)\},\\ \mathcal{X}_1^1 &= \{(1,0,1,1),(1,1,0,1)\},\end{aligned}$$

and note that $|\mathcal{X}| = |\mathcal{X}_0^0| + |\mathcal{X}_1^0| + |\mathcal{X}_0^1| + |\mathcal{X}_1^1|$. *Furthermore*

$$\begin{aligned}(0,1)\circ\mathcal{X} = \{&(0,1,0,0,0,1),(0,1,0,0,1,1),(0,1,0,1,0,0),\\ &(0,1,0,1,0,1),(0,1,0,1,1,0),(0,1,1,0,0,0),\\ &(0,1,1,0,1,1),(0,1,1,1,0,1)\}.\end{aligned}$$

If $\boldsymbol{x} = (0,1,1,0,0,1)$ *then*

$$\mathcal{D}_3(\boldsymbol{x})_0^1 = 1 \circ \mathcal{D}_1((1,0)) \circ 0 = 1 \circ \{0,1\} \circ 0,$$

and $|\mathcal{D}_3(\boldsymbol{x})_0^1| = 2$.

For $\boldsymbol{x} = (0,1,1,0), \boldsymbol{y} = (0,1,0,1)$, *we have that*

$$\begin{aligned}&R(\boldsymbol{x}) = (0,1,1,0),\ R(\boldsymbol{y}) = (1,0,1,0),\\ &\bar{\boldsymbol{x}} = (1,0,0,1),\ \bar{\boldsymbol{y}} = (1,0,1,0).\end{aligned}$$

and

$$\mathcal{D}_1(\boldsymbol{x}) \cap \mathcal{D}_1(\boldsymbol{y}) = \{(0,1,0),(0,1,1)\}$$
$$\mathcal{D}_1(R(\boldsymbol{x})) \cap \mathcal{D}_1(R(\boldsymbol{y})) = \{(0,1,0),(1,1,0)\}$$
$$\mathcal{D}_1(\bar{\boldsymbol{x}}) \cap \mathcal{D}_1(\bar{\boldsymbol{y}}) = \{(1,0,1),(1,0,0)\}.$$

In particular,

$$|\mathcal{D}_1(\boldsymbol{x}) \cap \mathcal{D}_1(\boldsymbol{y})| = |\mathcal{D}_1(R(\boldsymbol{x})) \cap \mathcal{D}_1(R(\boldsymbol{y}))| = |\mathcal{D}_1(\bar{\boldsymbol{x}}) \cap \mathcal{D}_1(\bar{\boldsymbol{y}})|.$$

III. The Lower Bound

The main goal of this section is to establish the value of $N(n,t)$ as a lower bound on $N(n,2,t)$. We accomplish this task by showing in Theorem 1 that the sequences $\boldsymbol{x} = (1,0,\boldsymbol{a}_{n-4},0,1)$, $\boldsymbol{y} = (0,1,\boldsymbol{a}_{n-4},1,0)$ of Levenshtein distance two satisfy

$$|\mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})| = N(n,t).$$

We first begin with the following simple, yet useful, claim which is a restatement of a result from [15].

Claim 2. *(c.f. [15], Lemma II.5) Let $\ell < n$, $\boldsymbol{x} \in \mathbb{Z}_2^n, \boldsymbol{y} \in \mathbb{Z}_2^{n-\ell}$, where $\boldsymbol{y} \in \mathcal{D}_\ell(\boldsymbol{x})$ and $\ell < t$. Then, $\mathcal{D}_{t-\ell}(\boldsymbol{y}) \subseteq \mathcal{D}_t(\boldsymbol{x})$.*

The following three claims are the results we need in order to prove Theorem 1.

Claim 3. *For n and t two positive integers such that $t < n$,*

$$|\mathcal{D}_t(\boldsymbol{a}_n) \cap \mathcal{D}_t(\bar{\boldsymbol{a}}_n)| = 2D(n-2,t-1). \tag{8}$$

