The Error Probability of Maximum-Likelihood Decoding for the t-Deletion Channel

Daniella Bar-Lev[*†], **Yotam Gershon**[*‡], **Omer Sabary**[*§], and **Eitan Yaakobi**[†]
[†] Department of Computer Science, Technion — Israel Institute of Technology, Haifa, 3200003 Israel
[‡] Department of Electrical Engineering, Technion — Israel Institute of Technology, Haifa, 3200003 Israel
[§] Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093, USA
Emails: `{daniellalev,yaakobi}@cs.technion.ac.il, yotamgr@campus.technion.ac.il, osabary@ucsd.edu`

***Abstract*—This paper studies the error probability of maximum-likelihood (ML) decoding for a special case of the deletion channel, referred by the *t-deletion channel*, which deletes exactly t symbols of the transmitted word uniformly at random. The goal of the paper is to understand how an ML decoder operates in order to minimize the average decoding *error* probability (as opposed to the average decoding *failure* probability). A full characterization of the ML decoder for this setup is given for a channel that deletes one or two symbols, that is, $t = 1, 2$. For $t = 1$ it is shown that if the code is the entire space then the ML decoder is the *lazy decoder* which simply returns the channel output. Similarly, for $t = 2$ it is shown that the ML decoder acts as the lazy decoder in almost all cases and for the rest, when the longest run is significantly long, it prolongs the longest run by exactly one symbol.**

***Index Terms*—Deletion channel, insertion channel, sequence reconstruction.**

I. Introduction

Codes correcting insertions/deletions recently attracted considerable attention due to their relevance to the special error behavior in DNA-based data storage [3], [22], [25], [30], [32], [33], [42], [43]. These codes are relevant for other applications in communication models. For example, insertions/deletions happen during the synchronization of files and symbols of data streams [34] or due to over-sampling and under-sampling at the receiver side [12]. The algebraic concepts correcting insertions/deletions date back to the 1960s when Varshamov and Tenengolts designed a class of binary codes, nowadays called *VT codes* [39]. These codes were originally designed to correct a single asymmetric error and later were proven to correct a single insertion/deletion [26]. Extensions for multiple deletions were recently proposed in several studies; see e.g. [4], [16], [35], [36]. However, while codes correcting substitution errors were widely studied and efficient capacity achieving codes both for short and large block lengths are used conventionally, much less is known for codes correcting insertions/deletions. More than that, even the deletion channel capacity is far from being solved [6]–[8], [10], [28], [29], [31].

There are two main models which are studied for deletion errors. While in the first one, the goal is to correct a fixed number of deletions in the worse case, for the second one, which corresponds to the channel capacity of the deletion channel, one seeks to construct codes which correct a fraction p of deletions with high probability [5], [7], [9], [11], [13], [15], [23], [24], [29], [38], [40]. This paper considers a combination of these two models. In this channel, referred as the *t-deletion channel*, t symbols of the transmitted word are deleted uniformly at random. Consider for example the case of $t = 1$, i.e., one of the n transmitted symbols is deleted, each with the same probability. In case the transmitted word belongs to a single-deletion-correcting code then clearly it is possible to successfully decode the transmitted word. However, if such error correction capability is not guaranteed in the worst case, two approaches can be of interest. In the first, one may output a list of all possible transmitted words, that is, *list decoding* for deletion errors as was studied recently in several works; see e.g. [18]–[21], [23], [27], [41]. The second one, which is taken in the present work, seeks to output a word that minimizes the decoding error probability. This channel was also studied in several previous works. In [17], the author studied the maximal length of words that can be uniquely reconstructed using a sufficient number of channel outputs of the t-deletion channel and calculated this maximal length explicitly for $n - t \leqslant 6$. In [2], the goal was to study the entropy of the set of the potentially channel input words given a corrupted word from a channel that deletes either one or two bits. The minimum and maximum values of this entropy value were explored. Another variation of this channel was studied in [1].

Mathematically speaking, assume S is a channel that is characterized by a conditional probability $\mathsf{Pr}_{\mathsf{S}}\{\boldsymbol{y} \text{ rec. } | \boldsymbol{x} \text{ trans.}\}$, for every pair $(\boldsymbol{x}, \boldsymbol{y}) \in (\Sigma_q^*)^2$. A decoder for a code $\mathcal{C}$ with respect to the channel S is a function $\mathcal{D} : \Sigma_q^* \to \mathcal{C}$. Its *average decoding failure probability* is the probability that the decoder output is not the transmitted word. The *maximum-likelihood* (*ML*) *decoder* for $\mathcal{C}$ with respect to S, denoted by $\mathcal{D}_{\mathsf{ML}}$, outputs a codeword $\boldsymbol{c} \in \mathcal{C}$ that maximizes the probability $\mathsf{Pr}_{\mathsf{S}}\{\boldsymbol{y} \text{ rec. } | \boldsymbol{c} \text{ trans.}\}$. This decoder minimizes the average decoding *failure* probability and thus it outputs only codewords. However, if one seeks to minimize the average decoding *error* probability, then the decoder should consider non-codewords as well. The *average decoding error probability* is the average normalized distance between the transmitted word and the decoder's output, where the distance function depends upon the channel of interest. The goal of this work is to study the *ML* decoder*, which outputs words that minimize the average decoding error probability, for the t-deletion channel.

The rest of the paper is organized as follows. Section II presents the formal definition of channel transmission and

*The three first authors contributed equally to this work.

maximum likelihood decoding in order to minimize the average decoding error probability. This section introduces also the t-deletion channel. Section III studies the 1-deletion channel. It introduces two types of decoders. The first one, referred as the *embedding number decoder*, maximizes the so-called *embedding number* between the channel output and all possible codewords. The second one is called the *lazy decoder* and it simply returns the channel output. The main result of this section states that if the code is the entire space then the ML* decoder is the lazy decoder. Similarly, Section IV studies the 2-deletion channel where it is shown that in almost all cases the ML* decoder should act as the lazy decoder and in the rest of the cases it returns a length-$(n-1)$ word which maximizes the embedding number. Due to the lack of space some of the proofs are omitted from this paper, however they can be found in the extended version of the paper [**?**].

II. Definitions and Preliminaries

Denote by $\Sigma_q = \{0,\ldots,q-1\}$ the alphabet of size q and $\Sigma_q^* \triangleq \bigcup_{\ell=0}^{\infty}\Sigma_q^\ell$. The length of $\boldsymbol{x}\in\Sigma^n$ is denoted by $|\boldsymbol{x}| = n$. The *Levenshtein distance* between two words $\boldsymbol{x},\boldsymbol{y}\in\Sigma_q^*$, denoted by $d_L(\boldsymbol{x},\boldsymbol{y})$, is the minimum number of insertions and deletions required to transform $\boldsymbol{x}$ into $\boldsymbol{y}$, and $d_H(\boldsymbol{x},\boldsymbol{y})$ denotes the *Hamming distance* between $\boldsymbol{x}$ and $\boldsymbol{y}$, when $|\boldsymbol{x}| = |\boldsymbol{y}|$. For a word $\boldsymbol{x}\in\Sigma_q^*$ and a set of indices $I\subseteq[|\boldsymbol{x}|]$, the word $\boldsymbol{x}_I$ is the *projection* of $\boldsymbol{x}$ on the indices of I which is the subsequence of $\boldsymbol{x}$ received by the symbols in the entries of I. For two words $\boldsymbol{x},\boldsymbol{y}\in\Sigma_q^*$, the number of times that $\boldsymbol{y}$ can be received as a subsequence of $\boldsymbol{x}$ is called the *embedding number of* $\boldsymbol{y}$ *in* $\boldsymbol{x}$ [2], [14], [37], defined by $\mathsf{Emb}(\boldsymbol{x};\boldsymbol{y}) = |\{I\subseteq[|\boldsymbol{x}|] \mid \boldsymbol{x}_I = \boldsymbol{y}\}|$. Note that if $\boldsymbol{y}$ is not a subsequence of $\boldsymbol{x}$ then $\mathsf{Emb}(\boldsymbol{x};\boldsymbol{y}) = 0$.

The *radius-r insertion ball* of a word $\boldsymbol{x}\in\Sigma_q^*$, denoted by $I_r(\boldsymbol{x})$, is the set of all supersequences of $\boldsymbol{x}$ of length $|\boldsymbol{x}|+r$. From [26] it is known that $I_r(\boldsymbol{x}) = \sum_{i=0}^{r}\binom{|\boldsymbol{x}|+r}{i}$. Similarily, the *radius-r deletion ball* of a word $\boldsymbol{x}\in\Sigma_q^*$, denoted by $D_r(\boldsymbol{x})$, is the set of all subsequences of $\boldsymbol{x}$ of length $|\boldsymbol{x}|-r$.

We consider a channel S that is characterized by a conditional probability Pr_S, defined by $\mathsf{Pr}_\mathsf{S}\{\boldsymbol{y} \text{ rec. } |\boldsymbol{x} \text{ trans.}\}$, for all $(\boldsymbol{x},\boldsymbol{y})\in(\Sigma_q^*)^2$. Note that the lengths of the input and output words may not be the same as we consider deletions in this work. A decoder for a code $\mathcal{C}$ with respect to the channel S is a function $\mathcal{D}:\Sigma_q^*\to\mathcal{C}$. Its *average decoding failure probability* is defined by $\mathsf{P}_{\mathsf{fail}}(\mathsf{S},\mathcal{C},\mathcal{D}) = \frac{\sum_{c\in\mathcal{C}}\mathsf{P}_{\mathsf{fail}}(\boldsymbol{c})}{|\mathcal{C}|}$, where

$$\mathsf{P}_{\mathsf{fail}}(\boldsymbol{c}) = \sum_{\boldsymbol{y}:\mathcal{D}(\boldsymbol{y})\neq\boldsymbol{c}}\mathsf{Pr}_\mathsf{S}\{\boldsymbol{y} \text{ rec. } |\boldsymbol{c} \text{ trans.}\}.$$

We will mostly be interested in the *average decoding error probability* which is the average normalized distance between the transmitted word and the decoder's output. The distance will depend upon the channel. For example, for the BSC one should consider the Hamming distance, while for insertion/deletion channels, the Levenshtein distance will be of interest. Hence, for a channel S, distance function d, and a decoder $\mathcal{D}$, we let $\mathsf{P}_{\mathsf{err}}(\mathsf{S},\mathcal{C},\mathcal{D},d) = \frac{\sum_{c\in\mathcal{C}}\mathsf{P}_{\mathsf{err}}(\boldsymbol{c},d)}{|\mathcal{C}|}$, where

$$\mathsf{P}_{\mathsf{err}}(\boldsymbol{c},d) = \sum_{\boldsymbol{y}:\mathcal{D}(\boldsymbol{y})\neq\boldsymbol{c}}\frac{d(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c})}{|\boldsymbol{c}|}\cdot\mathsf{Pr}_\mathsf{S}\{\boldsymbol{y} \text{ rec. } |\boldsymbol{c} \text{ trans.}\}.$$

The *maximum-likelihood* (*ML*) *decoder* for $\mathcal{C}$ with respect to a channel S, denoted by $\mathcal{D}_{\mathsf{ML}}$, outputs a codeword $\boldsymbol{c}\in\mathcal{C}$ that maximizes the probability $\mathsf{Pr}_\mathsf{S}\{\boldsymbol{y} \text{ rec. } |\boldsymbol{c} \text{ trans.}\}$. That is, for $\boldsymbol{y}\in\Sigma_q^*$, $\mathcal{D}_{\mathsf{ML}}(\boldsymbol{y}) = \arg\max_{\boldsymbol{c}\in\mathcal{C}}\{\mathsf{Pr}_\mathsf{S}\{\boldsymbol{y} \text{ rec. } |\boldsymbol{c} \text{ trans.}\}\}$. It is well known that for the BSC, the ML decoder chooses the closest codeword with respect to the Hamming distance.

Note that channels which introduce deletions or insertions change the sequence's length. If the goal is to minimize the average decoding *failure* probability then clearly the decoder's output should be a codeword as there is no point in outputting a non-codeword. However, if one seeks to minimize the average decoding *error* probability, then the decoder should consider non-codewords as well. Therefore, we present here the ML* decoder, which is an alternative definition of the ML decoder that takes into account non-codewords and in particular words with different length than the code length. The *maximum-likelihood** (*ML**) *decoder* for $\mathcal{C}$ with respect to a channel S, denoted by $\mathcal{D}_{\mathsf{ML}^*}$, should output words that minimize the average decoding error probability $\mathsf{P}_{\mathsf{err}}(\mathsf{S},\mathcal{C},\mathcal{D},d)$:

$$\begin{aligned}\mathsf{P}_{\mathsf{err}}(\mathsf{S},\mathcal{C},\mathcal{D},d) &= \frac{1}{|\mathcal{C}|}\sum_{\boldsymbol{c}\in\mathcal{C}}\mathsf{P}_{\mathsf{err}}(\boldsymbol{c},d)\\ &= \frac{1}{|\mathcal{C}|}\sum_{\boldsymbol{c}\in\mathcal{C}}\sum_{\boldsymbol{y}:\mathcal{D}(\boldsymbol{y})\neq\boldsymbol{c}}\frac{d(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c})}{|\boldsymbol{c}|}\cdot\mathsf{Pr}_\mathsf{S}\{\boldsymbol{y} \text{ rec. } |\boldsymbol{c} \text{ trans.}\}\\ &\overset{\text{(a)}}{=} \frac{1}{|\mathcal{C}|}\sum_{\boldsymbol{y}\in\Sigma_q^*}\sum_{\boldsymbol{c}:\mathcal{D}(\boldsymbol{y})\neq\boldsymbol{c}}\frac{d(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c})}{|\boldsymbol{c}|}\mathsf{Pr}_\mathsf{S}\{\boldsymbol{y} \text{ rec. } |\boldsymbol{c} \text{ trans.}\},\end{aligned}$$

where in (a) we switched the summation order, while taking into account all possible channel's outputs. For every $\boldsymbol{y}\in\Sigma_q^*$, denote the value $\sum_{\boldsymbol{c}:\mathcal{D}(\boldsymbol{y})\neq\boldsymbol{c}}\frac{d(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c})}{|\boldsymbol{c}|}\mathsf{Pr}_\mathsf{S}\{\boldsymbol{y} \text{ rec. } |\boldsymbol{c} \text{ trans.}\}$ by $f_{\boldsymbol{y}}(\mathcal{D}(\boldsymbol{y}))$ and if $\mathcal{D}(\boldsymbol{y})$ is some arbitrary value $\boldsymbol{x}$, this value is denoted by $f_{\boldsymbol{y}}(\boldsymbol{x})$. Thus, the ML* decoder is defined as

$$\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y}) = \underset{\boldsymbol{x}\in\Sigma_q^*}{\operatorname{argmin}}\{f_{\boldsymbol{y}}(\boldsymbol{x})\}.$$

In this paper we study the ML* decoder for a special case of the deletion channel that is denoted by t-Del and is referred as the *t-deletion channel*. In this channel, defined also in [17], exactly t symbols of the transmitted word are deleted. The t symbols are selected randomly and independently out of the $\binom{n}{t}$ options to delete t out of the n symbols, where n is the word length. Note that it may happen that different deletion patterns will still result with the same output. In this work, whenever the set $\arg\min_{\boldsymbol{x}\in\Sigma_q^*}\{f_{\boldsymbol{y}}(\boldsymbol{x})\}$ contains more than one word, we assume that $\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})$ returns a word of minimum length. Section III is dedicated for the case of $t=1$, while in Section IV the $t=2$ case is solved. In both cases we provide a full characterization of the ML* decoder and its average decoding error probability when the code is Σ_2^n. In the analysis to follow in this paper, when the channel being discussed is clear from the context, the conditional probability $\mathsf{Pr}_\mathsf{S}\{\boldsymbol{y} \text{ rec. } |\boldsymbol{c} \text{ trans.}\}$ will be denoted by $p(\boldsymbol{y}|\boldsymbol{c})$.

