Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 12, Number 1 (2016), pp. 243-251

http://www.ripublication.com

On the cardinality of optimal t-deletion-correcting binary codes of length $2t + 3$

Emil Milanov Kolev
Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences, Sofia, Bulgarian.

Abstract

In this paper we consider binary deletion-correcting codes. For any positive integer t we find the value of $M_2(2t+3, t)$, the maximum cardinality of t-deletion-correcting binary code of length $2t + 3$.

Keywords: insertion/deletion codes, Varshamov-Tennengolts codes, multiple insertion/deletion codes AMS Mathematics Subject Classification: 94B05

Introduction

When a binary message is transmitted through a noisy channel some of its symbols may change. Error-correcting codes are designed to correct such errors.
In a case of symbol lost the receiver gets shorter message and he does not know which of the symbols were lost. Deletion-correcting codes are designed to correct such deletions. A code is called t-deletion-correcting if it corrects any t deletions [1], [2].

Example 1 Consider the binary code $C = \{00000, 11111, 00011, 11000, 10101, 01110\}$. For a given codeword we may delete any of its five symbols. As a result we obtain a set of vectors of length 4. Direct verification shows that all six sets obtained from the six codewords are disjoined. Therefore *C* is 1-deletion-correcting code.
The Levenstein distance $d_L(\mathbf{x}, \mathbf{y})$ of two binary vectors is the minimum number of deletions and insertions needed to transform **x** into **y**.
Deletion distance $\mathrm{dd}(\mathbf{u}, \mathbf{v})$ between two vectors $\mathbf{u}$ and $\mathbf{v}$ of equal length is defined as one-half of the smallest number of deletions and insertions needed to change $\mathbf{u}$ to $\mathbf{v}$, [3]. For example, $\mathrm{dd}(00000, 11111) = 5$ whereas $\mathrm{dd}(00011, 10101) = 2$. It is clear that for vectors $\mathbf{u}$ and $\mathbf{v}$ of equal length we have

$$\mathrm{dd}(\mathrm{u}, \mathrm{v}) = \frac{1}{2} d_L(\mathrm{u}, \mathrm{v}).$$

For a given code C the deletion distance $\mathrm{dd}(C)$ is defined as
$\mathrm{dd}(C) = \min\{\mathrm{dd}(\mathrm{u},\mathrm{v}) \mid \mathrm{u},\mathrm{v} \in C\}$.
Denote by $M_2(n,t)$ the maximum cardinality of a binary t-deletion-correcting code C of length n, i. e. for any two distinct codewords $\mathbf{u}$ and $\mathbf{v}$ we have $\mathrm{dd}(u,v) > t$ (or, equivalently $d_L(\mathbf{u},\mathbf{v}) > 2t$). A binary code C of length n and cardinality $M_2(n,t)$ is called optimal. For more information and useful results the reader is referred to [4], [5], [6], [7], [8], [9].
In this paper we consider optimal t-deletion-correcting codes of length $n = 2t+3$. The main result is given by the following theorem.

Theorem 1 It is true that:
(a) $M_2(5,1) = 6$, i. e. an optimal 1-deletion-correcting code of length 5 has cardinality 6;
(b) $M_2(2t+3,t) = 5$ for $t \geq 2$, i. e. an optimal t-deletion-correcting code of length $2t+3$ has cardinality 5.

Definitions and Preliminary Results

In the proof of the main theorem we repeatedly use the following Lemma.

Lemma 1 Consider two binary vectors
$\mathbf{x} = (x_1, x_2, \dots, x_{2t+3})$ and $\mathbf{y} = (y_1, y_2, \dots, y_{2t+3})$.
Suppose there exist $i,j \in \{1,2,\dots,2t+2\}$ such that $x_i = y_j$ and $x_{i+1} = y_{j+1}$. Let for some $p,q \in \{0,1\}$
$l = \min\{\,|\{m: m < i, x_m = p\}|, |\{n: n < j, y_n = p\}|\,\}$
$r = \min\{\,|\{m: m > i+1, x_m = q\}|, |\{n: n > j+1, y_n = q\}|\,\}$.
Then $\mathrm{dd}(\mathbf{x},\mathbf{y}) \leq 2t+1-l-r$.