Proof. Let us denote $\mathcal{X} = \mathcal{D}_t(\boldsymbol{a}_n) \cap \mathcal{D}_t(\bar{\boldsymbol{a}}_n)$. The string $(1,\bar{\boldsymbol{a}}_{n-2})$ is a subsequence of both $\boldsymbol{a}_n$ and $\bar{\boldsymbol{a}}_n$ that can be obtained by deleting a single bit from either $\boldsymbol{a}_n$ or $\bar{\boldsymbol{a}}_n$. Hence, from Claim 2, we have

$$\mathcal{D}_{t-1}((1,\bar{\boldsymbol{a}}_{n-2})) \subseteq \mathcal{D}_t(\boldsymbol{a}_n) \cap \mathcal{D}_t(\bar{\boldsymbol{a}}_n) = \mathcal{X},$$

and in particular $\mathcal{D}_{t-1}((1,\bar{\boldsymbol{a}}_{n-2}))^1 \subseteq \mathcal{X}^1$. Hence, we can write

$$|\mathcal{X}^1| \geq \left|\mathcal{D}_{t-1}((1,\bar{\boldsymbol{a}}_{n-2}))^1\right| \overset{(a)}{=} |1 \circ \mathcal{D}_{t-1}(\bar{\boldsymbol{a}}_{n-2})| \overset{(b)}{=} D(n-2,t-1),$$

where (a) results from Lemma 2 and (b) follows from (2).

Similarly, $(0,\boldsymbol{a}_{n-2})$ is a subsequence of both $\boldsymbol{a}_n$ and $\bar{\boldsymbol{a}}_n$ and thus we have that

$$|\mathcal{X}^0| \geq \left|\mathcal{D}_{t-1}((0,\boldsymbol{a}_{n-2}))^0\right| = |0 \circ \mathcal{D}_{t-1}(\boldsymbol{a}_{n-2})| = D(n-2,t-1).$$

Therefore, according to Lemma 1, we get that

$$|\mathcal{X}| = |\mathcal{X}^0| + |\mathcal{X}^1| \geq 2D(n-2,t-1).$$

However, from (5), for any $\boldsymbol{x},\boldsymbol{y} \in \mathbb{Z}_2^n$,

$$|\mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})| \leq 2D(n-2,t-1),$$

and hence the result in the claim holds. □

Claim 4. *For $n > 0$, $t < n$,*

$$|\mathcal{D}_t((\boldsymbol{a}_{n-1},0)) \cap \mathcal{D}_t((\bar{\boldsymbol{a}}_{n-1},0))| = 2D(n-3,t-1) + D(n-3,t-2).$$

Claim 5. *Let $n > 0$, $t < n$, where $n \geq 10$ is even, then*

$$|\mathcal{D}_t(\boldsymbol{a}_n,0) \cap \mathcal{D}_t(1,\boldsymbol{a}_n)| = 2D(n-3,t-1) + D(n-3,t-2) + D(n-2,t-2).$$

Proof. Let $\mathcal{X} = \mathcal{D}_t(\boldsymbol{a}_n,0) \cap \mathcal{D}_t(1,\boldsymbol{a}_n)$, Then,

$$|\mathcal{X}^1| \overset{(a)}{=} |1 \circ (\mathcal{D}_t(\bar{\boldsymbol{a}}_{n-1},0) \cap \mathcal{D}_t(\boldsymbol{a}_{n-1},0))| \overset{(b)}{=} 2D(n-3,t-1) + D(n-3,t-2).$$

where (a) follows from Lemma 2 and (b) follows from Claim 4. Similarly according to Lemma 2 we get

$$|\mathcal{X}^0| = |0 \circ (\mathcal{D}_{t-1}((\boldsymbol{a}_{n-2},0)) \cap \mathcal{D}_{t-2}(\boldsymbol{a}_{n-2}))| = |\mathcal{D}_{t-1}((\boldsymbol{a}_{n-2},0)) \cap \mathcal{D}_{t-2}(\boldsymbol{a}_{n-2})|.$$

Since $\boldsymbol{a}_{n-2} \in \mathcal{D}_1(\boldsymbol{a}_{n-2},0)$, we can apply Claim 2 to conclude that $|\mathcal{X}^0| = |\mathcal{D}_{t-2}(\boldsymbol{a}_{n-2})| = D(n-2,t-2)$. Lastly, from Lemma 1 it follows that

$$|\mathcal{X}| = |\mathcal{X}^0| + |\mathcal{X}^1| \leq 2D(n-3,t-1) + D(n-3,t-2) + D(n-2,t-2).$$

□

The next theorem constitutes the lower bound.