III. The 1-Deletion Channel

In this section we consider the 1-deletion channel which deletes one symbol randomly. Given a single-deletion-correcting code, any channel output can be easily decoded, and therefore for the rest of this section we assume that the given code is not a single-deletion-correcting code. We start by examining two types of decoders for this channel. The first decoder, referred as the *embedding number decoder* and denoted by $\mathcal{D}_{EN}$, returns for a channel output $\boldsymbol{y}$ the word $\mathcal{D}_{EN}(\boldsymbol{y})$ which is a codeword in the code $\mathcal{C}$ that maximizes the embedding number of $\boldsymbol{y}$ in $\mathcal{D}_{EN}(\boldsymbol{y})$. That is,

$$\mathcal{D}_{EN}(\boldsymbol{y}) = \arg\max_{\boldsymbol{c}\in\mathcal{C}}\{\mathsf{Emb}(\boldsymbol{c};\boldsymbol{y})\},$$

where, for now, if there is more than one such a word the decoder chooses one of them arbitrarily. The second decoder, referred as the *lazy decoder*, is denoted by $\mathcal{D}_{Lazy}$. For a channel output $\boldsymbol{y}$, $\mathcal{D}_{Lazy}$ simply returns $\boldsymbol{y}$ as the output, i.e., $\mathcal{D}_{Lazy}(\boldsymbol{y}) = \boldsymbol{y}$. Note that the lazy decoder does not return a codeword. Additionally, $d_L(\mathcal{D}_{Lazy}(\boldsymbol{y}),\boldsymbol{c}) = 1$ since $\boldsymbol{y}\in D_1(\boldsymbol{c})$ and hence, the average decoding error probability of the lazy decoder is $\frac{1}{n}$, when n is the code length.

In the main result of this section, presented in Theorem 11, we prove for $\mathsf{S} = \mathsf{1\text{-}Del}$ and $\mathcal{C} = \Sigma_2^n$, that $\mathcal{D}_{Lazy}$ performs at least as good as any other decoder, and hence $\mathcal{D}_{Lazy} = \mathcal{D}_{\mathsf{ML}^*}$.

For the rest of this section it is assumed that $\mathcal{C}\subseteq\Sigma_2^n$ and $\mathsf{S} = \mathsf{1\text{-}Del}$. Under this setup, the Levenshtein distance between the lazy decoder's output $\boldsymbol{y}$ and the transmitted word $\boldsymbol{c}$ is always $d_L(\boldsymbol{y},\boldsymbol{c}) = 1$, since $\boldsymbol{y}\in D_1(\boldsymbol{c})$. Hence, the following lemma follows immediately.

Lemma 1. *The average decoding error probability of the lazy decoder* $\mathcal{D}_{\text{Lazy}}$ *under the 1-deletion channel* $\mathsf{1\text{-}Del}$ *is*

$$P_{\text{err}}(\mathsf{1\text{-}Del},\mathcal{C},\mathcal{D}_{\text{Lazy}},d_L) = \frac{1}{n}.$$

Proof: The average decoding error probability of the lazy decoder for each codeword $\boldsymbol{c}$ is calculated as follows.

$$\begin{aligned} P_{\text{err}}(\boldsymbol{c},d_L) &= \sum_{\boldsymbol{y}:\mathcal{D}_{\text{Lazy}}(\boldsymbol{y})\neq\boldsymbol{c}} \frac{d_L\left(\mathcal{D}_{\text{Lazy}}(\boldsymbol{y}),\boldsymbol{c}\right)}{|\boldsymbol{c}|}p(\boldsymbol{y}|\boldsymbol{c}) \\ &= \sum_{\boldsymbol{y}\in D_1(\boldsymbol{c})}\frac{1}{n}p(\boldsymbol{y}|\boldsymbol{c}) = \frac{1}{n}. \end{aligned}$$

Since this is true for every $\boldsymbol{c}\in\mathcal{C}$, we get that

$$P_{\text{err}}(\mathsf{1\text{-}Del},\mathcal{C},\mathcal{D}_{\text{Lazy}},d_L) = \frac{1}{n}\cdot|\mathcal{C}|\cdot\frac{1}{|\mathcal{C}|} = \frac{1}{n}.$$

■

We can now show that the lazy decoder is preferable, with respect to the average decoding error probability, over any decoder that outputs a word of the same length as its input.

Lemma 2. *Let* $\mathcal{D}:(\Sigma_2)^{n-1}\to(\Sigma_2)^{n-1}$ *be a general decoder that preserves the channel's output length word length. It follows that*

$$P_{\text{err}}(\mathsf{1\text{-}Del},\mathcal{C},\mathcal{D},d_L) \geqslant P_{\text{err}}(\mathsf{1\text{-}Del},\mathcal{C},\mathcal{D}_{\text{Lazy}},d_L),$$

and for $\mathcal{C} = (\Sigma_2)^n$ *equality is obtained if and only if* $\mathcal{D} = \mathcal{D}_{\text{Lazy}}$.

Proof: Equality is trivial when $\mathcal{D} = \mathcal{D}_{\text{Lazy}}$. Furthermore, since for every $\boldsymbol{y}\in\mathcal{C}$ it holds that $|\mathcal{D}(\boldsymbol{y})| = n-1$, it is deduced that $d_L(\boldsymbol{c},\mathcal{D}(\boldsymbol{y}))\geqslant 1$. Hence, similarly to the proof of Lemma 1, it is easy to verify that

$$P_{\text{err}}(\mathsf{1\text{-}Del},\mathcal{C},\mathcal{D},d_L)\geqslant\frac{1}{n} = P_{\text{err}}(\mathsf{1\text{-}Del},\mathcal{C},\mathcal{D}_{\text{Lazy}},d_L),$$

where the last equality follows from Lemma 1.

Let us now assume that $\mathcal{D}\neq\mathcal{D}_{\text{Lazy}}$, i.e., there exists $\boldsymbol{z}\in(\Sigma_2)^{n-1}$ such that $\mathcal{D}(\boldsymbol{z}) = \boldsymbol{z}'\neq\boldsymbol{z}$. Since $\boldsymbol{z}'\neq\boldsymbol{z}$ we get that $I_1(\boldsymbol{z}')\neq I_1(\boldsymbol{z})$, i.e., there exists a word $\boldsymbol{c}\in(\Sigma_2)^n$ such that $\boldsymbol{c}\in I_1(\boldsymbol{z})$ and $\boldsymbol{c}\notin I_1(\boldsymbol{z}')$. Equivalently, $\boldsymbol{z}\in D_1(\boldsymbol{c})$ and $\boldsymbol{z}'\notin D_1(\boldsymbol{c})$, and so $d_L(\boldsymbol{c},\boldsymbol{z}')\geqslant 3$ (at least one more deletion and one more insertion are needed in addition to the insertion needed for every word in the deletion ball).

Hence, it is derived that

$$\begin{aligned} P_{\text{err}}(\boldsymbol{c},d_L) &= \sum_{\boldsymbol{y}\in D_1(\boldsymbol{c})}\frac{d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c})}{n}p(\boldsymbol{y}|\boldsymbol{c}) \\ &\geqslant \sum_{\boldsymbol{y}\in D_1(\boldsymbol{c})\setminus\{\boldsymbol{z}\}}\frac{1}{n}p(\boldsymbol{y}|\boldsymbol{c}) + \frac{d_L(\mathcal{D}(\boldsymbol{z})=\boldsymbol{z}',\boldsymbol{c})}{n}\cdot p(\boldsymbol{y}|\boldsymbol{c}) \\ &> \sum_{\boldsymbol{y}\in D_1(\boldsymbol{c})}\frac{1}{n}p(\boldsymbol{y}|\boldsymbol{c}) = \frac{1}{n}. \end{aligned}$$

If $\mathcal{C} = (\Sigma_2)^n$ it must hold that $\boldsymbol{c}\in\mathcal{C}$, and so

$$P_{\text{err}}(\mathsf{1\text{-}Del},(\Sigma_2)^n,\mathcal{D},d_L)\geqslant\frac{|\mathcal{C}|-1}{|\mathcal{C}|}\cdot\frac{1}{n}+\frac{1}{|\mathcal{C}|}P_{\text{err}}(\boldsymbol{c},d_L) > \frac{1}{n}.$$

Combining with Lemma 1 again completes the proof. ■

Before examining the performance of the embedding number decoder, we first discuss its properties over the 1-deletion channel. It is first shown that a decoder that prolongs an arbitrary run of maximal length within the input word is equivalent to the embedding number decoder.

Lemma 3. *Given* $\boldsymbol{y}\in(\Sigma_2)^{n-1}$, *the word* $\widehat{\boldsymbol{x}}\in(\Sigma_2)^n$ *obtained by prolonging a run of maximal length in* $\boldsymbol{y}$ *satisfies*

$$\mathsf{Emb}(\widehat{\boldsymbol{x}};\boldsymbol{y}) = \max_{\boldsymbol{x}\in\Sigma_2^n}\{\mathsf{Emb}(\boldsymbol{x};\boldsymbol{y})\}.$$

Proof: Let $\boldsymbol{y}$ be a word with n_r runs of lengths $r_1, r_2, \ldots, r_{n_r}$. Let $\boldsymbol{x}_0$ be any word obtained from $\boldsymbol{y}$ by creating a new run of length one, and so $\mathsf{Emb}(\boldsymbol{x}_0;\boldsymbol{y}) = 1$. Let $\boldsymbol{x}_i$, $1\leqslant i\leqslant n_r$ be the word obtained from $\boldsymbol{y}$ by prolonging the i-th run by one, and so $\mathsf{Emb}(\boldsymbol{x}_i;\boldsymbol{y}) = r_i+1$. Hence, it follows that

$$\arg\max_{0\leqslant i\leqslant n_r}\{\mathsf{Emb}(\boldsymbol{x}_i;\boldsymbol{y})\} = \arg\max_{0\leqslant i\leqslant n_r}\{r_i+1\},$$

where by definition $r_0 = 0$.

■

According to Lemma 3, we can arbitrarily choose the decoder that prolongs the first run of maximal length as the embedding number decoder.

Definition 4. *The* embedding number decoder $\mathcal{D}_{\text{EN}}$ *prolongs the first run of maximal length in* $\boldsymbol{y}$ *by one. A decoder* $\mathcal{D}$ *that prolongs one of the runs of maximal length in* $\boldsymbol{y}$ *by one is said to be* equivalent *to the embedding number decoder, and is denoted by* $\mathcal{D}\equiv\mathcal{D}_{\text{EN}}$.

The following lemmas will be stated for the embedding number decoder for the simplicity of the proofs, but unless stated otherwise they hold for any decoder $\mathcal{D}$ for which $\mathcal{D} \equiv \mathcal{D}_{\text{EN}}$.

Lemma 5. *For every codeword $\boldsymbol{c} \in \mathcal{C}$, the embedding number decoder satisfies*

$$P_{\text{err}}(\boldsymbol{c}, d_L) = \frac{2}{n} \cdot \sum_{\boldsymbol{y} \in D_1(\boldsymbol{c})} \frac{\mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y})}{n} \cdot \mathbb{I}\{\mathcal{D}_{\text{EN}}(\boldsymbol{y}) \neq \boldsymbol{c}\}.$$

Proof: Let $\boldsymbol{c} \in \mathcal{C}$ be a codeword and let $\boldsymbol{y} \in D_1(\boldsymbol{c})$ be a channel output such that $\mathcal{D}_{\text{EN}}(\boldsymbol{y}) \neq \boldsymbol{c}$. Since $\mathcal{D}_{\text{EN}}(\boldsymbol{y})$ can be obtained from a word in $D_1(\boldsymbol{c})$ by one insertion, it follows that $d_L(\mathcal{D}_{\text{EN}}(\boldsymbol{y}), \boldsymbol{c}) = 2$. Thus,

$$\begin{aligned} P_{\text{err}}(\boldsymbol{c}, d_L) &= \sum_{\boldsymbol{y}: \mathcal{D}_{\text{EN}}(\boldsymbol{y}) \neq \boldsymbol{c}} \frac{d_L\left(\mathcal{D}_{\text{EN}}(\boldsymbol{y}), \boldsymbol{c}\right)}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\ &= \frac{2}{n} \sum_{\boldsymbol{y} \in D_1(\boldsymbol{c})} p(\boldsymbol{y}|\boldsymbol{c}) \cdot \mathbb{I}\{\mathcal{D}_{\text{EN}}(\boldsymbol{y}) \neq \boldsymbol{c}\} \\ &= \frac{2}{n} \cdot \sum_{\boldsymbol{y} \in D_1(\boldsymbol{c})} \frac{\mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y})}{n} \cdot \mathbb{I}\{\mathcal{D}_{\text{EN}}(\boldsymbol{y}) \neq \boldsymbol{c}\}. \end{aligned}$$

■

For $\boldsymbol{y} \in D_1(\boldsymbol{c})$, we get $\mathcal{D}_{\text{EN}}(\boldsymbol{y}) = \boldsymbol{c}$ if and only if the deletion occurred within the run corresponding to the first run of maximal length in $\boldsymbol{y}$. Hence, the embedding number decoder will fail at least for any deletion occurring outside of the first run of maximal length in $\boldsymbol{c}$. This observation will be used in the proof of the Lemma 6. Before presenting this proof, one more definition is introduced. For a word $\boldsymbol{x} \in \Sigma_2^n$, we denote by $\tau(\boldsymbol{x})$ the length of its maximal run. For example $\tau(00111010) = 3$ and $\tau(01010101) = 1$. For a code $\mathcal{C} \subseteq \Sigma_2^n$, we denote by $\tau(\mathcal{C})$ the average length of the maximal runs of its codewords. That is,

$$\tau(\mathcal{C}) = \frac{\sum_{\boldsymbol{c} \in \mathcal{C}} \tau(\boldsymbol{c})}{|\mathcal{C}|}.$$

Furthermore, if $N(r)$, for $1 \leqslant r \leqslant n$ denotes the number of codewords in $\mathcal{C}$ in which the length of their maximal run is r, then $\tau(\mathcal{C}) = \frac{\sum_{r=1}^n r \cdot N(r)}{|\mathcal{C}|}$. We are now ready to present a lower bound on the average decoding error probability of the embedding number decoder.

Lemma 6. *The average decoding error probability of the embedding number decoder $\mathcal{D}_{\text{EN}}$ satisfies*

$$P_{\text{err}}(\textsf{1-Del}, \mathcal{C}, \mathcal{D}_{\text{EN}}, d_L) \geqslant \frac{2}{n} \cdot \left(1 - \frac{\tau(\mathcal{C})}{n}\right).$$

Proof: Let $\mathcal{C}_r \subseteq \mathcal{C}$ be the subset of codewords with maximal run length of r, and let its size be denoted by $N(r)$. For any codeword $\boldsymbol{c}$, any deletion outside of the first run of maximal length will result in a decoding failure. Since the sum

$$\sum_{\boldsymbol{y} \in D_1(\boldsymbol{c})} \frac{\mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y})}{n} \cdot \mathbb{I}\{\mathcal{D}_{\text{EN}}(\boldsymbol{y}) \neq \boldsymbol{c}\}$$

is equivalent to counting the indices in $\boldsymbol{c}$ in which a deletion will result in a decoding failure, using Lemma 5 we get that for every $\boldsymbol{c} \in \mathcal{C}_r$,

$$P_{\text{err}}(\boldsymbol{c}, d_L) \geqslant \frac{2}{n} \cdot \frac{n-r}{n},$$

and the average decoding error probability becomes

$$\begin{aligned} &P_{\text{err}}(\textsf{1-Del}, \mathcal{C}, \mathcal{D}_{\text{EN}}, d_L) = \frac{1}{|\mathcal{C}|} \sum_{\boldsymbol{c} \in \mathcal{C}} P_{\text{err}}(\boldsymbol{c}, d_L) \\ &= \frac{1}{|\mathcal{C}|} \sum_{r=1}^{n} \sum_{\boldsymbol{c} \in \mathcal{C}_r} P_{\text{err}}(\boldsymbol{c}, d_L) \geqslant \frac{1}{|\mathcal{C}|} \sum_{r=1}^{n} \sum_{\boldsymbol{c} \in \mathcal{C}_r} \frac{2}{n} \cdot \frac{n-r}{n} \\ &= \frac{1}{|\mathcal{C}|} \frac{2}{n} \sum_{r=1}^{n} N(r) \left(1 - \frac{r}{n}\right) = \frac{2}{n} \left(1 - \frac{1}{n} \frac{\sum_{r=1}^n r \cdot N(r)}{|\mathcal{C}|}\right) \\ &= \frac{2}{n} \cdot \left(1 - \frac{\tau(\mathcal{C})}{n}\right). \end{aligned}$$

■

For the special case of $\mathcal{C} = (\Sigma_2)^n$, the next claim is proved in Appendix A.