Proof For both vectors $\mathbf{x}$ and $\mathbf{y}$ delete all elements distinct from p on the left of x_i and y_j and all elements distinct from q on the right of x_{i+1} and y_{j+1}. Hence we obtain from both $\mathbf{x}$ and $\mathbf{y}$ the following vector of length $l+r+2$:
$p^l x_i x_{i+1} q^r$.
The number of deleted symbols equals $2t+3-(l+2+r) = 2t+1-l-r$. Thus, $\mathrm{dd}(x,y) \leq 2t+1-l-r$. $\diamond$
The following notation facilitates the application of Lemma 1. For $p,q \in \{0,1\}$ and two positive integers l and r denote by $\mathbf{u}_{r,q}^{l,p}$ the vector obtained from $\mathbf{u}$ by deleting its first l symbols p and deleting its last r symbols q.
Consider a vector $\mathbf{u} = (u_1, u_2, \dots, u_{2t+3})$ of length $2t+3$ and two positive integers a and b such that $a+b = t+1$. The vectors $\mathbf{u_a} = (u_1, u_2, \dots, u_{2a})$ and $\mathbf{u_b} = (u_{2a+2}, u_{2a+3}, \dots, u_{2t+3})$ are called a and b components of $\mathbf{u}$, respectively. In other words $\mathbf{u_a}$ is the vector of the first $2a$ entries of $\mathbf{u}$ and $\mathbf{u_b}$ is the vector of the last $2b$ entries of $\mathbf{u}$. Note that the entry u_{2a+1} of the vector $\mathbf{u}$ is not included in neither $\mathbf{u_a}$or $\mathbf{u_b}$.

Consider a binary vector $\mathbf{v}$ of length $2m$. When $m = \mathrm{wt}(\mathbf{v})$ (for a vector $\mathbf{x}$ by $\mathrm{wt}(\mathbf{x})$ we denote the usual Hamming weight of $\mathbf{x}$) then $\mathbf{v}$ is called balanced. If $|m - \mathrm{wt}(\mathbf{v})| = k$ for some positive integer k then $\mathbf{v}$ is called k-unbalanced.
For fixed positive integers a and b with $a + b = t + 1$ we say that a vector $\mathbf{u}$ of length $2t + 3$ is:

- balanced, if both vectors $\mathbf{u_a}$ and $\mathbf{u_b}$ are balanced;
- k-unbalanced, if at least one of the components $\mathbf{u_a}$ and $\mathbf{u_b}$ is k-unbalanced.

Existence Results

In this section we show existence results for t-deletion-correcting codes of length $2t + 3$.

Proposition 1 **(a).** $M_2(5,1) \geq 6$, i. e. there exists 1-deletion-correcting code of length 5 and cardinality 6.
(b). $M_2(2t+3, t) \geq 5$ for any positive integer $t \geq 2$ i. e. there exists t-deletion-correcting code of length $2t + 3$ and cardinality 5.

Proof. **(a).** For the code $C_1 = \{00000, 11111, 00011, 11000, 10101, 01110\}$ from Example 1 we have $\mathrm{dd}(C_1) = 2$.
(b). Consider the code
$C_2 = \{\mathbf{u_1} = 0^{2t+3},\ \mathbf{u_2} = 1^{2t+3}, \mathbf{u_3} = 0^{t+1}1^{t+2}, \mathbf{u_4} = 1^{t+2}0^{t+1}, \mathbf{u_5} = (01)^{t+1}0\}$.

By t deletions from $\mathbf{u_1}$, $\mathbf{u_2}$, $\mathbf{u_3}$ and $\mathbf{u_4}$ one obtains vectors of the form 0^{t+3}, 1^{t+3}, 0^s1^{t+3-s}, $s \geq 1$, and 1^k0^{t+3-k}, $k \geq 2$, respectively. It is clear now that $\mathrm{dd}(\mathbf{u_i}, \mathbf{u_j}) \geq t + 1$ for $i, j \in \{1,2,3,4\}, i \neq j$.

It remains to prove that $\mathrm{dd}(\mathbf{u_i}, \mathbf{u_5}) \geq t + 1$ for $i = 1,2,3,4$. Note that there are $2t + 2$ pairs (p, q), $p \neq q$ of consecutive terms in $\mathbf{u_5}$. A single deletion decreases this number by at most two. Therefore after t deletions in $\mathbf{u_5}$ there exist at least two such pairs of consecutive terms. None of the vectors obtained from $\mathbf{u_i}$ for $i = 1,2,3,4$ by t deletions has more that one such pair. Hence $\mathrm{dd}(\mathbf{u_i}, \mathbf{u_5}) \geq t + 1$ for $i = 1,2,3,4$ and therefore C_2 is t-deletion-correcting code of length $2t + 3$ and cardinality 5.