Theorem 1. *For all $n \geq 10$ and t, let $\boldsymbol{x} = (1,0,\boldsymbol{a}_{n-4},0,1)$, $\boldsymbol{y} = (0,1,\boldsymbol{a}_{n-4},1,0)$. Then, $|\mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})| = N(n,t)$.*

Proof. We show the proof only for n even. Let us denote $\mathcal{X} = \mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})$. We first consider the set $\mathcal{X}_0^0$.

$$|\mathcal{X}_0^0| \overset{(a)}{=} |0 \circ (\mathcal{D}_{t-2}(\boldsymbol{a}_{n-4}) \cap \mathcal{D}_t((1,\boldsymbol{a}_{n-4},1))) \circ 0| = |\mathcal{D}_{t-2}(\boldsymbol{a}_{n-4}) \cap \mathcal{D}_t((1,\boldsymbol{a}_{n-4},1))| \overset{(b)}{=} |\mathcal{D}_{t-2}(\boldsymbol{a}_{n-4})| = D(n-4,t-2), \tag{9}$$

where (a) follows from Lemma 2 and (b) follows by applying Claim 2 since $\boldsymbol{a}_{n-4} \in \mathcal{D}_2((1,\boldsymbol{a}_{n-4},1))$. By repeating the same steps of $\mathcal{X}_0^0$, we get that

$$|\mathcal{X}_1^1| = D(n-4,t-2). \tag{10}$$

For the set $\mathcal{X}_0^1$, we have from Lemma 2

$$|\mathcal{X}_0^1| = |1 \circ (\mathcal{D}_{t-1}(\bar{\boldsymbol{a}}_{n-3}) \cap \mathcal{D}_{t-1}(\boldsymbol{a}_{n-3})) \circ 0| = |\mathcal{D}_{t-1}(\bar{\boldsymbol{a}}_{n-3}) \cap \mathcal{D}_{t-1}(\boldsymbol{a}_{n-3})| \overset{(a)}{=} 2D(n-5,t-2), \tag{11}$$

where (a) follows from Claim 3. Similarly, for the set $\mathcal{X}_1^0$, we have

$$|\mathcal{X}_1^0| = |0 \circ (\mathcal{D}_{t-1}(\boldsymbol{a}_{n-4},0) \cap \mathcal{D}_{t-1}(1,\boldsymbol{a}_{n-4})) \circ 1| = |\mathcal{D}_{t-1}(\boldsymbol{a}_{n-4},0) \cap \mathcal{D}_{t-1}(1,\boldsymbol{a}_{n-4})| \overset{(a)}{=} 2D(n-7,t-2) + D(n-6,t-3) + D(n-7,t-3), \tag{12}$$

where (a) is a result of Claim 5. Finally, by applying Lemma 1 and summing (9), (10), (11), and (12) we get the result in the statement of the theorem. □

The following corollary summarizes this result.

Corollary 1. *For n and t where $n \geq 10$, we have*

$$N(n,2,t) \geq N(n,t).$$

IV. The Upper Bound

We now prove that the lower bound from the previous section is also an upper bound. That is, we show that for any pair of sequences, $\boldsymbol{x},\boldsymbol{y}$ where $d_L(\boldsymbol{x},\boldsymbol{y}) \geq 2$, the intersection of their deletion balls is at most $N(n,t)$. We proceed by considering different possibilities for $\boldsymbol{x},\boldsymbol{y}$ and showing that this result holds in all cases. To improve the flow of the section and due to the lack of space these proofs are omitted and will appear in the long version of the paper.

At a high level, the logical flow of the section is the following. We first assume $d_L(\boldsymbol{x},\boldsymbol{y}) \geq 2$ and that $\boldsymbol{x},\boldsymbol{y}$ differ on the first and last bits. Lemma 3 shows that if $\boldsymbol{x},\boldsymbol{y}$ have more than $N(n,t)$ common subsequences, then $\boldsymbol{x}$

cannot have the same value in positions one and two. As a consequence of Lemma 3, if $\boldsymbol{x}, \boldsymbol{y}$ have more than $N(n,t)$ common subsequences then the sequences can be written as $\boldsymbol{x} = (1, 0, \boldsymbol{x}', \bar{a}, a) \in \mathbb{Z}_2^n, \boldsymbol{y} = (0, 1, \boldsymbol{y}', a, \bar{a}) \in \mathbb{Z}_2^n$ for some $a \in \mathbb{Z}_2$. Lemma 5 then shows that the intersection of the deletion balls for $\boldsymbol{x}, \boldsymbol{y}$ (as described in the previous sentence) is at most $N(n,t)$. Finally, in Theorem 2, we remove the restriction that $\boldsymbol{x}$ and $\boldsymbol{y}$ differ in the first bit, and show that when $d_L(\boldsymbol{x}, \boldsymbol{y}) \geq 2$ for any $\boldsymbol{x}, \boldsymbol{y}$ the intersection of their deletion balls is at most $N(n,t)$.