Claim 7. *For all $n \geqslant 1$ it holds that $\tau((\Sigma_2)^n) \leqslant 2\log_2(n)$.*

The rest of this section will focus on the case for which $\mathcal{C} = (\Sigma_2)^n$. We will now show that the embedding number decoder is preferable over any decoder that outputs a word of the original codeword length.

Lemma 8. *Let $\mathcal{D} : (\Sigma_2)^{n-1} \to (\Sigma_2)^n$ be a general decoder that prolongs the input length by one. It follows that*

$$P_{\text{err}}(\textsf{1-Del}, \mathcal{C}, \mathcal{D}, d_L) \geqslant P_{\text{err}}(\textsf{1-Del}, \mathcal{C}, \mathcal{D}_{\text{EN}}, d_L). \quad (1)$$

and equality is obtained if and only if $\mathcal{D} \equiv \mathcal{D}_{\text{EN}}$.

Proof: We have the following sequence of equalities and inequalities

$$
\begin{aligned}
&P_{\text{err}}(\text{1-Del}, \mathcal{C}, \mathcal{D}, d_L) = \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}}\sum_{y:\mathcal{D}(y)\neq c}\frac{d_L(\mathcal{D}(y),c)}{|c|}p(y|c)\\
&\overset{(a)}{=} \frac{1}{|\mathcal{C}|}\sum_{y\in(\Sigma_2)^{n-1}}\sum_{c\in I_1(y)}\frac{d_L(\mathcal{D}(y),c)}{|c|}p(y|c)\\
&\overset{(b)}{\geqslant} \frac{1}{|\mathcal{C}|}\sum_{y\in(\Sigma_2)^{n-1}}\frac{2}{n}\left(\left(\sum_{c\in I_1(y)}p(y|c)\right)-p(y|\mathcal{D}(y))\right)\\
&= \frac{2}{n|\mathcal{C}|}\sum_{y\in(\Sigma_2)^{n-1}}\sum_{c\in I_1(y)}p(y|c) - \frac{2}{n|\mathcal{C}|}\sum_{y\in(\Sigma_2)^{n-1}}p(y|\mathcal{D}(y))\\
&\overset{(c)}{=} \frac{2}{n|\mathcal{C}|}\sum_{y\in(\Sigma_2)^{n-1}}\sum_{c\in I_1(y)}p(y|c)\\
&\quad - \frac{2}{n^2|\mathcal{C}|}\sum_{y\in(\Sigma_2)^{n-1}}\mathsf{Emb}(\mathcal{D}(y);y)\\
&\overset{(d)}{\geqslant} \frac{2}{n|\mathcal{C}|}\sum_{y\in(\Sigma_2)^{n-1}}\sum_{c\in I_1(y)}p(y|c)\\
&\quad - \frac{2}{n^2|\mathcal{C}|}\sum_{y\in(\Sigma_2)^{n-1}}\max_{c\in\mathcal{C}}\{\mathsf{Emb}(c;y)\}\\
&\overset{(e)}{\geqslant} \frac{2}{n|\mathcal{C}|}\sum_{y\in(\Sigma_2)^{n-1}}\sum_{c\in I_1(y)}p(y|c)\\
&\quad - \frac{2}{n^2|\mathcal{C}|}\sum_{y\in(\Sigma_2)^{n-1}}\mathsf{Emb}(\mathcal{D}_{EN}(y);y)\\
&= P_{\text{err}}(\text{1-Del}, \mathcal{C}, \mathcal{D}_{\text{EN}}, d_L),
\end{aligned}
$$

where (a) is a result of replacing the order of summation, (b) holds since for every c such that $\mathcal{D}(y) \neq c$ we get $d_L(\mathcal{D}(y), c) \geqslant 2$, and for $c^* = \mathcal{D}(y)$ we get $d_L(\mathcal{D}(y), c^*) = 0$, (c) is obtained by the definition of the 1-deletion channel, and in (d) we simply choose the word that maximizes the value of $\mathsf{Emb}(c; y)$, which is the definition of the ML decoder as derived in step (e). From steps (b) and (e) it also follows that equality is obtained if and only if $\mathcal{D} \equiv \mathcal{D}_{\text{EN}}$. ∎

It can now be shown that in this case, the lazy decoder is preferable over the embedding number decoder.

Lemma 9. *For every $n \geqslant 17$ it holds that*

$$P_{\text{err}}(\text{1-Del}, (\Sigma_2)^n, \mathcal{D}_{\text{EN}}, d_L) > P_{\text{err}}(\text{1-Del}, (\Sigma_2)^n, \mathcal{D}_{\text{Lazy}}, d_L).$$

Proof: For $\mathcal{C} = (\Sigma_2)^n$, from Claim 7, we have that $\tau(\mathcal{C}) \leqslant 2\log_2 n$. Since for every $n \geqslant 17$ it follows that $2\log_2(n) < n/2$, using Lemma 6 we get that

$$P_{\text{err}}(\text{1-Del}, (\Sigma_2)^n, \mathcal{D}_{\text{EN}}, d_L) \geqslant \frac{2}{n}\cdot\left(1-\frac{2\log_2(n)}{n}\right) > \frac{1}{n}.$$

∎

For the rest of this paper we assume $n \geqslant 17$. Next, we examine a *hybrid decoder* which returns words of length either $n-1$ or n and it will be shown that the lazy decoder is preferable over any hybrid decoder either.

Lemma 10. *Let $\mathcal{D} : (\Sigma_2)^{n-1} \to (\Sigma_2)^{n-1} \cup (\Sigma_2)^n$ be a general decoder that either preserves the word length or prolongs it by one. Then, it holds that*

$$P_{\text{err}}(\text{1-Del}, (\Sigma_2)^n, \mathcal{D}, d_L) \geqslant P_{\text{err}}(\text{1-Del}, (\Sigma_2)^n, \mathcal{D}_{\text{Lazy}}, d_L).$$

Proof: Let $\mathcal{D}$ be a decoder as defined in the lemma. Similarly to the proof of Lemma 8, by definition,

$$
\begin{aligned}
&P_{\text{err}}(\text{1-Del}, (\Sigma_2)^n, \mathcal{D}, d_L) = \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}}\sum_{y:\mathcal{D}(y)\neq c}\frac{d_L(\mathcal{D}(y),c)}{|c|}p(y|c)\\
&= \frac{1}{|\mathcal{C}|}\sum_{y\in(\Sigma_2)^{n-1}}\sum_{c\in I_1(y)}\frac{d_L(\mathcal{D}(y),c)}{|c|}p(y|c)\\
&\geqslant \frac{2}{n|\mathcal{C}|}\sum_{\substack{y\in(\Sigma_2)^{n-1}\\|\mathcal{D}(y)|=n}}\left(\sum_{c\in I_1(y)}p(y|c)-p(y|\mathcal{D}(y))\right)\\
&\quad + \frac{1}{n|\mathcal{C}|}\sum_{\substack{y\in(\Sigma_2)^{n-1}\\|\mathcal{D}(y)|=n-1}}\sum_{c\in I_1(y)}p(y|c).
\end{aligned}
\tag{2}
$$

We first show that for each $y \in (\Sigma_2)^{n-1}$ such that $|\mathcal{D}(y)| = n$ it holds that

$$2\sum_{c\in I_1(y)}p(y|c) - 2p(y|\mathcal{D}(y)) \geqslant \sum_{c\in I_1(y)}p(y|c). \tag{3}$$

This is proved by verifying that

$$
\begin{aligned}
&2\sum_{c\in I_1(y)}p(y|c) - 2p(y|\mathcal{D}(y)) - \sum_{c\in I_1(y)}p(y|c)\\
&= \sum_{c\in I_1(y)}p(y|c) - 2p(y|\mathcal{D}(y))\\
&\overset{(a)}{=} \sum_{\substack{c\in I_1(y)\\c\neq\mathcal{D}(y)}}p(y|c) + p(y|\mathcal{D}(y)) - 2p(y|\mathcal{D}(y))\\
&\overset{(b)}{\geqslant} \sum_{\substack{c\in I_1(y)\\c\neq\mathcal{D}(y)}}\frac{1}{n} - p(y|\mathcal{D}(y))\\
&\overset{(c)}{\geqslant} 1 - p(y|\mathcal{D}(y) \geqslant 0,
\end{aligned}
$$

where in (a) we split the summation of $c \in I_1(y)$ into two parts when $\mathcal{D}(y) \in I_1(y)$ and note that this equality holds also when $\mathcal{D}(y) \notin I_1(y)$. In (b) we used the inequality $p(y|c) \geqslant 1/n$ when $c \in I_1(y)$ and lastly in (c) the size of the set $I_1(y) \setminus \{\mathcal{D}(y)\}$ is at least n since $I_1(y)| = n+1$.

Lastly, combining (2) and (3) and remembering that

$d_L(\boldsymbol{c}, \mathcal{D}_{\text{Lazy}}(\boldsymbol{y})) = 1$ we have that

$$\begin{aligned}
&P_{\text{err}}(\text{1-Del}, (\Sigma_2)^n, \mathcal{D}, d_L) \\
&\geqslant \frac{1}{n|\mathcal{C}|}\left(\sum_{\substack{\boldsymbol{y}\in(\Sigma_2)^{n-1}\\ |\mathcal{D}(\boldsymbol{y})|=n}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} p(\boldsymbol{y}|\boldsymbol{c}) + \sum_{\substack{\boldsymbol{y}\in(\Sigma_2)^{n-1}\\ |\mathcal{D}(\boldsymbol{y})|=n-1}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} p(\boldsymbol{y}|\boldsymbol{c})\right) \\
&= \frac{1}{n|\mathcal{C}|} \sum_{\boldsymbol{y}\in(\Sigma_2)^{n-1}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} p(\boldsymbol{y}|\boldsymbol{c}) \\
&= \frac{1}{|\mathcal{C}|} \sum_{\boldsymbol{y}\in(\Sigma_2)^{n-1}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} \frac{d_L(\boldsymbol{c}, \mathcal{D}_{\text{Lazy}}(\boldsymbol{y}))}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\
&= \frac{1}{|\mathcal{C}|} \sum_{\boldsymbol{c}\in\mathcal{C}} \sum_{\boldsymbol{y}:\mathcal{D}_{\text{Lazy}}(\boldsymbol{y})\neq\boldsymbol{c}} \frac{d_L(\boldsymbol{c}, \mathcal{D}_{\text{Lazy}}(\boldsymbol{y}))}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\
&= P_{\text{err}}(\text{1-Del}, (\Sigma_2)^n, \mathcal{D}_{\text{Lazy}}, d_L).
\end{aligned}$$

■

Finally, it is shown that the lazy decoder is preferable over any other type of decoder that returns words of any length.

Theorem 11. *For any decoder* $\mathcal{D} : \Sigma_2^{n-1} \to \Sigma_2^*$,

$$P_{\text{err}}(\text{1-Del}, \Sigma_2^n, \mathcal{D}, d_L) \geqslant P_{\text{err}}(\text{1-Del}, \Sigma_2^n, \mathcal{D}_{\text{Lazy}}, d_L).$$

Proof: Let $\mathcal{D}$ be a decoder as defined in the theorem. By Lemma 10, the theorem holds for any hybrid decoder and therefore we can assume that $\mathcal{D}$ is not a hybrid decoder. Hence, there exists at least one channel output $\boldsymbol{y}'$, such that, $\mathcal{D}(\boldsymbol{y}')$ is neither of length n, nor of length $n-1$. We consider the following two cases.

Case 1: $|\mathcal{D}(\boldsymbol{y}')| \notin \{n-1, n, n+1\}$, and thus $d_L(\mathcal{D}(\boldsymbol{y}'), \boldsymbol{c}) \geqslant 2$. As was done in the proof of Lemma 8, by definition we have the following

$$\begin{aligned}
&P_{\text{err}}(\text{1-Del}, (\Sigma_2)^n, \mathcal{D}, d_L) \\
&= \frac{1}{|\mathcal{C}|} \sum_{\boldsymbol{y}\in(\Sigma_2)^{n-1}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} \frac{d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c})}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\
&\overset{(a)}{\geqslant} \frac{1}{n|\mathcal{C}|} \sum_{\substack{\boldsymbol{y}\in(\Sigma_2)^{n-1}\\ |\mathcal{D}(\boldsymbol{y})|=n\pm1}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} p(\boldsymbol{y}|\boldsymbol{c}) \\
&+ \frac{1}{|\mathcal{C}|} \sum_{\substack{\boldsymbol{y}\in(\Sigma_2)^{n-1}\\ |\mathcal{D}(\boldsymbol{y})|=n}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} \frac{d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c})}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\
&+ \frac{1}{|\mathcal{C}|} \sum_{\substack{\boldsymbol{y}\in(\Sigma_2)^{n-1}\\ |\mathcal{D}(\boldsymbol{y})|\notin\{n,n\pm1\}}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} \frac{d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c})}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\
&\overset{(b)}{\geqslant} \frac{1}{n|\mathcal{C}|} \sum_{\substack{\boldsymbol{y}\in(\Sigma_2)^{n-1}\\ |\mathcal{D}(\boldsymbol{y})|\in\{n,n\pm1\}}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} p(\boldsymbol{y}|\boldsymbol{c}) \\
&+ \frac{1}{|\mathcal{C}|} \sum_{\substack{\boldsymbol{y}\in(\Sigma_2)^{n-1}\\ |\mathcal{D}(\boldsymbol{y})|\notin\{n,n\pm1\}}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} \frac{d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c})}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\
&\overset{(c)}{>} \frac{1}{n|\mathcal{C}|} \sum_{\boldsymbol{y}\in(\Sigma_2)^{n-1}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} p(\boldsymbol{y}|\boldsymbol{c}) = \frac{1}{n},
\end{aligned}$$

where (a) follows from the face that if $|\mathcal{D}(\boldsymbol{y})| = n \pm 1$, then $d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c}) \geqslant 1$ for each $\boldsymbol{c} \in I_1(\boldsymbol{y})$, (b) is obtained using the inequalities in (2) and (3), and (c) is obtained from the fact that whenever $|\mathcal{D}(\boldsymbol{y})| \notin \{n, n\pm 1\}$, we have that $d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c}) \geqslant 2$ for each $\boldsymbol{c} \in I_1(\boldsymbol{y})$ and this summation is not empty. The last equality results from the summing over all probabilities.