Proof of Theorem 1
Suppose there exists t-deletion-correcting code C of length $2t + 3$ and cardinality 6. Since there exists at most one codeword $\mathbf{u} \in C$ of weight $\mathrm{wt}(\mathbf{u}) \geq t + 3$ (respectively $\mathrm{wt}(\mathbf{u}) \leq t$) we may assume that all 1's vector (respectively all 0's vector) is a codeword. It follows now that any of the remaining four codewords is of weight $t + 1$ or $t + 2$.
Let a and b be fixed positive integers such that $a + b = t + 1$. Note that after t dele- tions a vector $\mathbf{x}$ of length $2t + 3$ is transformed into a vector of length $t + 3 = a + b + 2$.

Proposition 2 There exist at most two unbalanced codewords.

Proof. We prove first that if there exists k-unbalanced codeword $\mathbf{u}$ then by $t-2k+2$ deletions in $\mathbf{u}$ one can obtain one of the two vectors:
$0^{a+k}1^{b+k}$ or $1^{a+k}0^{b+k}$
Assume for a codeword $\mathbf{u}$ the vector $\mathbf{u_a}$ is k-unbalanced (the case $\mathbf{u_b}$ k-unbalanced is settled in the same way). This implies that $\mathrm{wt}(\mathbf{u_a}) = a-k$ or $\mathrm{wt}(\mathbf{u_a}) = a+k$ and the two cases are equivalent modulo interchanging 0 and 1 in all codewords.
Let $\mathrm{wt}(\mathbf{u_a}) = a-k$. Since $\mathrm{wt}(\mathbf{u}) = t+1$ or $t+2$ we have
$\mathrm{wt}(\mathbf{u_b}) \in \{b+k-1, b+k, b+k+1\}$.
Also, if $\mathrm{wt}(\mathbf{u_b}) = b+k-1$ then $u_{2a+1} = 1$; if $\mathrm{wt}(\mathbf{u_b}) = b+k$ then $u_{2a+1} = 0$ or 1 and if $\mathrm{wt}(\mathbf{u_b}) = b+k+1$ then $u_{2a+1} = 0$. Therefore

- if $\mathrm{wt}(\mathbf{u_b}) = b+k-1$ then $\mathbf{u}_{b-k+1,0}^{a-k,1} = 0^{a+k}u_{2a+1}1^{b+k-1} = 0^{a+k}1^{b+k}$ and we have exactly $t-2k+2$ deletions;
- if $\mathrm{wt}(\mathbf{u_b}) = b+k$ then $\mathbf{u}_{b-k,0}^{a-k,1} = 0^{a+k}u_{2a+1}1^{b+k}$, we have exactly $t-2k+1$ deletions and it remains to delete the symbol u_{2a+1};
- if $\mathrm{wt}(\mathbf{u_b}) = b+k+1$ then $\mathbf{u}_{b-k-1,0}^{a-k,1} = 0^{a+k}u_{2a+1}1^{b+k+1}$, we have exactly $t-2k$ deletions and it remains to delete the symbol u_{2a+2} and one symbol 1.

Suppose there exist three unbalanced codewords. It follows from the above that for two of them, say $\mathbf{u}$, $\mathbf{v}$, there exist $\{p,q\} = \{0,1\}$ such that one obtains $p^{a+1}q^{b+1}$ by t deletions from both $\mathbf{u}$ and $\mathbf{v}$. Therefore $\mathrm{dd}(\mathbf{u},\mathbf{v}) \leq t$, a contradiction. This completes the proof. ◊

Proposition 3 There exist at most two balanced codewords.

Proof. Suppose there exist three balanced codewords $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$. Thus, $\mathrm{wt}(\mathbf{u_a}) = \mathrm{wt}(\mathbf{v_a}) = \mathrm{wt}(\mathbf{w_a}) = a$ and $\mathrm{wt}(\mathbf{u_b}) = \mathrm{wt}(\mathbf{v_b}) = \mathrm{wt}(\mathbf{w_b}) = b$.
Without lost of generality we have to consider two cases: $u_{2a+1} = v_{2a+1} = w_{2a+1} = 0$ and $u_{2a+1} = v_{2a+1} = 0$, $w_{2a+1} = 1$.
In the first case we may assume that $u_{2a+2} = v_{2a+2}$. Consequently

- if $u_{2a+2} = v_{2a+2} = 0$ then $\mathbf{u}_{b-1,0}^{a,0} = \mathbf{v}_{b-1,0}^{a,0} = 1^a001^b$,
- if $u_{2a+2} = v_{2a+2} = 1$ then $\mathbf{u}_{b-1,1}^{a,0} = \mathbf{v}_{b-1,1}^{a,0} = 1^a010^b$.