We introduce a few identities that will be used throughout the section. From [12], we have

$$D(n,t) \leq D(n+1, t+1), \tag{13}$$

and

$$D(n,t) \leq D(n+1, t). \tag{14}$$

The following corollary follows from (3) and (6).

Corollary 2. *For n and t where $n \geq 10$, we have*

$$N(n,t) = N(n-1,t) + N(n-2,t-1). \tag{15}$$

We begin by restricting our attention to sequences $\boldsymbol{x}, \boldsymbol{y}$ that differ on the first bit and the last bit.

Lemma 3. *Let n,t be integers where $n \geq 7$. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}_2^n$ be such that $x_1 = 1$, $y_1 = 0$ and*
1) *$d_L(\boldsymbol{x}, \boldsymbol{y}) \geq 2$,*
2) *$x_n \neq y_n$, and*
3) *$x_1 = x_2$.*

Then, $|\mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})| \leq N(n,t)$.

Proof. Assume first that $x_n = 0$, so $y_n = 1$ and we can write

$$\boldsymbol{x} = (1, 1, \boldsymbol{x}', 0), \quad \boldsymbol{y} = (0, \boldsymbol{y}', 1),$$

where $\boldsymbol{x}' = (x'_1, \ldots, x'_{n-3}) \in \mathbb{Z}_2^{n-3}, \boldsymbol{y}' = (y'_1, \ldots, y'_{n-2}) \in \mathbb{Z}_2^{n-2}$. Let $\mathcal{X} = \mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})$. We first consider the set $|\mathcal{X}^0|$, where we have

$$|\mathcal{X}^0| \leq |\mathcal{D}_t(\boldsymbol{x})^0| \leq |\mathcal{D}_{t-2}(\boldsymbol{x}', 0)^0| \leq D(n-3, t-2).$$

Similarly,

$$|\mathcal{X}_0^1| \leq |\mathcal{D}_t(\boldsymbol{y})_0^1| \leq |\mathcal{D}_{t-2}(\boldsymbol{y}')_0^1| \leq D(n-4, t-2).$$

As for the set $|\mathcal{X}_1^1|$, assume first that $x'_{n-3} = 1$ and $y'_1 = 1$. Then we get

$$\begin{aligned}|\mathcal{X}_1^1| &= \left|1 \circ \left(\mathcal{D}_{t-1}((1, x'_1, \ldots, x'_{n-4})) \cap \mathcal{D}_{t-1}((y'_2, \ldots, y'_{n-2}))\right) \circ 1\right| \\ &= \left|\left(\mathcal{D}_{t-1}((1, x'_1, \ldots, x'_{n-4})) \cap \mathcal{D}_{t-1}((y'_2, \ldots, y'_{n-2}))\right)\right| \\ &\leq 2D(n-5, t-2),\end{aligned}$$

where the last inequality holds since $(1, x'_1, \ldots, x'_{n-4}) \neq (y'_2, \ldots, y'_{n-2})$ as otherwise we won't have $d_L(\boldsymbol{x}, \boldsymbol{y}) \geq 2$. To complete this part assume that $x'_{n-3} = 0$, then we get

$$\begin{aligned}|\mathcal{X}_1^1| &\leq |\mathcal{D}_{t-1}((1, \boldsymbol{x}'))_1| \leq \left|\mathcal{D}_{t-2}((1, x'_1, \ldots, x'_{n-4}))_1\right| \\ &\leq D(n-4, t-2) \leq 2D(n-5, t-1).\end{aligned}$$

The same proof holds for the case $y'_1 = 0$.

We now consider the case $x_n = 1$, so $y_n = 0$ and

$$\begin{aligned}\boldsymbol{x} &= (1, 1, \boldsymbol{x}', 1), \\ \boldsymbol{y} &= (0, \boldsymbol{y}', 0),\end{aligned}$$

and let $\mathcal{X} = \mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})$. As before, we get

$$|\mathcal{X}^0| \leq D(n-3, t-2), |\mathcal{X}_1^1| \leq D(n-4, t-2).$$

and $\mathcal{X}_0^1| \leq 2D(n-5, t-2)$.