Case 2: $|\mathcal{D}(\boldsymbol{y}')| = n+1$. If $\mathcal{D}(\boldsymbol{y}')$ is not the alternating word, then $|D_1(\mathcal{D}(\boldsymbol{y}'))| \leqslant n$, i.e., there are at most n words of length n of distance 1 from $\mathcal{D}(\boldsymbol{y}')$. Since $|I_1(\boldsymbol{y}')| = n+1$, there is at least one word $\boldsymbol{c} \in I_1(\boldsymbol{y}')$ such that $d_L(\mathcal{D}(\boldsymbol{y}'), \boldsymbol{c}) > 1$. Using this observation and as was done in the first case of this proof we derive that

$$\begin{aligned}
&P_{\text{err}}(\text{1-Del}, (\Sigma_2)^n, \mathcal{D}, d_L) \\
&= \frac{1}{|\mathcal{C}|} \sum_{\boldsymbol{y}\in(\Sigma_2)^{n-1}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} \frac{d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c})}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\
&\geqslant \frac{1}{n|\mathcal{C}|} \sum_{\substack{\boldsymbol{y}\in(\Sigma_2)^{n-1}\\ |\mathcal{D}(\boldsymbol{y})|=n-1}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} p(\boldsymbol{y}|\boldsymbol{c}) \\
&+ \frac{1}{|\mathcal{C}|} \sum_{\substack{\boldsymbol{y}\in(\Sigma_2)^{n-1}\\ |\mathcal{D}(\boldsymbol{y})|=n}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} \frac{d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c})}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\
&+ \frac{1}{|\mathcal{C}|} \sum_{\substack{\boldsymbol{y}\in(\Sigma_2)^{n-1}\\ |\mathcal{D}(\boldsymbol{y})|=n+1}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} \frac{d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c})}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\
&> \frac{1}{n|\mathcal{C}|} \sum_{\boldsymbol{y}\in(\Sigma_2)^{n-1}} \sum_{\boldsymbol{c}\in I_1(\boldsymbol{y})} p(\boldsymbol{y}|\boldsymbol{c}) = \frac{1}{n},
\end{aligned}$$

where the last inequality results from the words $\boldsymbol{y}', \boldsymbol{c}$ which satisfy $d_L(\mathcal{D}(\boldsymbol{y}'), \boldsymbol{c}) > 1$ That is, it is concluded that

$$P_{\text{err}}(\text{1-Del}, (\Sigma_2)^n, \mathcal{D}, d_L) > \frac{1}{n} = P_{\text{err}}(\text{1-Del}, (\Sigma_2)^n, \mathcal{D}_{\text{Lazy}}, d_L).$$

Note that, for the special case where $\mathcal{D}(\boldsymbol{y}')$ is the alternating sequence of length $n+1$, $|I_1(\boldsymbol{y}')| = |D_1(\mathcal{D}(\boldsymbol{y}'))| = n+1$, which implies that inequality (a) is a weak inequality. ■

Theorem 11 verifies that $\mathcal{D}_{Lazy}$ minimizes the average decoding error probability for the case when $\mathcal{C} = \Sigma_2^n$, which implies that $\mathcal{D}_{Lazy}$ is the ML* decoder for the 1-deletion channel.

IV. The 2-Deletion Channel

In this section we consider the case of a single 2-deletion channel over a code which is the entire space, i.e., $\mathcal{C} = \Sigma_2^n$. In this setup, a word $\boldsymbol{x} \in \Sigma_2^n$ is transmitted over the channel 2-Del, where exactly 2 symbols from $\boldsymbol{x}$ are selected and deleted, resulting in the channel output $\boldsymbol{y} \in \Sigma_2^{n-2}$. We construct a decoder that is based on the lazy decoder and on a variant of the embedding number decoder and prove that it minimizes the average decoding error probability, that is, we explicitly find the ML* decoder for the 2-Del channel.

Recall that the average decoding error probability of a decoder $\mathcal{D}$ over a single 2-deletion channel is defined as

$$\begin{aligned}\mathsf{P}_{\text{err}}(\mathcal{D}) &= \frac{1}{|\mathcal{C}|}\sum_{\boldsymbol{c}\in\mathcal{C}}\sum_{\boldsymbol{y}\in\Sigma_2^{n-2}} P_{\text{err}}(\boldsymbol{c}) \\ &= \frac{1}{|\mathcal{C}|\cdot|\boldsymbol{c}|}\sum_{\boldsymbol{c}\in\mathcal{C}}\sum_{\boldsymbol{y}:\mathcal{D}(\boldsymbol{y})\neq \boldsymbol{c}} d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c})\cdot p(\boldsymbol{y}|\boldsymbol{c}).\end{aligned}$$

We can rearrange the sum as follows

$$\mathsf{P}_{\text{err}}(\mathcal{D}) = \frac{1}{|\mathcal{C}|\cdot|\boldsymbol{c}|}\sum_{\boldsymbol{y}\in\Sigma_2^{n-2}}\sum_{\boldsymbol{c}\in\mathcal{C}} d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c})\cdot p(\boldsymbol{y}|\boldsymbol{c}).$$

As mentioned before we denote $\sum_{\boldsymbol{c}:\mathcal{D}(\boldsymbol{y})\neq \boldsymbol{c}} \frac{d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c})}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c})$ by $f_{\boldsymbol{y}}(\mathcal{D}(\boldsymbol{y}))$. Recall that, a decoder that minimizes $f_{\boldsymbol{y}}(\mathcal{D}(\boldsymbol{y}))$ for any channel output $\boldsymbol{y} \in \Sigma_2^{n-2}$, also minimizes the average decoding error probability. Hence, when comparing two decoders, it is enough to compare $f_{\boldsymbol{y}}(\mathcal{D}(\boldsymbol{y}))$ for each channel output $\boldsymbol{y}$.

Before we continue, two more families of decoders are introduced. The *maximum likelihood* of length m*, denoted by $\mathcal{D}^m_{\mathsf{ML}^*}$, is the decoder that for any given channel output $\boldsymbol{y}$ returns a word $\boldsymbol{x}$ of length m that minimizes $f_{\boldsymbol{y}}(\boldsymbol{x})$. That is,

$$\mathcal{D}^m_{\mathsf{ML}^*}(\boldsymbol{y}) = \underset{\boldsymbol{x}\in\Sigma_2^m}{\arg\min}\{f_{\boldsymbol{y}}(\boldsymbol{x})\}.$$

The *embedding number decoder of length m*, denoted by $\mathcal{D}^m_{EN}$, is the decoder that for any given channel output $\boldsymbol{y}$ returns a word $\boldsymbol{x}$ of length m that maximizes the embedding number of $\boldsymbol{y}$ in $\boldsymbol{x}$. That is,

$$\mathcal{D}^m_{EN}(\boldsymbol{y}) = \underset{\boldsymbol{x}\in\Sigma_2^m}{\arg\max}\{\mathsf{Emb}(\boldsymbol{x};\boldsymbol{y})\}.$$

In these decoders' definitions, and unless stated otherwise, if there is more than one word $\boldsymbol{x}$ that optimizes these expressions, the decoder chooses one of them arbitrarily.

Similarly to the analysis of the 1-Del channel in Section III, any embedding number decoder prolongs existing runs in the word $\boldsymbol{y}$. The following lemma proves that any embedding number decoder of length $m > |\boldsymbol{y}|$ prolongs at least one of the longest runs in $\boldsymbol{y}$ by at least one symbol.

Lemma 12. *Let $\boldsymbol{y} \in \Sigma_2^{n-2}$ be a channel output. For any $m > |\boldsymbol{y}|$, the decoder $\mathcal{D}^m_{EN}$ prolongs one of the longest runs of $\boldsymbol{y}$ by at least one symbol.*

Proof: Assume that the number of runs in $\boldsymbol{y}$ is $\rho(\boldsymbol{y}) = t$ and let r_j denote the length of the j-th run for $1 \leqslant j \leqslant t$. Assume to the contrary that none of the longest runs in $\boldsymbol{y}$ was prolonged and let i' be one of the indices of the longest runs in $\boldsymbol{y}$ that was prolonged by the decoder $\mathcal{D}^m_{EN}$ and note that $r_i > r_{i'}$. However, if the decoder will instead prolong the i-th run in $\boldsymbol{y}$ with the same number of symbols as the i'-th run was prolonged, we will get a word that strictly increases the embedding number, in contradiction. ■

For simplicity, we assume that in the case where there are two or more longest runs in $\boldsymbol{y}$, the embedding number decoder $\mathcal{D}^m_{EN}$ for $m > |\boldsymbol{y}|$ necessarily chooses to prolong the first ones. Moreover, if there is more than one option that maximize the embedding number, the embedding number decoder $\mathcal{D}^m_{EN}$ will choose the one that prolongs the least number of runs.

In the following lemma a useful property about $\mathcal{D}^n_{EN}$, the embedding number decoder of length n, is given.

Lemma 13. *Let $\boldsymbol{y} \in \Sigma_2^{n-2}$ be a channel output. Assume that the number of runs in $\boldsymbol{y}$ is $\rho(\boldsymbol{y}) = k$ and let r_i denote the length of the i-th run for $1 \leqslant i \leqslant t$. In addition, let the i-th and the j-th runs be the two longest runs in $\boldsymbol{y}$, such that $r_i \geqslant r_j$. The decoder $\mathcal{D}^n_{EN}$ operates as follows.*

1) *If $r_i \geqslant 2r_j$, the decoder prolongs the i-th run by two symbols.*
2) *If $r_i < 2r_j$, the decoder prolongs the i-th and the j-th runs, each by one symbol.*

Proof: The embedding number decoder in this case has two options. The first one is to prolong one of the runs in $\boldsymbol{y}$ by two symbols and the second is to prolong two runs in $\boldsymbol{y}$ each by one symbol. We ignore the option of creating new runs since it won't increase the embedding number. Thus, the maximum embedding number value is given by

$$\begin{aligned}&\max\left\{\max_{1\leqslant s<\ell\leqslant k}\left\{\binom{r_s+1}{1}\cdot\binom{r_\ell+1}{1}\right\}, \max_{1\leqslant s\leqslant k}\binom{r_s+2}{2}\right\} \\ &= \max\left\{\max_{1\leqslant s<\ell\leqslant k}\{(r_s+1)(r_\ell+1)\}, \max_{1\leqslant s\leqslant k}\left\{\frac{(r_s+1)(r_s+2)}{2}\right\}\right\} \\ &= \max\left\{(r_i+1)(r_j+1), \frac{(r_i+1)(r_i+2)}{2}\right\}.\end{aligned}$$

Finally, in order to determine the option which maximizes the embedding number, it is left to compare between $(r_i+1)(r_j+1)$ and $\frac{(r_i+1)(r_i+2)}{2}$. Thus, the decoder $\mathcal{D}^n_{EN}$ choses the first option, i.e., prolonging the longest run in two symbols, if and only if $\frac{r_i+2}{2} \geqslant (r_j+1)$ which is equivalent to $r_i \geqslant 2r_j$. ■

In the rest of this section we prove several properties on $\mathcal{D}_{\mathsf{ML}^*}$, the ML* decoder for a single 2-deletion channel and lastly in Theorem 25 we construct this decoder explicitly. Unless specified otherwise, we assume that $\mathcal{D}_{\mathsf{ML}^*}$ returns a word with minimum length that minimizes $f_{\boldsymbol{y}}(\mathcal{D}(\boldsymbol{y}))$.

Lemma 14. *For a channel output $\boldsymbol{y} \in \Sigma_2^{n-2}$, it holds that*

$$n-2 \leqslant |\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})| \leqslant n+1.$$

Proof: Let $\boldsymbol{y} \in \Sigma_2^{n-2}$ be a channel output and assume to the contrary that $|\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})| \geqslant n+2$ or $|\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})| \leqslant n-3$. In order to show a contradiction, we prove that

$$\sum_{\boldsymbol{c} \in I_2(\boldsymbol{y})} \frac{d_L(\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y}), \boldsymbol{c})}{|\boldsymbol{c}|} \cdot p(\boldsymbol{y}|\boldsymbol{c}) \geqslant \sum_{\boldsymbol{c} \in I_2(\boldsymbol{y})} \frac{d_L(\mathcal{D}_{Lazy}(\boldsymbol{y}), \boldsymbol{c})}{|\boldsymbol{c}|} \cdot p(\boldsymbol{y}|\boldsymbol{c}),$$

and equality can be obtained only in the case $|\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})| = n+2$. If $|\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})| \leqslant n-3$ or $|\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})| \geqslant n+3$, then $d_L(\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y}), \boldsymbol{c}) \geqslant 3$ and since $d_L(\mathcal{D}_{Lazy}(\boldsymbol{y}), \boldsymbol{c}) = 2$ a strict inequality holds for each $\boldsymbol{y}$. In case $|\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})| = n+2$, $d_L(\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y}), \boldsymbol{c}) \geqslant 2$ and the inequality holds. Recall that $\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})$ returns a word with minimum length which implies that $|\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})| \leqslant n+1$. ∎

For $\boldsymbol{y} \in \Sigma_2^{n-2}$, Lemma 14 implies that $m = |\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})| \in \{n-2, n-1, n, n+1\}$. In the following lemmas, we show that for any $m \in \{n-2, n-1, n\}$,

$$\mathcal{D}_{\mathsf{ML}^*}^m = \mathcal{D}_{EN}^m.$$

Lemma 15. *It holds that*

$$\mathcal{D}_{\mathsf{ML}^*}^m = \mathcal{D}_{EN}^{n-2} = \mathcal{D}_{Lazy}.$$

Proof: Let $\boldsymbol{y} \in \Sigma_2^{n-2}$ be a channel output. Each $\boldsymbol{y}' \in \Sigma_2^{n-2}$ such that $\boldsymbol{y}' \neq \boldsymbol{y}$ satisfies $\mathsf{Emb}(\boldsymbol{y}'; \boldsymbol{y}) = 0$. Hence $\mathcal{D}_{EN}^{n-2}(\boldsymbol{y}) = \boldsymbol{y}$, which implies that $\mathcal{D}_{EN}^{n-2} = \mathcal{D}_{Lazy}$.

In order to show that $\mathcal{D}_{Lazy} = \mathcal{D}_{\mathsf{ML}^*}^{n-2}$, let us consider any decoder $\mathcal{D}$ that outputs words of length $n-2$ such that $\mathcal{D} \neq \mathcal{D}_{\text{Lazy}}$, i.e., there exists $\boldsymbol{y} \in \Sigma_2^{n-2}$ such that $\mathcal{D}(\boldsymbol{y}) = \boldsymbol{y}' \neq \boldsymbol{y}$. Since $\boldsymbol{y}' \neq \boldsymbol{y}$ it holds that $I_2(\boldsymbol{y}') \neq I_2(\boldsymbol{y})$ and hence, there exists a codeword $\boldsymbol{c} \in \Sigma_2^n$ such that $\boldsymbol{c} \in I_2(\boldsymbol{y})$ and $\boldsymbol{c} \notin I_2(\boldsymbol{y}')$. Equivalently, $\boldsymbol{y} \in D_2(\boldsymbol{c})$, $\boldsymbol{y}' \notin D_2(\boldsymbol{c})$ and therefore $d_L(\boldsymbol{c}, \boldsymbol{y}') \geqslant 4$ (at least one more deletion and one more insertion are needed in addition to the two insertions needed for every word in the deletion ball). Hence,

$$\begin{aligned}
f_{\boldsymbol{y}}(\mathcal{D}(\boldsymbol{y})) &= \sum_{\boldsymbol{c}' \in \Sigma_2^n} \frac{d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c}')}{|\boldsymbol{c}'|} p(\boldsymbol{y}|\boldsymbol{c}') \\
&= \sum_{\substack{\boldsymbol{c}' \in \Sigma_2^n \\ \boldsymbol{c}' \neq \boldsymbol{c}}} \frac{d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c}')}{|\boldsymbol{c}'|} p(\boldsymbol{y}|\boldsymbol{c}') + \frac{d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c})}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\
&\geqslant \sum_{\substack{\boldsymbol{c}' \in \Sigma_2^n \\ \boldsymbol{c}' \neq \boldsymbol{c}}} \frac{2}{|\boldsymbol{c}'|} p(\boldsymbol{y}|\boldsymbol{c}') + \frac{d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c})}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\
&\geqslant \sum_{\substack{\boldsymbol{c}' \in \Sigma_2^n \\ \boldsymbol{c}' \neq \boldsymbol{c}}} \frac{2}{|\boldsymbol{c}'|} p(\boldsymbol{y}|\boldsymbol{c}') + \frac{4}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\
&> \sum_{\boldsymbol{c}' \in \Sigma_2^n} \frac{2}{|\boldsymbol{c}'|} p(\boldsymbol{y}|\boldsymbol{c}') = f_{\boldsymbol{y}}(\mathcal{D}_{\text{Lazy}}(\boldsymbol{y})) = f_{\boldsymbol{y}}(\mathcal{D}_{EN}^{n-2}(\boldsymbol{y})).
\end{aligned}$$