In the latter case if $u_{2a+2} = v_{2a+2}$ we have a contradiction as above. Hence, suppose $u_{2a+2} = 0$ and $v_{2a+2} = 1$. If $v_{2a} = 0$ then $\mathbf{u}_{b-1,0}^{a,0} = \mathbf{v}_{b,0}^{a-1,0} = 1^a001^b$. Therefore $v_{2a} = 1$. If $w_{2a+2} = 0$ then $\mathbf{v}_{b,0}^{a-1,1} = \mathbf{w}_{b-1,0}^{a,1} = 0^a101^b$. If $w_{2a} = 0$ then $\mathbf{v}_{b-1,1}^{a,0} = \mathbf{w}_{b,1}^{a-1,0} = 1^a010^b$. Therefore $w_{2a} = w_{2a+2} = 1$. Now $\mathbf{v}_{b-1,1}^{a-1,1} = 0^a1010^b$ and $\mathbf{w}_{b-1,1}^{a-1,1} = 0^a1110^b$ and thus from both vectors we obtain 0^a110^b. ◊
It follows from Propositions 2 and 3 that the code C consists of all 0's vector, all 1's vector, two balanced and two unbalanced codewords. In particular $M_2(5,1) \leq 6$ and Proposition 1(a) implies $M_2(5,1) = 6$.

Proposition 4 A k-unbalanced codeword for $k \geq 2$ does not exist.

Proof Suppose such a codeword exists. It follows from the proof of Proposition 2 that for $k \geq 2$ without lost of generality after $t-2$ deletions we obtain from a k-unbalanced codeword the vector
$1^{a+2}0^{b+2}$.
Note that deleting further two elements one obtains 1^a0^{b+2} or $1^{a+2}0^b$. Let $\mathbf{u}$ and $\mathbf{v}$ be the two balanced codewords. If $u_{2a+1} = u_{2a+2} = 1$ then $\mathbf{u}_{b-1,1}^{a,0} = 1^{a+2}0^b$, a contradiction. By analogy, if $u_{2a-1} = u_{2a} = 0$ then $\mathbf{u}_{b,1}^{a-1,0} = 1^a0^{b+2}$, a contradiction. The above two observations apply also for the codeword $\mathbf{v}$. It follows now that if $u_{2a+1} = v_{2a+1} = 0$ then $u_{2a-1} = v_{2a-1} = 1$ and therefore $\mathbf{u}_{b,1}^{a-1,1} = \mathbf{v}_{b,1}^{a-1,1} = 0^a10^{b+1}$. If $u_{2a+1} = v_{2a+1} = 1$ then $u_{2a+1} = v_{2a+1} = 0$ and therefore $\mathbf{u}_{b-1,0}^{a,1} = \mathbf{v}_{b-1,1}^{a,1} = 0^a101^b$. Hence, without lost of generality $u_{2a+1} = 0$ and $v_{2a+1} = 1$ implying that $u_{2a-1} = 1$ and $v_{2a+1} = 0$. Hence $\mathbf{u}_{b,0}^{a-1,1} = \mathbf{v}_{b-1,0}^{a,1} = 0^a101^b$. In all cases we reached a contradiction and therefore a k-unbalanced codeword for $k \geq 2$ does not exist. ◊
It follows from Proposition 4 that the two unbalanced codewords for t-deletion-correcting code of length $2t+3$ are 1-unbalanced.
Let $\mathbf{u}$ and $\mathbf{v}$ be the two unbalanced codewords, $\mathbf{w}$ and $\mathbf{t}$ be the two balanced codewords. Therefore at least one of $\mathbf{u_a}$ and $\mathbf{u_b}$ ($\mathbf{v_a}$ and $\mathbf{v_b}$ respectively) is 1-unbalanced.

Proposition 5 All four components of the two unbalanced codewords are 1-unbalanced.

Proof We have to consider the following cases:
Case A. Only $\mathbf{u_a}$ and $\mathbf{v_a}$ unbalanced.
Case B. Only $\mathbf{u_a}$ and $\mathbf{v_b}$ unbalanced.
Case C. Only $\mathbf{u_a}$, $\mathbf{u_b}$ and $\mathbf{v_a}$ unbalanced.