Finally, we conclude that in both cases

$$|\mathcal{X}| \leq D(n-3, t-2) + D(n-4, t-2) + 2D(n-5, t-2) \leq N(n,t),$$

where the last inequality holds since

$$\begin{aligned}&N(n,t) - (D(n-3,t-2) + D(n-4,t-2) + 2D(n-5,t-2)) \\ =&N(n,t) - (2D(n-4,t-2) + D(n-5,t-3) + 2D(n-5,t-2)) \\ =&2D(n-7,t-2) + D(n-6,t-3) + D(n-7,t-3) \\ &-(D(n-6,t-3) + D(n-7,t-4)) \\ =&2D(n-7,t-2) + D(n-7,t-3) - D(n-7,t-4) \geq 0.\end{aligned}$$

□

As a result of Lemma 3, if $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}_2^n$, $x_1 \neq y_1$, and $|\mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})| > N(n,t)$, then $\boldsymbol{x} = (1, 0, \boldsymbol{x}', 1, 0), \boldsymbol{y} = (0, 1, \boldsymbol{x}', 0, 1)$ or $\boldsymbol{x} = (1, 0, \boldsymbol{x}', 0, 1), \boldsymbol{y} = (0, 1, \boldsymbol{x}', 1, 0)$ where $\boldsymbol{x}', \boldsymbol{y}' \in \mathbb{Z}_2^{n-4}$. The purpose of the next lemma is to show that in either case $\boldsymbol{x}', \boldsymbol{y}' \in \{\boldsymbol{a}_{n-4}, \bar{\boldsymbol{a}}_{n-4}\}$. In order to establish this result we first state the following three useful claims.

Claim 6. *Suppose the number of runs in $\boldsymbol{x} \in \mathbb{Z}_2^n$ is at most $n-1$. Then for $t \leq n/2$,*

$$|\mathcal{D}_t(\boldsymbol{x})| \leq D(n-2,t) + D(n-4,t-2) + D(n-2,t-1).$$

Claim 7. *Suppose n,t are integers where $n \geq 10$ and $t \leq n/2$. Suppose $\boldsymbol{x} \in \mathbb{Z}_2^n, \boldsymbol{y} \in \mathbb{Z}_2^n$ are such that $\boldsymbol{x} \neq \boldsymbol{y}$, the number of runs in $\boldsymbol{x} = (x_1, \ldots, x_n)$ is $n-1$, $\boldsymbol{y}$ has at least $n-1$ runs, and $x_2 \neq x_3, x_{n-1} \neq x_{n-2}$. Then,*

$$\begin{aligned}|\mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})| \leq &D(n-2,t-1) + D(n-4,t-1) \\ &+ D(n-4,t-2) + D(n-6,t-3).\end{aligned}$$

Claim 8. *Suppose n,t are integers where $n \geq 10$ and $t \leq n/2$. Suppose $\boldsymbol{x} \in \mathbb{Z}_2^n, \boldsymbol{y} \in \mathbb{Z}_2^n$ are such that $\boldsymbol{x} \neq \boldsymbol{y}$ and the number of runs in $\boldsymbol{x}$ is at most $n-2$. Then,*

$$\begin{aligned}|\mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})| \leq &D(n-2,t-1) + D(n-4,t-1) \\ &+ D(n-4,t-2) + D(n-6,t-3).\end{aligned}$$

Using Claims 7 and 8, we prove the following lemma.

Lemma 4. *Let n,t be integers where $n \geq 10$ and let $a \in \mathbb{Z}_2$. Let $\boldsymbol{x} = (1, 0, \boldsymbol{x}', \bar{a}, a) \in \mathbb{Z}_2^n, \boldsymbol{y} = (0, 1, \boldsymbol{y}', a, \bar{a}) \in \mathbb{Z}_2^n$ be such that $d_L(\boldsymbol{x}, \boldsymbol{y}) \geq 2$. If*

$$|\mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})| \geq N(n,t),$$

then $\boldsymbol{x}', \boldsymbol{y}' \in \{\boldsymbol{a}_{n-4}, \bar{\boldsymbol{a}}_{n-4}\}$.