These inequalities state that $\mathcal{D}_{Lazy}$ is the decoder that minimizes $f_{\boldsymbol{y}}(\mathcal{D}(\boldsymbol{y}))$ for any $\boldsymbol{y} \in \Sigma_2^{n-2}$ among all decoders that return words of length $n-2$. Hence, we deduce that the ML* decoder of length $n-2$ is $\mathcal{D}_{Lazy}$. ∎

For the rest of this section we use the following observation, given two decoder $\mathcal{D}_1$ and $\mathcal{D}_2$,

$$\begin{aligned}
&f_{\boldsymbol{y}}(\mathcal{D}_1(\boldsymbol{y})) - f_{\boldsymbol{y}}(\mathcal{D}_2(\boldsymbol{y})) \\
&= \sum_{\boldsymbol{c}: \mathcal{D}_1(\boldsymbol{y}) \neq \boldsymbol{c}} \frac{d_L(\mathcal{D}_1(\boldsymbol{y}), \boldsymbol{c})}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) - \sum_{\boldsymbol{c}: \mathcal{D}_2(\boldsymbol{y}) \neq \boldsymbol{c}} \frac{d_L(\mathcal{D}_2(\boldsymbol{y}), \boldsymbol{c})}{|\boldsymbol{c}|} p(\boldsymbol{y}|\boldsymbol{c}) \\
&= \frac{1}{|\boldsymbol{c}|} \left(\sum_{\boldsymbol{c} \in \Sigma_2^n} d_L(\mathcal{D}_1(\boldsymbol{y}), \boldsymbol{c}) p(\boldsymbol{y}|\boldsymbol{c}) - \sum_{\boldsymbol{c} \in \Sigma_2^n} d_L(\mathcal{D}_2(\boldsymbol{y}), \boldsymbol{c}) p(\boldsymbol{y}|\boldsymbol{c}) \right) \\
&= \frac{1}{|\boldsymbol{c}|} \sum_{\boldsymbol{c} \in \Sigma_2^n} p(\boldsymbol{y}|\boldsymbol{c}) \Big(d_L(\mathcal{D}_1(\boldsymbol{y}), \boldsymbol{c}) - d_L(\mathcal{D}_2(\boldsymbol{y}), \boldsymbol{c}) \Big) \\
&= \frac{1}{\binom{n}{2} |\boldsymbol{c}|} \sum_{\boldsymbol{c} \in \Sigma_2^n} \mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y}) \Big(d_L(\mathcal{D}_1(\boldsymbol{y}), \boldsymbol{c}) - d_L(\mathcal{D}_2(\boldsymbol{y}), \boldsymbol{c}) \Big) \\
&= \frac{1}{\binom{n}{2} |\boldsymbol{c}|} \sum_{\boldsymbol{c} \in I_2(\boldsymbol{y})} \mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y}) \Big(d_L(\mathcal{D}_1(\boldsymbol{y}), \boldsymbol{c}) - d_L(\mathcal{D}_2(\boldsymbol{y}), \boldsymbol{c}) \Big),
\end{aligned}$$

where the last equality holds since for any $\boldsymbol{c} \in \Sigma_2^n$ such that $\boldsymbol{c} \notin I_2(\boldsymbol{y})$ it holds that $\mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y}) = 0$. Hence when comparing the average decoding error probability of two decoders $\mathcal{D}_1$ and $\mathcal{D}_2$, it holds that,

$$f_{\boldsymbol{y}}(\mathcal{D}_1(\boldsymbol{y})) \geqslant f_{\boldsymbol{y}}(\mathcal{D}_2(\boldsymbol{y}))$$

if and only if

$$\sum_{\boldsymbol{c} \in I_2(\boldsymbol{y})} \mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y}) \Big(d_L(\mathcal{D}_1(\boldsymbol{y}), \boldsymbol{c}) - d_L(\mathcal{D}_2(\boldsymbol{y}), \boldsymbol{c}) \Big) \geqslant 0. \quad (4)$$

Lemma 16. *It holds that*

$$\mathcal{D}_{\mathsf{ML}^*}^{n-1} = \mathcal{D}_{EN}^{n-1}.$$

Proof: By similar arguments to those presented in Lemma 12, for any channel output $\boldsymbol{y}$, $\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})$ is obtained from $\boldsymbol{y}$ by prolonging the first longest run of $\boldsymbol{y}$ by one symbol. Let $\boldsymbol{y}$ be the channel output and let $\mathcal{D}$ be a decoder such that $|\mathcal{D}(\boldsymbol{y})| = n-1$. Our goal is to prove that the inequality stated in (4) holds for all $\boldsymbol{y}$ when $\mathcal{D}_1 = \mathcal{D}$ and $\mathcal{D}_2 = \mathcal{D}_{EN}^{n-1}$.

This verifies the lemma's statement. This will be verified in the following claims.

Claim 17. *For any decoder $\mathcal{D}$ such that $\mathcal{D}(\boldsymbol{y}) \neq \mathcal{D}_{EN}^{n-1}(\boldsymbol{y})$ and $|\mathcal{D}(\boldsymbol{y})| = n-1$, where $\mathcal{D}(\boldsymbol{y})$ is obtained from $\boldsymbol{y}$ by prolonging one of the runs in $\boldsymbol{y}$, the inequality stated in (4) holds.*

Proof: Assume that the number of runs in $\boldsymbol{y}$ is $\rho(\boldsymbol{y}) = k$, let r_j denote the length of the j-th run for $1 \leqslant j \leqslant k$, and let the i-th run of $\boldsymbol{y}$ be the first longest run of $\boldsymbol{y}$. Assume that $\mathcal{D}(\boldsymbol{y})$ is obtained by prolonging the j-th run of $\boldsymbol{y}$ by one symbol. Since $\mathcal{D}(\boldsymbol{y}) \neq \mathcal{D}_{EN}^{n-1}(\boldsymbol{y})$ it holds that $j \neq i$. Note that

$$|I_1(\mathcal{D}(\boldsymbol{y})) \cap I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))| = 1$$

since the only word $\boldsymbol{c}$ in this set is the word that is obtained by prolonging the i-th and j-th runs of $\boldsymbol{y}$. It holds that $d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c}) = d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}), \boldsymbol{c}) = 1$ and hence this word can be eliminated from inequality (4). Similarly for words $\boldsymbol{c}$ such that $\boldsymbol{c} \notin I_1(\mathcal{D}(\boldsymbol{y}))$ and $\boldsymbol{c} \notin I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))$, we get that $d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c}) = d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}), \boldsymbol{c}) = 3$ and therefore these words can also be eliminated from inequality (4). Note that the number of such words is

$$\begin{aligned}
&|I_2(\boldsymbol{y})| - \left|I_1\left(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})\right)\right| \\
&\quad - |I_1(\mathcal{D}(\boldsymbol{y}))| + \left|I_1(\mathcal{D}(\boldsymbol{y})) \cap I_1\left(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})\right)\right| \\
&\quad = \binom{n}{2} + n + 1 - 2(n+1) + 1 = \binom{n}{2} - n.
\end{aligned}$$

Let us consider the remaining $2n$ words in $I_2(\boldsymbol{y})$.

1) $\boldsymbol{c} \in I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))$ and $\boldsymbol{c} \notin I_1(\mathcal{D}(\boldsymbol{y}))$: Since the embedding number decoder prolongs a run in $\boldsymbol{y}$, $I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})) \subseteq I_2(\boldsymbol{y})$. Therefore, there are

$$|I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))| - |I_1(\mathcal{D}(\boldsymbol{y})) \cap I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))| = n + 1 - 1 = n$$

such words and for each one of them,

$$d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c})) = 3 \text{ and } d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}), \boldsymbol{c}) = 1.$$

We consider three possible options for the word $\boldsymbol{c}$ in this case. If $\boldsymbol{c}$ is the word obtained by prolonging the i-th run of $\boldsymbol{y}$ by two symbols, then $\mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y}) = \binom{r_i+2}{2}$. Let $\boldsymbol{c} = \boldsymbol{c}_h$ be the word obtained by prolonging the i-th and the h-th run for $h \neq i, j$. Since there are $t-2$ runs other than the i-th and the j-th run, the number of such words is $t-2$, while $\mathsf{Emb}(\boldsymbol{c}_h; \boldsymbol{y}) = (r_i+1)(r_h+1)$. Lastly, if $\boldsymbol{c}$ is obtained by prolonging the i-th run and creating a new run in $\boldsymbol{y}$ then $\mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y}) = r_i + 1$, and the number of such words is $n - t + 1$. Thus,

$$\begin{aligned}
&\sum_{\substack{\boldsymbol{c} \in I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})) \\ \boldsymbol{c} \notin I_1(\mathcal{D}(\boldsymbol{y}))}} \mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y}) \left(d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c}) - d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}), \boldsymbol{c})\right) \\
&= 2\left(\binom{r_i+2}{2} + \sum_{\substack{h=1 \\ h \neq j,i}}^{k} (r_h+1)(r_i+1) + (n-k+1)(r_i+1)\right) \\
&= 2\left(\binom{r_i+2}{2} + (r_i+1)(n-2-r_j-r_i+k-2) + (n-k+1)(r_i+1)\right) \\
&= 2\left(\binom{r_i+2}{2} + (r_i+1)(n-r_j-r_i+k-4) + (n-k+1)(r_i+1)\right).
\end{aligned}$$

2) $\boldsymbol{c} \notin I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))$ and $\boldsymbol{c} \in I_1(\mathcal{D}(\boldsymbol{y}))$: The decoder $\mathcal{D}$ prolongs a run in $\boldsymbol{y}$, and therefore $I_1(\mathcal{D}(\boldsymbol{y})) \subseteq I_2(\boldsymbol{y})$. Similarily to Case 1, there are n such words, and

$$\begin{aligned}
&\sum_{\substack{\boldsymbol{c} \notin I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})) \\ \boldsymbol{c} \in I_1(\mathcal{D}(\boldsymbol{y}))}} \mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y}) \left(d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c}) - d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}), \boldsymbol{c})\right) \\
&= 2\left(\binom{r_j+2}{2} + \sum_{\substack{h=1 \\ h \neq j,i}}^{k} (r_h+1)(r_j+1) + (n-k+1)(r_j+1)\right) \\
&= 2\left(\binom{r_j+2}{2} + (r_j+1)(n-2-r_j-r_i+k-2) + (n-k+1)(r_j+1)\right) \\
&= 2\left(\binom{r_j+2}{2} + (r_j+1)(n-r_j-r_i+k-4) + (n-k+1)(r_j+1)\right).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\sum_{\boldsymbol{c} \in I_2(\boldsymbol{y})} \mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y}) \left(d_L(\mathcal{D}_1(\boldsymbol{y}), \boldsymbol{c}) - d_L(\mathcal{D}_2(\boldsymbol{y}), \boldsymbol{c})\right) \\
&= 2\left(\binom{r_i+2}{2} + (r_i+1)(n-r_j-r_i+k-4) + (n-k+1)(r_i+1)\right) \\
&- 2\left(\binom{r_j+2}{2} + (r_j+1)(n-r_j-r_i+k-4) + (n-k+1)(r_j+1)\right) \\
&\geqslant 0,
\end{aligned}$$

where the last inequality holds since $r_i \geqslant r_j$. ■

Claim 18. *For any decoder $\mathcal{D}$ such that $\mathcal{D}(\boldsymbol{y}) \neq \mathcal{D}_{EN}^{n-1}(\boldsymbol{y})$ and $|\mathcal{D}(\boldsymbol{y})| = n-1$, where $\mathcal{D}(\boldsymbol{y})$ is obtained from $\boldsymbol{y}$ by creating a new run of one symbol in $\boldsymbol{y}$, the inequality stated in (4) holds.*

Proof: Assume that the number of runs in $\boldsymbol{y}$ is $\rho(\boldsymbol{y}) = k$, let r_j denote the length of the j-th run for $1 \leqslant j \leqslant k$, and let the i-th run of $\boldsymbol{y}$ be the first longest run of $\boldsymbol{y}$. As in Claim 17, if $\boldsymbol{c} \in \left(I_1(\mathcal{D}(\boldsymbol{y})) \cap I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))\right)$, then $\boldsymbol{c}$ can be eliminated from (4). In addition, any word $\boldsymbol{c}$ such that $\boldsymbol{c} \notin I_1(\mathcal{D}(\boldsymbol{y}))$ and $\boldsymbol{c} \notin I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))$ can be eliminated from (4). Let us consider the remaining $2n$ words in $I_2(\boldsymbol{y})$:

1) $\boldsymbol{c} \in I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))$ and $\boldsymbol{c} \notin I_1(\mathcal{D}(\boldsymbol{y}))$: From arguments similar to those presented in Claim 17, there are n such words and

$$\begin{aligned}
&\sum_{\substack{\boldsymbol{c} \in I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})) \\ \boldsymbol{c} \notin I_1(\mathcal{D}(\boldsymbol{y}))}} \mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y}) \left(d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c}) - d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}), \boldsymbol{c})\right) \\
&= 2\left(\binom{r_i+2}{2} + (r_i+1)(n-r_j-r_i+k-3) + (n-k+1)(r_i+1)\right).
\end{aligned}$$

2) $\boldsymbol{c} \notin I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))$ and $\boldsymbol{c} \in I_1(\mathcal{D}(\boldsymbol{y}))$: As in Claim 17, the number of such words is n, and for each of these words,

$$d_L(\mathcal{D}(\boldsymbol{y}), \boldsymbol{c}) = 3 \text{ and } d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}), \boldsymbol{c}) = 1.$$

We consider three possible options for the word $\boldsymbol{c}$ in this case. If $\boldsymbol{c}$ is the word obtained by prolonging the new run of $\mathcal{D}(\boldsymbol{y})$ by additional symbol then $\mathsf{Emb}(\boldsymbol{c}; \boldsymbol{y}) = 1$. Let $\boldsymbol{c} = \boldsymbol{c}_h$ be the word obtained by prolonging the h-th run of $\boldsymbol{y}$ for $h \neq i$ and creating the same new run of one symbol as in $\mathcal{D}(\boldsymbol{y})$. Since there are $t-1$ runs other than the i-th run, the number of such words is $t-1$, while

$\mathsf{Emb}(\boldsymbol{c}_h;\boldsymbol{y}) = (r_h+1)$. Lastly, if $\boldsymbol{c}$ is obtained by by creating additional new run in $\mathcal{D}(\boldsymbol{y})$ then $\mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}) \leqslant 2$ [1], and the number of such words is $n-t$. Hence,

$$\begin{aligned}
&\sum_{\substack{\boldsymbol{c}\notin I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))\\ \boldsymbol{c}\in I_1(\mathcal{D}(\boldsymbol{y}))}} \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}) \left(d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c}) - d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c})\right)\\
&\geqslant 1\left(1+\sum_{\substack{h=1\\h\neq i}}^{k}(r_h+1)+1(n-k)\right)\\
&\quad -3\left(1+\sum_{\substack{h=1\\h\neq i}}^{k}(r_h+1)+2(n-k)\right)\\
&= -2-2\sum_{\substack{h=1\\h\neq i}}^{t}(r_h+1)-5(n-k).\\
&= -2-2(n-2-r_i+k-1)-5(n-k).\\
&= -2(n-r_i+k-2)-5(n-k).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\sum_{\boldsymbol{c}\in I_2(\boldsymbol{y})} \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y})\Big(d_L(\mathcal{D}_1(\boldsymbol{y}),\boldsymbol{c}) - d_L(\mathcal{D}_2(\boldsymbol{y}),\boldsymbol{c})\Big)\\
&\geqslant 2\Big((r_i+2)+(r_i+1)(n-2-r_i+k-1)+(n-k+1)(r_i+1)\Big)\\
&\quad -2(n-r_i+k-2)-5(n-k)\\
&= 2\Big((r_i+2)+(r_i+1)(n-r_i+k-3)\\
&\quad +(n-k+1)(r_i+1)-(n-r_i+k-2)\Big)-5(n-k)\\
&= 2\Big((r_i+2)+(r_i+1)(2n-r_i-2)-(n-r_i+k-2)\Big)-5(n-k)\\
&\geqslant 0,
\end{aligned}$$

where the last inequality holds for any $1\leqslant r_i, k\leqslant n$. ■

Claim 19. *For any decoder $\mathcal{D}$ such that $\mathcal{D}(\boldsymbol{y}) \neq \mathcal{D}_{EN}^{n-1}(\boldsymbol{y})$ and $|\mathcal{D}(\boldsymbol{y})| = n-1$, where $\mathcal{D}(\boldsymbol{y})$ is not a supersequence of $\boldsymbol{y}$, the inequality stated in (4) holds.*

Proof: By definition $\mathcal{D}(\boldsymbol{y})$ is not a supersequence of $\boldsymbol{y}$ which implies that $\boldsymbol{y}\notin D_1(\mathcal{D}(\boldsymbol{y}))$. Note that for any word $\boldsymbol{c}\in I_2(\boldsymbol{y})$ such that $\boldsymbol{c}\notin I_1(\mathcal{D}(\boldsymbol{y}))$, it holds that $d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c})\geqslant 3$, while $d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c})\leqslant 3$. Hence, if $I_2(\boldsymbol{y})\cap I_1(\mathcal{D}(\boldsymbol{y}))=\emptyset$ then,

$$\begin{aligned}
&\sum_{\boldsymbol{c}\in I_2(\boldsymbol{y})} \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y})\Big(d_L(\mathcal{D}_1(\boldsymbol{y}),\boldsymbol{c}) - d_L(\mathcal{D}_2(\boldsymbol{y}),\boldsymbol{c})\Big)\\
&\qquad \geqslant \sum_{\boldsymbol{c}\in I_2(\boldsymbol{y})} \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y})\Big(3-3\Big) = 0.
\end{aligned}$$

Otherwise, let $\boldsymbol{c}$ be a word such that $\boldsymbol{c}\in\Big(I_2(\boldsymbol{y})\cap I_1(\mathcal{D}(\boldsymbol{y}))\Big)$, let $\rho(\boldsymbol{c})=k'$ be the number of runs in $\boldsymbol{c}$ and denote by r'_j the length of the j-th run in $\boldsymbol{c}$. Let the i-th run in $\boldsymbol{c}$ be the first longest run in $\boldsymbol{c}$. Note that $\boldsymbol{y}\in D_2(\boldsymbol{c})$ and $\mathcal{D}(\boldsymbol{y})\in D_1(\boldsymbol{c})$. Consider the following distinct cases.