Case A. Suppose $\mathrm{wt}(\mathbf{u_a}) = \mathrm{wt}(\mathbf{v_a}) = a+1$. Since $\mathbf{u_b}$ is balanced (meaning that $\mathrm{wt}(\mathbf{u_b}) = b$) we have $\mathrm{wt}(\mathbf{u}) = a+b+1+u_{2a+1} = t+2+u_{2a+1} \leq t+2$ implying $u_{2a+1} = 0$. Analogously $v_{2a+1} = 0$ and we have $\mathbf{u}_{b,1}^{a-1,0} = \mathbf{v}_{b,1}^{a-1,0} = 1^{a+1}0^{b+1}$, a contradiction. The case $\mathrm{wt}(\mathbf{u_a}) = \mathrm{wt}(\mathbf{v_a}) = a-1$ is treated in the same way.
Therefore without lost of generality it remains to consider the case $\mathrm{wt}(\mathbf{u_a}) = a-1$ and $\mathrm{wt}(\mathbf{v_a}) = a+1$. Note that $\mathrm{wt}(\mathbf{u_b}) = \mathrm{wt}(\mathbf{v_b}) = b$ and as above we have $u_{2a+1} = 1$ and analogously $v_{2a+1} = 0$. Thus $\mathbf{u}_{b,1}^{a-1,1} = 0^{a+1}10^b$ and $\mathbf{v}_{b,0}^{a-1,0} = 1^{a+1}01^b$.
If $(w_{2a+1}, w_{2a+2}) = (0,1)$ ($(1,0)$, respectively) then $\mathbf{w}_{b-1,1}^{a,1} = 0^{a+1}10^b$ ($\mathbf{w}_{b-1,0}^{a,0} = 1^{a+1}01^b$, respectively), a contradiction to the above.
Therefore $w_{2a+1} = w_{2a+2}$ and similarly $t_{2a+1} = t_{2a+2}$. If $w_{2a+1} = t_{2a+1} = 0$ then $\mathbf{w}_{b-1,0}^{a,0} = \mathbf{t}_{b-1,0}^{a,0} = 1^a001^b$ and if $w_{2a+1} = t_{2a+1} = 1$ then $\mathbf{w}_{b-1,1}^{a,0} = \mathbf{t}_{b-1,1}^{a,0} =$

$1^{a+2}0^b$.
Thus, without lost of generality $w_{2a+1} = w_{2a+2} = 0$ and $t_{2a+1} = t_{2a+2} = 1$. Finally,

- if $v_{2a+2} = 0$ then $\mathbf{v}_{b-1,0}^{a-1,0} = 1^{a+1}001^b$ and $\mathbf{w}_{b-1,0}^{a,0} = 1^a001^b$;
- if $v_{2a+2} = 1$ then $\mathbf{v}_{b-1,1}^{a-1,0} = 1^{a+1}010^b$ and $\mathbf{t}_{b-1,1}^{a,0} = 1^{a+2}0^b$.

In all cases we have a contradiction.

Case B. Suppose $\mathrm{wt}(\mathbf{u_a}) = a - 1$ and $\mathrm{wt}(\mathbf{v_b}) = b + 1$. As in Case A. we obtain $u_{2a+1} = 1$ and $v_{2a+1} = 0$. Then $\mathbf{u}_{b,0}^{a-1,1} = \mathbf{v}_{b-1,0}^{a,1} = 0^{a+1}1^{b+1}$, a contradiction. The case $\mathrm{wt}(\mathbf{u_a}) = a + 1$ and $\mathrm{wt}(\mathbf{v_b}) = b - 1$ is settled in the same way.
Therefore we have to consider two cases: $\mathrm{wt}(\mathbf{u_a}) = a - 1, \mathrm{wt}(\mathbf{v_b}) = b - 1$ and $\mathrm{wt}(\mathbf{u_a}) = a + 1, \mathrm{wt}(\mathbf{v_b}) = b + 1$. By swapping $0 \leftrightarrow 1$ we see that the two cases are equivalent.
Thus, suppose that $\mathrm{wt}(\mathbf{u_a}) = a - 1$ and $\mathrm{wt}(\mathbf{v_b}) = b - 1$. Again, as in Case A. we obtain $u_{2a+1} = 1$ and $v_{2a+1} = 1$.
If $(t_{2a+1}, t_{2a+2}) = (0,1)$ then $\mathbf{t}_{b-1,1}^{a,1} = \mathbf{u}_{b,1}^{a-1,1} = 0^{a+1}10^b$ and if $(t_{2a}, t_{2a+1}) = (1,0)$ then $\mathbf{t}_{b-1,1}^{a-1,1} = \mathbf{v}_{b-1,1}^{a,1} = 0^a10^{b+1}$. It follows from the above that if $t_{2a+1} = 0$ ($w_{2a+1} = 0$, respectively) then $t_{2a} = t_{2a+2} = 0$ ($w_{2a} = w_{2a+2} = 0$, respectively).
We prove now that $w_{2a+1} \neq t_{2a+1}$.
If $w_{2a+1} = t_{2a+1} = 0$ then $w_{2a} = t_{2a} = w_{2a+2} = t_{2a+2} = 0$ and $\mathbf{t}_{b-1,0}^{a-1,0} = \mathbf{w}_{b-1,0}^{a-1,0} = 1^a0001^b$, a contradiction.
If $w_{2a+1} = t_{2a+1} = 1$ then as in the proof of Proposition 3 we obtain $w_{2a} \neq t_{2a}$. Assuming $w_{2a} = 0$ and $t_{2a} = 1$ we have