Proof. Similar to before, let $\mathcal{X} = \mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})$. Then, since $\boldsymbol{x}' \notin \{\boldsymbol{a}_{n-4}, \bar{\boldsymbol{a}}_{n-4}\}$, we can apply Claim 6 so that

$$\begin{aligned}|\mathcal{X}_{\bar{a}}^0| =&|0 \circ (\mathcal{D}_{t-2}(\boldsymbol{x}') \cap \mathcal{D}_t(1, \boldsymbol{y}', a)) \circ \bar{a}| \\ \leq&|\mathcal{D}_{t-2}(\boldsymbol{x}')| \leq D(n-6,t-2) \\ &+ D(n-6,t-3) + D(n-8,t-4).\end{aligned}$$

We also have

$$|\mathcal{X}_a^1| \leq |1 \circ \mathcal{D}_{t-2}(\boldsymbol{y}') \circ a| \leq D(n-4,t-2).$$

Furthermore, notice

$$|\mathcal{X}_{\bar{a}}^1| = |1 \circ (\mathcal{D}_{t-1}(0, \boldsymbol{x}') \cap \mathcal{D}_{t-1}(\boldsymbol{y}', a)) \circ \bar{a}|,$$

where $(\boldsymbol{y}', a) \neq (0, \boldsymbol{x}')$ so that we can apply (5) to get that

$$|\mathcal{X}_{\bar{a}}^1| \leq 2D(n-5,t-2).$$

Next we consider $|\mathcal{X}_a^0|$ where

$$|\mathcal{X}_a^0| = |0 \circ (\mathcal{D}_{t-1}(\boldsymbol{x}', \bar{a}) \cap \mathcal{D}_{t-1}(1, \boldsymbol{y}')) \circ a|$$

If $(\boldsymbol{x}', \bar{a}), (1, \boldsymbol{y}')$ satisfy the conditions of Claim 7 or Claim 8, then we can write

$$|\mathcal{X}_a^0| \leq D(n-5, t-2) + D(n-7, t-2) + D(n-7, t-3) + D(n-9, t-4).$$

We now consider the case where $(\boldsymbol{x}', \bar{a}), (1, \boldsymbol{y}')$ do not satisfy the conditions in Claim 7 or Claim 8, and show that in this case the (different) vectors $(0, \boldsymbol{x}'), (\boldsymbol{y}', a)$ satisfy the conditions in Claim 7. If $(\boldsymbol{x}', \bar{a}), (1, \boldsymbol{y}')$ do not satisfy the conditions in Claim 8, then $(\boldsymbol{x}', \bar{a}), (1, \boldsymbol{y}')$ each have at least $(n-3)-1 = n-4$ runs and, in particular, $\boldsymbol{x}'$ has $(n-4)-1 = n-5$ runs by assumption since $\boldsymbol{x}' \notin \{\boldsymbol{a}_{n-4}, \bar{\boldsymbol{a}}_{n-4}\}$. If $(\boldsymbol{x}', \bar{a}), (1, \boldsymbol{y}')$ do not satisfy the conditions in Claim 7 then $x'_2 = x'_3$ or $x'_{n-4} = x'_{n-5}$. Since $\boldsymbol{x}'$ has $n-5$ runs, then in either case $x'_1 \neq x'_2$ and $x'_{n-6} \neq x'_{n-5}$. Thus, $(0, \boldsymbol{x}'), (\boldsymbol{y}', a)$ satisfy the conditions in Claim 7. Therefore, using the same logic as before

$$|\mathcal{X}_a^0| \leq 2D(n-5, t-2),$$

and

$$|\mathcal{X}_{\bar{a}}^1| \leq D(n-5, t-2) + D(n-7, t-2) + D(n-7, t-3) + D(n-9, t-4).$$

Thus, applying Lemma 1, it is possible to show that if $\boldsymbol{x}, \boldsymbol{y}$ satisfy the conditions in this lemma, then

$$|\mathcal{X}| \leq D(n-4, t-2) + 3D(n-5, t-2) + D(n-6, t-2) + D(n-7, t-2) + D(n-6, t-3) + D(n-7, t-3) + D(n-8, t-4) + D(n-9, t-4) = N(n, t).$$

□

The next lemma will be directly used in the derivation of the upper bound. The proof relies on simply enumerating the possible choices for $\boldsymbol{x}, \boldsymbol{y}$ and applying previous results such as Lemma 4.