[1]This value equals two if and only if $\boldsymbol{c}$ is the alternating sequence.

1) There exists an index $1\leqslant j\leqslant k'$ such that $\boldsymbol{y}$ is obtained from $\boldsymbol{c}$ by deleting two symbols from the j-th run of $\boldsymbol{c}$. In this case, since $\mathcal{D}(\boldsymbol{y})$ is not a supersequence of $\boldsymbol{y}$, $\mathcal{D}(\boldsymbol{y})$ must be obtained from $\boldsymbol{c}$ by deleting one symbol from the h-th run of $\boldsymbol{c}$ for some $h\neq j$. Hence, $\boldsymbol{c}$ is the unique word that is obtained by inserting to $\boldsymbol{y}$ the two symbols that were deleted from the j-th run of $\boldsymbol{c}$, that is,

$$I_2(\boldsymbol{y})\cap I_1(\mathcal{D}(\boldsymbol{y}))=\{\boldsymbol{c}\}.$$

Note that, $\mathsf{Emb}(\boldsymbol{c};\boldsymbol{y})=\binom{r'_j}{2}\leqslant\binom{r'_i}{2}$ and $d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c})=1$, while $d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c})\in\{1,3\}$. If $d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c})=1$, (4) holds (since $\boldsymbol{c}$ is the only word in the intersection). Otherwise $d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c})=3$ and our goal is to find $\boldsymbol{c}'\in I_2(\boldsymbol{y})$ such that

$$\begin{aligned}
&\sum_{\boldsymbol{w}\in I_2(\boldsymbol{y})} \mathsf{Emb}(\boldsymbol{w};\boldsymbol{y}) \left(d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{w}) - d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{w})\right)\\
&= \sum_{\substack{\boldsymbol{w}\in I_2(\boldsymbol{y})\\ \boldsymbol{w}\neq\boldsymbol{c},\boldsymbol{c}'}} \mathsf{Emb}(\boldsymbol{w};\boldsymbol{y}) \left(d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{w}) - d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{w})\right)\\
&\quad + \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}) \left(d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c}) - d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c})\right)\\
&\quad + \mathsf{Emb}(\boldsymbol{c}';\boldsymbol{y}) \left(d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c}') - d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c}')\right) \geqslant 0.
\end{aligned}$$

Since $d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{w}) - d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{w})\geqslant 0$ for every $\boldsymbol{w}\neq\boldsymbol{c}$, it is enough to find $\boldsymbol{c}'\in I_2(\boldsymbol{y})$ such that,

$$\begin{aligned}
&\mathsf{Emb}(\boldsymbol{c};\boldsymbol{y})\Big(d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c}) - d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c})\Big)\\
&+\mathsf{Emb}(\boldsymbol{c}';\boldsymbol{y})\Big(d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c}') - d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c}')\Big)\geqslant 0.
\end{aligned}$$

Recall that the embedding number decoder prolongs the first longest run in $\boldsymbol{y}$. If the first longest run in $\boldsymbol{c}$, which is the i-th run, satisfies $i\neq j$, this run is also the first longest run in $\boldsymbol{y}$. In this case, let $\boldsymbol{c}'$ be the word obtained from $\boldsymbol{y}$ by prolonging this run by two symbols. It holds that, $d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c}')=1$, $d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c}')=5$, and $\mathsf{Emb}(\boldsymbol{c}';\boldsymbol{y})=\binom{r'_i+2}{2}$. Recall that $r'_i\geqslant r'_j$ and hence,

$$-2\binom{r'_j}{2}+4\binom{r'_i+2}{2}\geqslant 0.$$

Else, if the first longest run in $\boldsymbol{c}$ is the j-th run (i.e., $i=j$) and all the other runs in $\boldsymbol{c}$ are strictly shorter in more than two symbols from the j-th run. Then, the j-th run is also the first longest run in $\boldsymbol{y}$. In this case $\mathcal{D}(\boldsymbol{y})=\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})$ which is a contradiction to the definition of $\mathcal{D}(\boldsymbol{y})$. Otherwise, the longest run in $\boldsymbol{c}$ is the j-th run and there exists $s<j$ such that $r'_s+2\geqslant r'_j$, which implies that the s-th run is the first longest run in $\boldsymbol{y}$. By Lemma 13, D_{EN}^{n-1} prolongs the s-th run of $\boldsymbol{y}$ by one symbol. Let $\boldsymbol{c}'$ be the word that is obtained from $\boldsymbol{y}$ by prolonging the s-th run by two symbols, it holds that $d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c}')=1$, $d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c}')=5$ and

$$\mathsf{Emb}(\boldsymbol{c}';\boldsymbol{y})=\binom{r'_s+2}{2}\geqslant\binom{r'_j}{2}=\mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}).$$

Which implies that ,

$$-2\binom{r'_j}{2}+4\binom{r'_s+2}{2}\geqslant 0.$$

2) There exist $1\leqslant j<j'\leqslant k'$ such that $\boldsymbol{y}$ is obtained from $\boldsymbol{c}$ by deleting one symbol from the j-th run and one symbol from the j'-th run. Similarly to the previous case, $\mathcal{D}(\boldsymbol{y})$ must be obtained from $\boldsymbol{c}$ by deleting one symbol from the h-th run for some $h\neq j,j'$. Hence, $\boldsymbol{c}$ is the unique word that is obtained from $\boldsymbol{y}$ by inserting one symbol to the j-th run, and one symbol to the j'-th run, that is,

$$I_2(\boldsymbol{y})\cap I_1(\mathcal{D}(\boldsymbol{y}))=\{\boldsymbol{c}\}.$$

Note that $\mathsf{Emb}(\boldsymbol{c};\boldsymbol{y})=r'_jr'_{j'}$ and that $d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c})=1$ and $d_L(\mathcal{D}^{n-1}_{EN}(\boldsymbol{y}),\boldsymbol{c})\in\{1,3\}$. Similarly to the previous case we can assume that $d_L(\mathcal{D}^{n-1}_{EN}(\boldsymbol{y}),\boldsymbol{c})=3$ and our goal is to find a word $\boldsymbol{c}'\in I_2(\boldsymbol{y})$ such that,

$$\begin{aligned}&\mathsf{Emb}(\boldsymbol{c};\boldsymbol{y})\left(d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c})-d_L(\mathcal{D}^{n-1}_{EN}(\boldsymbol{y}),\boldsymbol{c})\right)\\&+\mathsf{Emb}(\boldsymbol{c}';\boldsymbol{y})\left(d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c}')-d_L(\mathcal{D}^{n-1}_{EN}(\boldsymbol{y}),\boldsymbol{c}')\right)\geqslant 0.\end{aligned}$$

Similarly to the previous case, if the i-th run, which is the first longest run in $\boldsymbol{c}$ satisfies $i\neq j,j'$, the same run is also the first longest run in $\boldsymbol{y}$. Let $\boldsymbol{c}'$ be the word that is obtained from $\boldsymbol{y}$ by prolonging this longest run by two symbols. It holds that $d_L(\mathcal{D}^{n-1}_{EN}(\boldsymbol{y}),\boldsymbol{c}')=1$, $d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c}')=5$ and $\mathsf{Emb}(\boldsymbol{c}';\boldsymbol{y})=\binom{r'_i+2}{2}$, and since, $r'_i\geqslant r'_j,r'_{j'}$,

$$-2r'_jr'_{j'}+4\binom{r'_i+2}{2}\geqslant 0.$$

Else, if the first longest run in $\boldsymbol{c}$ is the j-th run, or the j'-th run (i.e., $i\in\{j,j'\}$), and the same run is also the first longest run in $\boldsymbol{y}$. Then, similarly to the previous case $\mathcal{D}(\boldsymbol{y})=\mathcal{D}^{n-1}_{EN}(\boldsymbol{y})$ which contradicts the definition of $\mathcal{D}(\boldsymbol{y})$. Otherwise, $i\in\{j,j'\}$, and there exists $s\neq j,j'$ such that $r'_s+1\geqslant r'_j,r'_{j'}$. In other words this run is the first longest run in $\boldsymbol{y}$. By Lemma 13, $\mathcal{D}^{n-1}_{EN}$ prolongs this run by one symbol. Assume w.l.o.g. that $r'_j\geqslant r'_{j'}$ and let $\boldsymbol{c}'$ be the word obtained from $\boldsymbol{c}$ by deleting one symbol from the j'-th run and prolonging the s-th run by one symbol. In this case $d_L(\mathcal{D}^{n-1}_{EN}(\boldsymbol{y}),\boldsymbol{c}')=1$, $d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c}')=3$ and

$$\mathsf{Emb}(\boldsymbol{c}';\boldsymbol{y})=r'_j(r'_s+1)\geqslant r'_jr'_{j'}=\mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}).$$

Therefore,

$$-2r'_jr'_{j'}+2r'_j(r'_s+1)\geqslant 0.$$

■

■

Lemma 20. *It holds that*

$$\mathcal{D}^n_{\mathsf{ML}^*}=\mathcal{D}^n_{EN}.$$

Proof: For any channel output $\boldsymbol{y}\in\Sigma_2^{n-2}$ and for any decoder $\mathcal{D}$, such that $|\mathcal{D}(\boldsymbol{y})|=n$, we have the following sequence of equalities and inequalities,

$$\begin{aligned}f_{\boldsymbol{y}}(\mathcal{D}(\boldsymbol{y}))&=\sum_{\boldsymbol{c}\in\Sigma_2^n}\frac{d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c})}{|\boldsymbol{c}|}p(\boldsymbol{y}|\boldsymbol{c})\\&=\sum_{\boldsymbol{c}\in I_2(\boldsymbol{y})}\frac{d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c})}{|\boldsymbol{c}|}p(\boldsymbol{y}|\boldsymbol{c})\\&\overset{(a)}{\geqslant}\frac{2}{n}\sum_{\boldsymbol{c}\in I_2(\boldsymbol{y})}p(\boldsymbol{y}|\boldsymbol{c})-\frac{2}{n}p(\boldsymbol{y}|\mathcal{D}(\boldsymbol{y}))\\&\overset{(b)}{=}\frac{2}{n}\sum_{\boldsymbol{c}\in I_2(\boldsymbol{y})}p(\boldsymbol{y}|\boldsymbol{c})-\frac{2}{n}\frac{\mathsf{Emb}(\mathcal{D}(\boldsymbol{y});\boldsymbol{y})}{\binom{n}{2}}\\&\overset{(c)}{\geqslant}\frac{2}{n}\sum_{\boldsymbol{c}\in I_2(\boldsymbol{y})}p(\boldsymbol{y}|\boldsymbol{c})-\frac{2}{n}\max_{\boldsymbol{c}\in\Sigma_2^n}\left\{\frac{\mathsf{Emb}(\boldsymbol{c};\boldsymbol{y})}{\binom{n}{2}}\right\}\\&\overset{(d)}{=}\frac{2}{n}\sum_{\boldsymbol{c}\in I_2(\boldsymbol{y})}p(\boldsymbol{y}|\boldsymbol{c})-\frac{2}{n}\frac{\mathsf{Emb}(\mathcal{D}^n_{EN}(\boldsymbol{y});\boldsymbol{y})}{\binom{n}{2}}\\&=f_{\boldsymbol{y}}(\mathcal{D}^n_{EN}(\boldsymbol{y})),\end{aligned}$$

where (a) holds since for every $\boldsymbol{c}$ such that $\mathcal{D}(\boldsymbol{y})\neq\boldsymbol{c}$ it holds that $d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c})\geqslant 2$, and $d_L(\mathcal{D}(\boldsymbol{y}),\mathcal{D}(\boldsymbol{y}))=0$, (b) is obtained by the definition of the 2-deletion channel, and in (c) we simply choose the word that maximizes the value of $\mathsf{Emb}(\boldsymbol{c};\boldsymbol{y})$, which is the definition of the EN decoder of length n as derived in step (d). This verifies the lemma's statement. ■

Lemma 21. *Let $\boldsymbol{y}\in\Sigma_2^{n-2}$ be a channel output. It holds that*

$$|\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})|\neq n.$$

Proof: Assume to the contrary that $|\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})|=n$. We show that

$$f_{\boldsymbol{y}}(\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y}))\geqslant f_{\boldsymbol{y}}(\mathcal{D}_{Lazy}(\boldsymbol{y})),$$

which is a contradiction to the definition of the ML* decoder (since the ML* decoder is defined to return the shortest word that minimizes $f_{\boldsymbol{y}}(\cdot)$). By Lemma 20, $\mathcal{D}^n_{EN}(\boldsymbol{y})$ is the decoder that minimizes $f_{\boldsymbol{y}}(\mathcal{D}(\boldsymbol{y}))$ among all other decoders that return a word of length n for the channel output $\boldsymbol{y}$. Hence, it is enough to show that $f_{\boldsymbol{y}}(\mathcal{D}^n_{EN}(\boldsymbol{y}))-f_{\boldsymbol{y}}(\mathcal{D}_{Lazy}(\boldsymbol{y}))\geqslant 0$.