- if $v_{2a} = 1$ then $\mathbf{v}_{b-1,1}^{a-1,1} = 0^a110^{b+1}$ and $\mathbf{t}_{b-1,1}^{a-1,1} = 0^a1110^b$ give a contradiction;
- if $v_{2a} = 0$ then $\mathbf{v}_{b-1,1}^{a-1,0} = 1^a010^{b+1}$ and $\mathbf{w}_{b,1}^{a-1,0} = 1^a010^b$ give a contradiction.

Thus, without lost of generality $w_{2a+1} = 1$, $t_{2a+1} = 0$. Then $t_{2a} = t_{2a+2} = 0$ and

- if $u_{2a+2} = 0$ then $\mathbf{u}_{b-1,0}^{a-1,1} = 0^{a+1}101^b$ and $\mathbf{t}_{b-1,0}^{a-1,1} = 0^{a+2}1^b$, a contradiction (the vector $0^{a+2}1^b$ can be obtained from both $\mathbf{u}$ and $\mathbf{t}$);
- if $v_{2a-1} = 0$ then $\mathbf{v}_{b-1,1}^{a-1,0} = 1^a010^{b+1}$ and $\mathbf{t}_{b,1}^{a-1,0} = 1^a0^{b+2}$, a contradiction (the vector 0^a1^{b+2} can be obtained from both $\mathbf{v}$ and $\mathbf{t}$).

Therefore $u_{2a+2} = 1$ and $v_{2a-1} = 1$. Then $\mathbf{u}_{b-1,1}^{a-1,1} = 0^{a+1}110^b$ and $\mathbf{v}_{b-1,1}^{a-1,1} = 0^a110^{b+1}$. Hence, the vector 0^a110^b can be obtained from both $\mathbf{u}$ and $\mathbf{v}$. This completes the proof of Case B.

Case C. Without loss of generality assume $\mathrm{wt}(\mathbf{u_a}) = a - 1$ (otherwise swap 0 and 1 in all codewords). Hence, $\mathrm{wt}(\mathbf{u_b}) = b + 1$. It follows as in Case A. that $\mathrm{wt}(\mathbf{u_a}) = a + 1$. Since $\mathbf{v_b}$ is balanced we conclude that $v_{2a+1} = 0$.
Assume first that $u_{2a+1} = 1$.

If $w_{2a+1} = t_{2a+1} = 0$ then as in the proof of Proposition 3 we obtain $w_{2a} = w_{2a+2} = 0$ and $t_{2a} = t_{2a+2} = 1$. Now:

- If $v_{2a+2} = 0$ then $\mathbf{v}_{b-1,0}^{a-1,0} = 1^{a+1}001^b$ and $\mathbf{w}_{b-1,0}^{a,0} = 1^a001^b$ and after deleting an 1 from the first vector we get a contradiction;
- if $v_{2a+2} = 1$ then $\mathbf{v}_{b-1,1}^{a-1,0} = 1^{a+1}010^b$ and $\mathbf{t}_{b-1,1}^{a,0} = 1^a010^b$ and after deleting an 1 from the first vector we get a contradiction.

Therefore at least one of w_{2a+1} and t_{2a+1} is not 0 and assume that $w_{2a+1} = 1$. Then $w_{2a} = 0$ (if $w_{2a} = 1$ then $\mathbf{u}_{b-1,0}^{a-1,1} = 0^{a+1}1^{b+2}$ and $\mathbf{w}_{b,0}^{a-1,1} = 0^a1^{b+2}$) and $w_{2a+2} = 1$ (if $w_{2a+2} = 0$ then $\mathbf{v}_{b,0}^{a-1,0} = \mathbf{w}_{b-1,0}^{a,0} = 1^{a+1}01^b$).
If $t_{2a+1} = 1$ then by analogy $t_{2a} = 0$ and $t_{2a+2} = 1$. Hence, $\mathbf{w}_{b-1,1}^{a-1,0} = \mathbf{t}_{b-1,1}^{a-1,0} = 1^a0110^b$, a contradiction.
Let $t_{2a+1} = 0$. If $t_{2a+2} = 1$ then $\mathbf{w}_{b,1}^{a-1,0} = \mathbf{t}_{b-1,1}^{a,0} = 1^a010^b$. Hence, $t_{2a+2} = 0$. Now:

- if $v_{2a+2} = 0$ then $\mathbf{v}_{b-1,0}^{a-1,0} = 1^{a+1}001^b$ and $\mathbf{t}_{b-1,0}^{a,0} = 1^a001^b$;
- if $v_{2a+2} = 1$ then $\mathbf{v}_{b-1,1}^{a-1,0} = 1^{a+1}010^b$ and $\mathbf{w}_{b,1}^{a-1,0} = 1^a010^b$.