Lemma 5. *Let n, t be integers where $n \geq 7$. Let $\boldsymbol{x} = (1, 0, \boldsymbol{x}', \bar{a}, a) \in \mathbb{Z}_2^n, \boldsymbol{y} = (0, 1, \boldsymbol{y}', a, \bar{a}) \in \mathbb{Z}_2^n$ be such that $d_L(\boldsymbol{x}, \boldsymbol{y}) \geq 2$. Then,*

$$|\mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})| \leq N(n, t).$$

We now have the main result of this paper.

Theorem 2. *Let n, t be integers where $n \geq 7$. Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{Z}_2^n, \boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in \mathbb{Z}_2^n$ be such that $d_L(\boldsymbol{x}, \boldsymbol{y}) \geq 2$. Then,*

$$|\mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})| \leq N(n, t).$$

Proof. The proof will be by induction on the lengths of the sequences $\boldsymbol{x}, \boldsymbol{y}$. The base cases for $n = 10, 11$ were verified using a computerized search.

Suppose the result holds for all $m < n$ where n is the length of $\boldsymbol{x}, \boldsymbol{y}$. Let us write $\boldsymbol{x} = (a, \boldsymbol{x}'), \boldsymbol{y} = (b, \boldsymbol{y}')$ (for $a, b \in \mathbb{Z}_2$) and denote $\mathcal{X} = \mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})$. There are two cases:

1) $x_1 = y_1$, or
2) $x_1 \neq y_1$.

We first consider case 1). Then, $d_L(\boldsymbol{x}', \boldsymbol{y}') \geq 2$ and from the inductive hypothesis we get

$$\begin{aligned} |\mathcal{X}^a| &= |a \circ (\mathcal{D}_t(\boldsymbol{x}') \cap \mathcal{D}_t(\boldsymbol{y}'))| = |\mathcal{D}_t(\boldsymbol{x}') \cap \mathcal{D}_t(\boldsymbol{y}')| \\ &\leq \max_{\boldsymbol{x}^*, \boldsymbol{y}^* \in \mathbb{Z}_2^{n-1}, d_L(\boldsymbol{x}^*, \boldsymbol{y}^*) \geq 2} \{|\mathcal{D}_t(\boldsymbol{x}^*) \cap \mathcal{D}_t(\boldsymbol{y}^*)|\} \\ &\leq N(n-1, t). \end{aligned}$$

Suppose x_k is the first occurrence of the symbol $\bar{a}$ in $\boldsymbol{x}$ and the symbol $\bar{a}$ appears in $\boldsymbol{x}$ not after it appears in $\boldsymbol{y}$. Notice that under this setup, we have $(x_1, x_2, \ldots, x_{k-1}) = (y_1, y_2, \ldots, y_{k-1})$ and so $d_L(\boldsymbol{x}'', \boldsymbol{y}'') \geq 2$, where $\boldsymbol{x}'' = (x_k, \ldots, x_n)$ and $\boldsymbol{y}'' = (y_k, \ldots, y_n)$. Since $k \geq 2$, applying the inductive hypothesis along with Lemma 2 gives

$$\begin{aligned} |\mathcal{X}^{\bar{a}}| &\leq |\bar{a} \circ (\mathcal{D}_{t-(k-1)}(\boldsymbol{x}'') \cap \mathcal{D}_{t-(k-1)}(\boldsymbol{y}''))| \\ &= |\mathcal{D}_{t-(k-1)}(\boldsymbol{x}'') \cap \mathcal{D}_{t-(k-1)}(\boldsymbol{y}'')| \\ &\leq \max_{\boldsymbol{x}^*, \boldsymbol{y}^* \in \mathbb{Z}_2^{n-k}, d_L(\boldsymbol{x}^*, \boldsymbol{y}^*) \geq 2} \{|\mathcal{D}_{t-(k-1)}(\boldsymbol{x}^*) \cap \mathcal{D}_{t-(k-1)}(\boldsymbol{y}^*)|\} \\ &\leq N(n-k, t-k+1) \\ &\leq N(n-2, t-1). \end{aligned}$$

Since $N(n, t) = N(n-1, t) + N(n-2, t-1)$ from Corollary 2 for the case where $x_1 = y_1$.

In order to prove case 2), suppose that $x_1 \neq y_1$. Then, we also assume $x_n \neq y_n$ since otherwise if $x_n = y_n$ we can reverse the sequences and apply the same logic as above. However, if $x_1 \neq y_1$ and $x_n \neq y_n$ then from Lemmas 3 and 5, we have $|\mathcal{D}_t(\boldsymbol{x}) \cap \mathcal{D}_t(\boldsymbol{y})| \leq N(n, t)$ in this case as well and so the result holds. □

Acknowledgment: The work or Ryan Gabrys was supported in part by the NISE program at SSC Pacific. The work of Eitan Yaakobi was supported in part by the Israel Science Foundation (ISF) Grant No. 1624/14.