Note that since $\mathcal{D}^n_{EN}$ reutrns a word of length n that is a supersequence of $\boldsymbol{y}$ and therefore any possible output of $\mathcal{D}^n_{EN}$ is either of distance $0,2$, or 4 from the transmitted word $\boldsymbol{c}$. Hence,

$$
\begin{aligned}
&f_{\boldsymbol{y}}(\mathcal{D}^n_{EN}(\boldsymbol{y})) - f_{\boldsymbol{y}}(\mathcal{D}_{Lazy}(\boldsymbol{y})) \\
&\quad= \sum_{\boldsymbol{c}\in I_2(\boldsymbol{y})} \frac{p(\boldsymbol{y}|\boldsymbol{c})}{|\boldsymbol{c}|}\left(d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c}) - d_L(\mathcal{D}_{Lazy}(\boldsymbol{y}),\boldsymbol{c})\right) \\
&\quad\overset{(a)}{=} \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c})=4}} \frac{p(\boldsymbol{y}|\boldsymbol{c})}{|\boldsymbol{c}|}(4-2) \\
&\qquad+ \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c})=2}} \frac{p(\boldsymbol{y}|\boldsymbol{c})}{|\boldsymbol{c}|}(2-2) \\
&\qquad+ \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c})=0}} \frac{p(\boldsymbol{y}|\boldsymbol{c})}{|\boldsymbol{c}|}(0-2) \\
&\quad\overset{(b)}{=} \frac{2}{n}\left(\sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c})=4}} p(\boldsymbol{y}|\boldsymbol{c}) - \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c})=0}} p(\boldsymbol{y}|\boldsymbol{c})\right),
\end{aligned}
$$

where (a) holds since $d_L(\mathcal{D}_{Lazy}(\boldsymbol{y}),\boldsymbol{c}) = 2$ for every $\boldsymbol{c} \in I_2(\boldsymbol{y})$ and (b) holds since $|\boldsymbol{c}| = n$.

Denote,

$$
\begin{aligned}
\mathcal{S}um_4 &\triangleq \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c})=4}} p(\boldsymbol{y}|\boldsymbol{c}),\\
\mathcal{P}_0 &\triangleq \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c})=0}} p(\boldsymbol{y}|\boldsymbol{c}) = p\left(\boldsymbol{y}|\mathcal{D}^n_{EN}(\boldsymbol{y})\right).
\end{aligned}
$$

From the above discussion, our objective is to prove that $\mathcal{S}um_4 \geqslant \mathcal{P}_0$. Recall that $|I_2(\boldsymbol{y})| = \binom{n}{2} + n + 1$. Let the i-th, i'-th run be the first, second longest run of $\boldsymbol{y}$, respectively, and denote their lengths by $r_i \geqslant r_{i'}$. We will bound the number of possible words $\boldsymbol{c} \in I_2(\boldsymbol{y})$ such that $d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c}) = 4$.

Case 1: $\mathcal{D}^n_{EN}$ prolongs the i-th run by two symbols. There is one word $\boldsymbol{c} \in I_2(\boldsymbol{y})$ such that $d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c}) = 0$. Note that the set of words $\boldsymbol{c} \in I_2(\boldsymbol{y})$ such that $d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c}) = 2$ consists of words $\boldsymbol{c}$ that can be obtained from $\boldsymbol{y}$ by prolonging the i-th run by exactly one symbol. Consider the word $\boldsymbol{y}'$, which is the word obtained from $\boldsymbol{y}$ by prolonging the i-th run by exactly one symbol. $\boldsymbol{y}'$ is a word of length $n-1$, and the words $\boldsymbol{c}$, such that $d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c}) = 2$ are all the words in the radius-1 insertion ball centered at $\boldsymbol{y}'$ expect to the word $\mathcal{D}^n_{EN}(\boldsymbol{y})$. The number of such words is

$$
I_1(\boldsymbol{y}') - 1 = n + 1 - 1 = n.
$$

Hence, there are $\binom{n}{2}$ words $\boldsymbol{c} \in I_2(\boldsymbol{y})$ which satisfy that $d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c}) = 4$ and the conditional probability of each of these words is $p(\boldsymbol{y}|\boldsymbol{c}) \geqslant \frac{1}{\binom{n}{2}}$. Therefore,

$$
\mathcal{S}um_4 = \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c})=4}} p(\boldsymbol{y}|\boldsymbol{c}) \geqslant \binom{n}{2} \cdot \frac{1}{\binom{n}{2}} = 1.
$$

On the other hand,

$$
\mathcal{P}_0 = \frac{\binom{r_i+2}{2}}{\binom{n}{2}} \leqslant 1,
$$

which implies $\mathcal{S}um_4 \geqslant \mathcal{P}_0$ for every $n > 0$ and thus,

$$
f_{\boldsymbol{y}}(\mathcal{D}^n_{EN}(\boldsymbol{y})) - f_{\boldsymbol{y}}(\mathcal{D}_{Lazy}(\boldsymbol{y})) \geqslant 0.
$$

Case 2: $\mathcal{D}^n_{EN}(\boldsymbol{y})$ prolongs both the i-th run and the i'-th run, each by one symbol. By Lemma 13, we know that $\mathcal{D}^n_{EN}(\boldsymbol{y})$ prolongs these two runs if and only if $\binom{r_i}{2} \leqslant r_i r_{i'}$, and consequently, $\frac{r_i - 1}{2} \leqslant r_{i'} < r_i$.

The only word $\boldsymbol{c} \in I_2(\boldsymbol{y})$ that satisfies $d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c}) = 0$ is the word $\boldsymbol{c} = \mathcal{D}^n_{EN}(\boldsymbol{y})$. In addition the set of words $\boldsymbol{c} \in I_2(\boldsymbol{y})$ such that $d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c}) = 2$ consists of words $\boldsymbol{c}$ that can be obtained from $\boldsymbol{y}$ by prolonging either the i-th run or the i'-run by exactly one symbol. Let $\boldsymbol{y}'$ be the word obtained from $\boldsymbol{y}$ by prolonging the i-th run by one symbol and let $\boldsymbol{y}''$ be the word obtained from $\boldsymbol{y}$ by prolonging the i'-th run by one symbol. Similarly to the first case the number of such words is

$$
I_1(\boldsymbol{y}') - 1 + I_1(\boldsymbol{y}'') - 1 = 2n,
$$

which implies that the number of words $\boldsymbol{c} \in I_2(\boldsymbol{y})$ such that $d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c}) = 4$ is $\binom{n}{2} - n$ and the conditional probabilities of these words satisfy $p(\boldsymbol{y}|\boldsymbol{c}) \geqslant \frac{1}{\binom{n}{2}}$. Hence,

$$
\mathcal{S}um_4 = \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}^n_{EN}(\boldsymbol{y}),\boldsymbol{c})=4}} p(\boldsymbol{y}|\boldsymbol{c}) \geqslant \frac{\binom{n}{2} - n}{\binom{n}{2}}.
$$

On the other hand,

$$
\begin{aligned}
\mathcal{P}_0 &= \frac{(r_i+1)(r_{i'}+1)}{\binom{n}{2}} \overset{(a)}{\leqslant} \frac{(r_i+1)(n-r_i-1)}{\binom{n}{2}}\\
&\overset{(b)}{\leqslant} \frac{(\frac{n}{2}-1)^2}{\binom{n}{2}} = \frac{\frac{n^2}{4} - n + 1}{\binom{n}{2}},
\end{aligned}
$$

where (a) holds since $r_{i'} + r_i \leqslant n - 2$ and (b) holds since the maximum of the function $f(x) = x(n-x)$ is achieved for $x = n/2$. Hence, $\mathcal{S}um_4 \geqslant \mathcal{P}_0$ when $\frac{n^2}{4} - n + 1 \leqslant \binom{n}{2} - n$, which holds for all $n \geqslant 4$. Thus, for $n \geqslant 4$,

$$
f_{\boldsymbol{y}}(\mathcal{D}^n_{EN}(\boldsymbol{y})) - f_{\boldsymbol{y}}(\mathcal{D}_{Lazy}(\boldsymbol{y})) \geqslant 0.
$$

■

Lemma 22. *Let $\boldsymbol{y} \in \Sigma_2^{n-2}$ be a channel output. For any decoder $\mathcal{D}$, such that $\mathcal{D}(\boldsymbol{y})$ is not a supersequence of $\boldsymbol{y}$ and $|\mathcal{D}(\boldsymbol{y})| = n+1$, it holds that*

$$
f_{\boldsymbol{y}}(\mathcal{D}(\boldsymbol{y})) \geqslant f_{\boldsymbol{y}}(\mathcal{D}^{n-1}_{EN}(\boldsymbol{y})).
$$

Proof: Since $\mathcal{D}(\boldsymbol{y})$ is not a supersequence of $\boldsymbol{y}$, it is also not a supersequence of the transmitted word $\boldsymbol{c}$. Therefore, for each $\boldsymbol{c} \in I_2(\boldsymbol{y})$ it holds that $d_L(\mathcal{D}(\boldsymbol{y}),\boldsymbol{c}) \geqslant 3$, while $d_L(\mathcal{D}^{n-1}_{EN}(\boldsymbol{y}),\boldsymbol{c}) \leqslant 3$. Thus,

$$
\begin{aligned}
&f_{y}(\mathcal{D}(y)) - f_{y}(\mathcal{D}_{EN}^{n-1}(y))\\
&= \sum_{c\in I_2(y)} \frac{d_L(\mathcal{D}(y),c)}{|c|} p(y|c) - \sum_{c\in I_2(y)} \frac{d_L(\mathcal{D}_{EN}^{n-1}(y),c)}{|c|} p(y|c)\\
&= \frac{1}{|c|}\left(\sum_{c\in I_2(y)} d_L(\mathcal{D}(y),c)p(y|c) - \sum_{c\in I_2(y)} d_L(\mathcal{D}_{EN}^{n-1}(y),c)p(y|c)\right)\\
&= \frac{1}{|c|}\sum_{c\in I_2(y)} p(y|c)\left(d_L(\mathcal{D}(y),c) - d_L(\mathcal{D}_{EN}^{n-1}(y),c)\right)\\
&\geqslant \frac{1}{|c|}\sum_{c\in I_2(y)} p(y|c)\,(3-3) \geqslant 0.
\end{aligned}
$$

■

Lemma 23. *Let $y \in \Sigma_2^{n-2}$ be a channel output. For any decoder $\mathcal{D}$, such that $\mathcal{D}(y)$ is a supersequence of y and $|\mathcal{D}(y)| = n+1$, it holds that*

$$f_y(\mathcal{D}(y)) \geqslant f_y(\mathcal{D}_{EN}^{n-1}(y)).$$

Proof: From similar arguments to those presented in Lemma 16, our goal is to prove that (4) holds for $\mathcal{D}(y)$ and $\mathcal{D}_{EN}^{n-1}(y)$, i.e., to prove that

$$\sum_{c\in I_2(y)} \mathsf{Emb}(c;y)\left(d_L(\mathcal{D}(y),c) - d_L(\mathcal{D}_{EN}^{n-1}(y),c)\right) \geqslant 0.$$

Assume that the number of runs in y is $\rho(y) = k$, let r_j denote the length of the j-th run for $1 \leqslant j \leqslant k$, and let the i-th run of y be the first longest run of y. Note that the Levenshtein distance of $\mathcal{D}_{EN}^{n-1}(y)$ from the transmitted word c can be either 1 or 3. Similarly, $\mathcal{D}(y)$ can have distance of 1, 3 or 5 from c. Recall that $\mathcal{D}_{EN}^{n-1}$ prolongs the i-th run by one symbol and that $I_1(\mathcal{D}_{EN}^{n-1}(y)) \subseteq I_2(y)$. $\mathcal{D}(y)$ is a supersequence of y, and hence $\mathcal{D}(y)$ is obtained from y by prolonging existing runs or by creating new runs in y. From the discussion above, for every word $c \in I_2(y)$ such that

$$c \notin \left(I_1(\mathcal{D}_{EN}^{n-1}(y)) \cup D_1(\mathcal{D}(y))\right),$$

it holds that $d_L(\mathcal{D}_{EN}^{n-1}(y),c) = 3$ while $d_L(\mathcal{D}(y),c) \geqslant 3$. Additionally, every word $c \in I_2(y)$ such that

$$c \in \left(I_1(\mathcal{D}_{EN}^{n-1}(y)) \cap D_1(\mathcal{D}(y))\right),$$

satisfies $d_L(\mathcal{D}_{EN}^{n-1}(y),c) = d_L(\mathcal{D}(y),c) = 1$. Hence, for these words it holds that $d_L(\mathcal{D}(y),c) - d_L(\mathcal{D}_{EN}^{n-1}(y),c) \geqslant 0$ and they can be eliminated from inequality (4). In order to complete the proof, the words $c \in I_2(y)$ such that

$$c \in I_1(\mathcal{D}_{EN}^{n-1}(y)) \text{ and } c \notin D_1(\mathcal{D}(y))$$

and the words $c \in I_2(y)$ such that

$$c \notin I_1(\mathcal{D}_{EN}^{n-1}(y)) \text{ and } c \in D_1(\mathcal{D}(y))$$

should be considered. For words in the first case it holds that $d_L(\mathcal{D}_{EN}^{n-1}(y),c) = 1$ and $d_L(\mathcal{D}(y),c) \geqslant 3$, while for words in the latter case, $d_L(\mathcal{D}_{EN}^{n-1}(y),c) = 3$ and $d_L(\mathcal{D}(y),c) \geqslant 1$. Hence,

$$
\begin{aligned}
&\sum_{c\in I_2(y)} \mathsf{Emb}(c;y)\left(d_L(\mathcal{D}(y),c) - d_L(\mathcal{D}_{EN}^{n-1}(y),c)\right)\\
&\geqslant \sum_{\substack{c\in I_2(y)\\ c\in I_1(\mathcal{D}_{EN}^{n-1}(y))\\ c\notin D_1(\mathcal{D}(y))}} \mathsf{Emb}(c;y)\left(d_L(\mathcal{D}(y),c) - d_L(\mathcal{D}_{EN}^{n-1}(y),c)\right)\\
&+ \sum_{\substack{c\in I_2(y)\\ c\notin I_1(\mathcal{D}_{EN}^{n-1}(y))\\ c\in D_1(\mathcal{D}(y))}} \mathsf{Emb}(c;y)\left(d_L(\mathcal{D}(y),c) - d_L(\mathcal{D}_{EN}^{n-1}(y),c)\right)\\
&\geqslant 2\sum_{\substack{c\in I_2(y)\\ c\in I_1(\mathcal{D}_{EN}^{n-1}(y))\\ c\notin D_1(\mathcal{D}(y))}} \mathsf{Emb}(c;y) - 2\sum_{\substack{c\in I_2(y)\\ c\notin I_1(\mathcal{D}_{EN}^{n-1}(y))\\ c\in D_1(\mathcal{D}(y))}} \mathsf{Emb}(c;y).
\end{aligned}
$$

We first assume that $\mathcal{D}(y)$ is obtained from y by prolonging the i-th run by exactly one symbol. Let $c \in I_2(y)$ and consider the cases mentioned above.

1) $c \in I_1(\mathcal{D}_{EN}^{n-1}(y))$ and $c \notin D_1(\mathcal{D}(y))$: Recall that both decoders return supersequences of y. By the assumption $\mathcal{D}(y)$ is obtained from y by prolonging the i-th run by one symbol and then performing two more insertions to the obtained word. Since $c \in I_1(\mathcal{D}_{EN}^{n-1}(y))$, c must be obtained from y by prolonging the i-th run and performing one more insertion. $c \notin D_1(\mathcal{D}(y))$, and therefore the number of such words equals to

$$|I_1(\mathcal{D}_{EN}^{n-1}(y))| - \left|\left\{c \in I_2(y) \ : \ c \in I_1(\mathcal{D}_{EN}^{n-1}(y)) \cap D_1(\mathcal{D}(y))\right\}\right|.$$

Note that

$$\left|\left\{c \in I_2(y) \ : \ c \in I_1(\mathcal{D}_{EN}^{n-1}(y)) \cap D_1(\mathcal{D}(y))\right\}\right| \leqslant 2$$

since the words in the latter intersection are the words that obtain from y by prolonging the i-th run by one symbol and then performing one of the two other insertions performed to receive $\mathcal{D}(y)$. Hence, there are at least $|I_1(\mathcal{D}_{EN}^{n-1}(y))| - 2 = n-1$ such words in this case and for each of them $\mathsf{Emb}(c;y) \geqslant (r_i+1)$. Recall that these words satisfy $d(\mathcal{D}_{EN}^{n-1}(y),c) = 1$ and $d(\mathcal{D}(y),c) \geqslant 3$.