Therefore $u_{2a+1} = 0$. Then $\mathbf{u}_{b-1,0}^{a-1,1} = 0^{a+2}1^{b+1}$and $\mathbf{v}_{b,0}^{a-1,0} = 1^{a+1}01^b$.
If $w_{2a+2} = 0$ then:

- if $w_{2a+1} = 0$ then $\mathbf{w}_{b,0}^{a,1} = 0^{a+2}1^b$;
- if $w_{2a+1} = 1$ then $\mathbf{w}_{b,0}^{a,0} = 1^{a+1}01^b$.

Therefore $w_{2a+2} = 1$ and by analogy $t_{2a+1} = 1$. If $w_{2a+1} = t_{2a+1} = p$ then $\mathbf{w}_{b-1,1}^{a,0} = \mathbf{t}_{b-1,1}^{a,0} = 1^ap10^b$. Therefore $w_{2a+1} \neq t_{2a+1}$ and without lost of generality assume $w_{2a+1} = 0$ and $t_{2a+1} = 1$.
If $t_{2a} = 0$ then $\mathbf{w}_{b-1,1}^{a,0} = \mathbf{t}_{b,1}^{a-1,0} = 1^a010^b$ and thus $t_{2a} = 1$.
If $w_{2a} = 1$then $\mathbf{w}_{b-1,1}^{a-1,1} = 0^a1w_{2a+1}10^b$ and$\mathbf{t}_{b-1,1}^{a-1,1} = 0^a1t_{2a+1}10^b$and deleting the coordinate in the middle gives contradiction. Therefore $w_{2a} = 0$. Now

- if $v_{2a+2} = 1$ then $\mathbf{v}_{b-1,1}^{a-1,0} = 1^{a+1}010^b$ and $\mathbf{w}_{b-1,1}^{a,0} = 1^a010^b$ and after deleting an 1 from the first vector we obtain a contradiction;
- if $v_{2a+2} = 0$ then $\mathbf{v}_{b-1,0}^{a-1,0} = 1^{a+1}001^b$ and $\mathbf{w}_{b-1,1}^{a-1,0} = 1^a001^b$ and after deleting an 1 from the first vector we obtain a contradiction.

In all cases we have reached a contradiction. This completes the proof of Proposition 5. ◊

Proposition 6 For any codeword $\mathbf{u}$ we have $u_3 = u_5 = \ldots = u_{2t+1}$.

Proof Let $\mathbf{u}$ be a codeword. For all 1's and all 0's codewords the statement is obvious. Recall that otherwise $\mathrm{wt}(\mathbf{u}) = t + 1$ or $t + 2$. Fix positive integers $a = 1, 2, \ldots, t$ and b such that $a + b = t + 1$. Suppose that for these particular a and b the

codeword $\mathbf{u}$ is balanced. Since $\mathrm{wt}(\mathbf{u_a}) = a$ and $\mathrm{wt}(\mathbf{u_b}) = b$ we have that u_{2a+1} appears $t+2$ times in $\mathbf{u}$. Next, suppose that for these particular a and b the codeword $\mathbf{u}$ is unbalanced. It follows from Proposition 5 that both $\mathbf{u_a}$ and $\mathbf{u_b}$ are 1-unbalanced. Note that $\mathrm{wt}(\mathbf{u_a}) = a - 1$ and $\mathrm{wt}(\mathbf{u_b}) = \mathbf{b} - \mathbf{1}$ is impossible since then $wt(u) = \mathrm{wt}(\mathbf{u_a}) + u_{2a+1} + \mathrm{wt}(\mathbf{u_b}) = a + b - 2 + u_{2a+1} = t - 1 + u_{2a+1} \leq t$.
Similarly, $\mathrm{wt}(\mathbf{u_a}) = a + 1$ and $\mathrm{wt}(\mathbf{u_b}) = b + 1$ is also impossible. Therefore $\mathrm{wt}(\mathbf{u_a}) + \mathrm{wt}(\mathbf{u_b}) = a + b = t + 1$ and we have that u_{2a+1} appears $t+2$ times in $\mathbf{u}$. Hence, for any $a = 1, \dots, t$ the symbol u_{2a+1} appears $t+2$ times in $\mathbf{u}$. Thus, $u_3 = \cdots = u_{2t+1}$. ⋄

Proposition 7 For $t = 2, 3$ an optimal t-deletion-correcting code of length $2t+3$ has cardinality 5.