References

[1] T. Batu, S. Kannan, S. Khanna, and A. McGregor, "Reconstructing strings from random traces," *Proc. of the Fifteenth Annual ACM-SIAM Symp. on Discrete Algorithms*, pp. 903–911, 2004.
[2] L. Calabi, "On the computation of Levenshtein's distances," TN-9-0030, Parke Math. Labs., Inc., Carlisle, MA, 1967.
[3] D. S. Hirschberg and M. Regnier, "Tight bounds on the number of string subsequences," *Journal of Discrete Algorithms*, vol. 1, pp. 123-132, 2000.
[4] B. Haeupler and M. Mitzenmacher, "Repeated deletion channels," *Proc. IEEE Informa. Theory Workshop*, pp. 152–156, Hobart, Tasmania, Australia, Nov. 2014.
[5] T. Holenstein, M. Mitzenmacher, R. Panigrahy, and U. Wieder, "Trace reconstruction with constant deletion probability and related results," *Proc. of the Nineteenth Annual ACM-SIAM Symp. on Discrete Algorithms*, pp. 389–398, 2008.
[6] T. Jiang and A. Vardy, "Asymptotic improvement of the GilbertVarshamov bound on the size of binary codes," *IEEE Trans. on Inform. Theory*, vol. 50, no. 8, pp. 1655–1664, Aug. 2004.
[7] S. Kannan and A. McGregor, "More on reconstructing strings from random traces: insertions and deletions," *Proc. IEEE Int. Sym. Inform. Theory*, pp. 297–301, Australia, Sep. 2005.
[8] E. Konstantinova, "Reconstruction of permutations distorted by single reversal errors," *Discrete Applied Math.*, vol. 155, pp. 2426–2434, 2007.
[9] E. Konstantinova, V. Levenshtein, and J. Siemons, "Reconstruction of permutations distorted by single transposition errors," *http://arxiv.org/abs/math/0702191v1*, Feb. 2007.
[10] E. Konstantinova, "On reconstruction of signed permutations distorted by reversal errors," *Discrete Mathematics*, vol. 308, pp. 974–984, 2008.
[11] V.I. Levenshtein, "Efficient reconstruction of sequences," *IEEE Trans. on Inform. Theory*, vol. 47, no. 1, pp. 2–22, Jan. 2001.
[12] V. I. Levenshtein, "Efficient reconstruction of sequences from their subsequences or supersequences" *Journal of Combin. Theory, Ser. A*, vol. 93, no. 2, pp. 310–332, 2001.
[13] V.I. Levenshtein, E. Konstantinova, E. Konstantinov, and S. Molodtsov, "Reconstruction of a graph from 2-vicinities of its vertices," *Discrete Applied Mathematics*, vol. 156, pp. 1399–1406, 2008.
[14] V.I. Levenshtein and J. Siemons, "Error graphs and the reconstruction of elements in groups," *Journal of Combin. Theory, Ser. A*, vol. 116, pp. 795–815, 2009.
[15] Y. Liron and M. Langberg, "A characterization of the number of subsequences obtained via the deletion channel," available at http://arxiv.org/abs/1202.1644, 2012.
[16] F. Sala, R. Gabrys, C. Schoeny, and L. Dolecek, "Three novel combinatorial theorems for the insertion/deletion channel," *Proc. Int. Symp. Inform. Theory*, Hong Kong, 2015.
[17] K. Viswanathan and R. Swaminathan, "Improved string reconstruction over insertion-deletion channels," *Proc. of the Nineteenth Annual ACM-SIAM Symp. on Discrete Algorithms*, pp. 399–408, 2008.
[18] E. Yaakobi and J. Bruck, "On the Uncertainty of Information Retrieval in Associative Memories," *Proc. Int. Symp. Inform. Theory*, pp. 106–110, Jul. 2012.
[19] E. Yaakobi, M. Schwartz, M. Langberg, J. Bruck, "Sequence reconstruction for Grassmann graphs and permutations," *Proc. Int. Symp. Inform. Theory*, pp. 874–878, Jul. 2013.