Sketch proof Let the code C of length 7 and cardinality 6 be an optimal 2-deletion-correcting code of length 7. Without lost of generality let $\mathbf{u} = (u_1, u_2, \dots, u_7)$ and $\mathbf{v} = (v_1, v_2, \dots, v_7)$ be two codewords for which $\mathrm{wt}(\mathbf{u}) = \mathrm{wt}(\mathbf{v}) = 4$. It follows that $u_3 = u_5 = v_3 = v_5 = 1$. Suppose $u_s = v_s = 1$ for some $s \neq 3, 5$. Since $\mathrm{wt}(\mathbf{u}) = \mathrm{wt}(\mathbf{v}) = 4$ we have that there exist only two coordinates p and q such that $u_p \neq v_p$ and $u_q \neq v_q$. By deleting these two coordinates from both $\mathbf{u}$ and $\mathbf{v}$ we obtain one and the same vector of length 5, a contradiction. If $u_1 = v_1 = 0$ (respectively $u_7 = v_7 = 0$) then by deleting all but the first (respectively the last) 0 from $\mathbf{u}$ and $\mathbf{v}$ we obtain one and the same vector. Therefore$u_1 \neq v_1$ and $u_7 \neq v_7$. An easy enumeration gives only two options for $\mathbf{u}$ and $\mathbf{v}$:
$\mathbf{u} = (0011110)$, $\mathbf{v} = (1010101)$
and
$\mathbf{u} = (0010111)$, $\mathbf{v} = (1110100)$
It is easy to see that for the remaining two codewords $\mathbf{w}$ and $\boldsymbol{t}$ we have $w_3 = w_5 = t_3 = t_5 = 0$. The same arguments (the details are left to the reader) as above imply all possible values of $\mathbf{w}$ and $\boldsymbol{t}$. In all cases there exist two codewords of deletion distance 2.
The case $t = 3$ is settled in similar way.

We are ready to proceed to the proof of the main part of Theorem 1.
Proof of Theorem 1 (b) for $t \geq 4$. We prove now that for $t \geq 4$ an optimal t-deletion-correcting code of length $2t+3$ has cardinality 5.
Assume C is t-deletion-correcting code of cardinality 6 for $t \geq 4$. We may assume that there exist two codewords $\mathbf{u}$ and $\mathbf{v}$ for which $\mathrm{wt}(\mathbf{u}) = \mathrm{wt}(\mathbf{v}) = t + 2$ and therefore $u_3 = u_5 = \ \dots = u_{2t+1} = v_3 = v_5 = \ \dots = v_{2t+1} = 1$
This implies that there are at most 4 positions in which $\mathbf{u}$ and $\mathbf{v}$ differ. By deleting all elements in $\mathbf{u}$ and $\mathbf{v}$ in these 4 positions we obtain two equal vectors. Since there are 4 deletions and $t \geq 4$ we have a contradiction with $\mathrm{dd}(C) > t$.
The claims of Proposition 1 and Proposition 7 complete the proof of the Theorem 1. ⋄

References

[1] Levenshtein V. I., Binary codes capable of correcting, deletions, insertions and reversals, *Doklady Akad. Nauk SSSR* **1965**(163), 845–848.

[2] Levenshtein V. I., Binary codes capable of correcting spurious insertions and deletions of ones, *Problemy Peredchi Informacii* **1965**(1), 12–25.

[3] Sloane N. J. A., On Single-Deletion-Coirrecting Codes, In: Codes and Designs: Proceedings of a Conference Honoring Professor Dijen Ray-Chaudhuri, Walter De Gruyter, **2002**, 273–291.

[4] Levenshtein V. I., On perfect codes in the deletion-insertion metric, *Diskretnaya Matematika* **1991**(3), 3–20.

[5] Levenshtein V. I., Efficient reconstruction of sequences, *IEEE Trans Inf. Theory* **2001**(47), 2–22.

[6] Swart T. G., Fereira H. C., A note on double insertion/deletion correcting codes, *IEEE Trans. Inf. Theory* **2003**(49), 269-272.

[7] Tolhuizen L., Upper bounds on the size of insertion/deletion codes, Proc 8th Int. Workshop on ACCT, tsarskoe selo, Russia, September 2002. 242–246.

[8] Helberg A. S. J., Fereira H. C., On multiple insertion/deletion correcting codes, *IEEE Trans. Inf. Theory* **2002**(48), 305–308.

[9] Landjev I., Haralambiev Kr., On multiple deletion codes, Serdica J. Computing **2006**(1), 13–26.