2) $c \notin I_1(\mathcal{D}_{EN}^{n-1}(y))$ and $c \in D_1(\mathcal{D}(y))$: By the assumption, $\mathcal{D}$ prolongs the i-th run by one symbol and performs two more insertions into the obtained word and $\mathcal{D}_{EN}^{n-1}$ prolongs the i-th run by one symbol. Hence, the words $c \in I_2(y)$ such that $c \notin I_1(\mathcal{D}_{EN}^{n-1}(y))$ and $c \in D_1(\mathcal{D}(y))$ can not be obtained from y by prolonging the i-th run. Therefore, it implies that c is the unique word obtained from $\mathcal{D}(y)$ by deleting the symbol that was inserted to the i-th run of y. It holds that $\mathsf{Emb}(c;y) \leqslant (r_i+1)^2$ and $d_L(\mathcal{D}_{EN}^{n-1}(y),c) = 3$ and $d_L(\mathcal{D}(y),c) = 1$.

Note that $r_i \leqslant n-2$ since it is the length of the i-th run of

$\boldsymbol{y} \in \Sigma_2^{n-2}$. Thus,

$$\begin{aligned}
&2 \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ \boldsymbol{c}\in I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))\\ \boldsymbol{c}\notin D_1(\mathcal{D}(\boldsymbol{y}))}} \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}) - 2 \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ \boldsymbol{c}\notin I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))\\ \boldsymbol{c}\in D_1(\mathcal{D}(\boldsymbol{y}))}} \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y})\\
&\geqslant 2(n-1)(r_i+1) - 2\cdot(r_i+1)^2\\
&\geqslant 2(r_i+1)^2 - 2\cdot(r_i+1)^2 \geqslant 0.
\end{aligned}$$

Second we assume that $\mathcal{D}(\boldsymbol{y})$ is obtained from $\boldsymbol{y}$ by prolonging the i-th run by at least two symbols. In this case, it holds that $(D_1(\mathcal{D}(\boldsymbol{y})) \cap I_2(\boldsymbol{y})) \subseteq I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))$, which implies that

$$\left|\left\{\boldsymbol{c} \in I_2(\boldsymbol{y}) \ : \ \boldsymbol{c} \notin I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})) \text{ and } \boldsymbol{c} \in D_1(\mathcal{D}(\boldsymbol{y}))\right\}\right| = 0,$$

and therefore,

$$2 \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ \boldsymbol{c}\in I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))\\ \boldsymbol{c}\notin D_1(\mathcal{D}(\boldsymbol{y}))}} \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}) - 2 \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ \boldsymbol{c}\notin I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))\\ \boldsymbol{c}\in D_1(\mathcal{D}(\boldsymbol{y}))}} \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}) \geqslant 0.$$

Lastly, we assume that $\mathcal{D}(\boldsymbol{y})$ is obtained from $\boldsymbol{y}$ by three insertions such that neither of these insertions prolongs the i-th run. It holds that,

$$\left|\left\{\boldsymbol{c} \in I_2(\boldsymbol{y}) \ : \ \boldsymbol{c} \in I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})) \cap D_1(\mathcal{D}(\boldsymbol{y}))\right\}\right| = 0.$$

Therefore the number of words $\boldsymbol{c} \in I_2(\boldsymbol{y})$ such that $\boldsymbol{c} \in I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))$ and $\boldsymbol{c} \notin D_1(\mathcal{D}(\boldsymbol{y}))$ equals to $|I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))| = n+1$. For any such word $\boldsymbol{c}$ it holds that $\mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}) \geqslant r_i+1$. Furthermore, $|D_1(\mathcal{D}(\boldsymbol{y}))|$ equals to the number of runs in $\mathcal{D}(\boldsymbol{y})$ [26] and any $\boldsymbol{c} \in D_1(\mathcal{D}(\boldsymbol{y})) \cap I_2(\boldsymbol{y})$ is obtained from $\mathcal{D}(\boldsymbol{y})$ by deleting one symbol of the three insertions to $\boldsymbol{y}$ in order to obtain $\mathcal{D}(\boldsymbol{y})$. Hence, there are at most three such words, and each is obtained by deleting one of the three inserted symbols. Let $\boldsymbol{c}$ be one of those words. If the two remaining symbols belong to the same run, then $\mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}) = \binom{m}{2}$ where m is the length of this run in $\boldsymbol{c}$ and $m \leqslant r_i+2$. In this case consider the word $\boldsymbol{c}'$ that is obtained by prolonging the i-th run of $\boldsymbol{y}$ by two symbols. It holds that,

$$\mathsf{Emb}(\boldsymbol{c}';\boldsymbol{y}) = \binom{r_i+2}{2} \geqslant \binom{m}{2} = \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}).$$

Otherwise, $\mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}) = m_1m_2$ where m_1 and m_2 are the lengths of the runs that include the remaining inserted symbols and $m_1, m_2 \leqslant r_i+1$. Let $\boldsymbol{c}'$ be the word that is obtained from $\boldsymbol{y}$ by prolonging the i-th run and the run of length $\max\{m_1-1, m_2-1\}$ that is prolonged by $\mathcal{D}$. In this case,

$$\mathsf{Emb}(\boldsymbol{c}';\boldsymbol{y}) = m_1(r_i+1) \geqslant m_1m_2 = \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}).$$

Note that there is at most one such word $\boldsymbol{c}$ that is obtained by prolonging the same run with two symbols, which implies that there is always a selection of words $\boldsymbol{c}'$ such that,

$$2 \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ \boldsymbol{c}\in I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))\\ \boldsymbol{c}\notin D_1(\mathcal{D}(\boldsymbol{y}))}} \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}) - 2 \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ \boldsymbol{c}\notin I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))\\ \boldsymbol{c}\in D_1(\mathcal{D}(\boldsymbol{y}))}} \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}) \geqslant 0.$$

We proved that for any decoder $\mathcal{D}$ such that $\mathcal{D}(\boldsymbol{y})$ is a supersequence $\boldsymbol{y}$ and $|\mathcal{D}(\boldsymbol{y})| = n+1$,

$$2 \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ \boldsymbol{c}\in I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))\\ \boldsymbol{c}\notin D_1(\mathcal{D}(\boldsymbol{y}))}} \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}) - 2 \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ \boldsymbol{c}\notin I_1(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}))\\ \boldsymbol{c}\in D_1(\mathcal{D}(\boldsymbol{y}))}} \mathsf{Emb}(\boldsymbol{c};\boldsymbol{y}) \geqslant 0.$$

Thus,

$$f_{\boldsymbol{y}}(\mathcal{D}(\boldsymbol{y})) - f_{\boldsymbol{y}}(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})) \geqslant 0.$$

■

From the previous lemmas it holds that for a given channel output $\boldsymbol{y} \in \Sigma_2^{n-2}$, the length of $\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})$ is either $n-1$ or n. Lemma 16 implies that if $|\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y})| = n-1$, then $\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y}) = \mathcal{D}_{EN}^{n-1}(\boldsymbol{y})$. In the following result we define a condition on the length of the longest run in $\boldsymbol{y}$ to decide whether prolonging it by one symbol can minimize the average decoding error probability. In other words, this result defines a criteria on a given channel output $\boldsymbol{y}$ to define whether using the same output as $\mathcal{D}_{Lazy}$ or using the same output as $\mathcal{D}_{EN}^{n-1}$ is better in terms of minimizing $f_{\boldsymbol{y}}(\mathcal{D}(\boldsymbol{y}))$ (and therefore minimizing the error probability). An immediate conclusion of this result is Theorem 25 which determines the ML* decoder for the case of a single 2-deletion channel.

Lemma 24. *Let $\boldsymbol{y} \in \Sigma_2^{n-2}$ be a channel output, such that the number of runs in $\boldsymbol{y}$ is $\rho(\boldsymbol{y}) = k$, and the first longest run in $\boldsymbol{y}$ is the i-th run. Denote by r_j the length of the j-th for $1 \leqslant j \leqslant k$. It holds that*

$$f_{\boldsymbol{y}}(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})) - f_{\boldsymbol{y}}(\mathcal{D}_{Lazy}(\boldsymbol{y})) \geqslant 0$$

if and only if

$$2n^2 - 4nr_i - 6n + r_i^2 + 3r_i + k + 1 \geqslant 0.$$

Proof: By Lemma 12, $\mathcal{D}_{EN}^{n-1}$ prolongs the i-th run of $\boldsymbol{y}$ by one symbol. Therefore, the Levenshtein distance of $\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})$ from the transmitted word $\boldsymbol{c}$ can be either 1 or 3. Hence,

$$\begin{aligned}
&f_{\boldsymbol{y}}(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y})) - f_{\boldsymbol{y}}(\mathcal{D}_{Lazy}(\boldsymbol{y}))\\
&= \sum_{\boldsymbol{c}\in I_2(\boldsymbol{y})} \frac{p(\boldsymbol{y}|\boldsymbol{c})}{|\boldsymbol{c}|}\left(d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c}) - d_L(\mathcal{D}_{Lazy}(\boldsymbol{y}),\boldsymbol{c})\right)\\
&= \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c})=3}} \frac{p(\boldsymbol{y}|\boldsymbol{c})}{|\boldsymbol{c}|}(3-2)\\
&\quad + \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c})=1}} \frac{p(\boldsymbol{y}|\boldsymbol{c})}{|\boldsymbol{c}|}(1-2)\\
&= \frac{1}{n}\left(\sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c})=3}} p(\boldsymbol{y}|\boldsymbol{c}) - \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c})=1}} p(\boldsymbol{y}|\boldsymbol{c})\right).
\end{aligned}$$

Denote

$$\mathcal{S}um_3 \triangleq \sum_{\substack{c\in I_2(y)\\ d_L(\mathcal{D}_{EN}^{n-1}(y),c)=3}} p(y|c),$$

$$\mathcal{S}um_1 \triangleq \sum_{\substack{c\in I_2(y)\\ d_L(\mathcal{D}_{EN}^{n-1}(y),c)=1}} p(y|c).$$

Let us prove that

$$2n^2 - 4nr_i - 6n + r_i^2 + 3r_i + k + 1 \geqslant 0$$

is a necessary and sufficient condition for the inequality $\mathcal{S}um_3 \geqslant \mathcal{S}um_1$ to hold. First, we count the number of words $c \in I_2(y)$ such that $d_L(\mathcal{D}_{EN}^{n-1}(y), c) = 1$. Each such c is a supersequence of $\mathcal{D}_{EN}^{n-1}(y)$ and therefore c can be obtained from y only by one of the three following ways. The first way is by prolonging the i-th run and the j-th of y for $j \neq i$, each by one symbol. The number of such words is $k-1$. The second way is by prolonging the i-th run in y by one symbol and creating a new run in y. The number of options to create a new run in y is $n-k+1$ and therefore, there are $n-k+1$ such words. The third way is by prolonging the i-th run by two symbols and there is only one such word. Hence, the total number of words $c \in I_2(y)$ such that $d_L(\mathcal{D}_{EN}^{n-1}(y), c) = 1$ is $n+1 = |I_1(\mathcal{D}_{EN}^{n-1}(y))|$. Among them, the $k-1$ words that are obtained by the first way has an embedding number of $\mathsf{Emb}(c; y) = (r_i+1)(r_j+1)$. Similarly the $n-k+1$ words that are obtained from y using the second way satisfy $\mathsf{Emb}(c; y) = r_i + 1$. Lastly, for the word c that is obtained by prolonging the i-th run of y by two symbols it holds that $\mathsf{Emb}(c; y) = \binom{r_i+2}{2}$. Hence,

$$\begin{aligned}
\mathcal{S}um_1 &= \sum_{\substack{c\in I_2(y)\\ d_L(\mathcal{D}_{EN}^{n-1}(y),c)=1}} p(y|c) \\
&= \frac{\binom{r_i+2}{2}}{\binom{n}{2}} + \sum_{\substack{1\leqslant j\leqslant k\\ j\neq i}} \frac{(r_i+1)(r_j+1)}{\binom{n}{2}} + \sum_{j=1}^{n-k+1} \frac{(r_i+1)}{\binom{n}{2}} \\
&\overset{(a)}{=} \frac{(r_i+2)(r_i+1)}{2\binom{n}{2}} + \frac{(n-r_i-2+k-1)(r_i+1)}{\binom{n}{2}} \\
&\quad + \frac{(n-k+1)(r_i+1)}{\binom{n}{2}} \\
&= \frac{(2n-\frac{r_i}{2}-1)\cdot(r_i+1)}{\binom{n}{2}} = \frac{(4n-r_i-2)\cdot(r_i+1)}{n\cdot(n-1)},
\end{aligned}$$

where (a) holds since $\sum_{j\neq i} r_j = n-2-r_i$.

Next, let us evaluate the summation $\mathcal{S}um_3$. Note that if $d_L(\mathcal{D}_{EN}^{n-1}(y), c) = 3$ then c is not in a supersequence of $\mathcal{D}_{EN}^{n-1}(y)$, and hence $c \notin I_1(\mathcal{D}_{EN}^{n-1}(y))$. The words that contribute to the summation $\mathcal{S}um_3$ can be divided to three different types of words $c \in I_2(y)$.

Case 1: Let $\mathcal{C}_1 \subseteq I_2(y)$ be the set of words $c \in \mathcal{C}_1$, such that c includes additional run(s) that does not appear in y. Such additional runs can be either one run of length 2, or two runs of length 1 each. The number of words such that the length of the new run is two is $n-k$. And the number of words with two additional runs is $\binom{n-k}{2}$. Additionally, for $c \in \mathcal{C}_1$, $\mathsf{Emb}(c; y) = 1$, which implies,

$$\begin{aligned}
\sum_{c\in\mathcal{C}_1} p(y|c) &= \sum_{c\in\mathcal{C}_1} \frac{1}{\binom{n}{2}} \\
&= \frac{1}{\binom{n}{2}}\left(\binom{n-k}{2} + n - k\right) \\
&= \frac{2}{n(n-1)}\left(\frac{(n-k-1)(n-k)}{2} + n - k\right) \\
&= \frac{(n-k)(n-k+1)}{n(n-1)}.
\end{aligned}$$

Case 2: Let $\mathcal{C}_2 \subseteq I_2(y)$ be the set of words $c \in \mathcal{C}_2$, such that c is obtained from y by prolonging the j-th run and by creating a new run in y. Note that the prolonged run cannot be the i-th run in order to ensure $d_L(\mathcal{D}_{EN}^{n-1}(y), c) = 3$, i.e., $j \neq i$. The number of words in $\mathcal{C}_2$ is $(k-1)(n-k+1)$, since there are $k-1$ options for the index j, and $n-k+1$ ways to create a new run in the obtained word. For such a word $c \in \mathcal{C}_2$, it holds that $\mathsf{Emb}(c; y) = r_j + 1$ and hence,

$$\begin{aligned}
\sum_{c\in\mathcal{C}_2} p(y|c) &= \sum_{\substack{1\leqslant j\leqslant k\\ j\neq i}} (n-k+1)\cdot\frac{r_j+1}{\binom{n}{2}} \\
&= \frac{(n-k+1)}{\binom{n}{2}} \sum_{\substack{1\leqslant j\leqslant k\\ j\neq i}} (r_j+1) \\
&= \frac{2(n-k+1)}{n(n-1)}(n-r_i+k-3).
\end{aligned}$$

Case 3: Let $\mathcal{C}_3 \subseteq I_2(y)$ be the set of words $c \in \mathcal{C}_3$, such that c is obtained from y by prolonging one or two existing runs in y (other than the i-th run). The number of words $c \in \mathcal{C}_3$ obtained from y by prolonging a single run by two symbols is $k-1$. If the j-th run is the prolonged run then $\mathsf{Emb}(c; y) = \binom{r_j+2}{2}$. Additionally, there are $\binom{k-1}{2}$ words in $\mathcal{C}_3$ that are obtained by prolonging the j-th and the j'-th runs of y, each by one symbol. These words satisfy $\mathsf{Emb}(c; y) = (r_j+1)(r_{j'}+1)$. Therefore,

$$\begin{aligned}
\sum_{c\in\mathcal{C}_3} p(y|c) &= \sum_{\substack{1\leqslant j\leqslant k\\ j\neq i}} \frac{\binom{r_j+2}{2}}{\binom{n}{2}} + \sum_{\substack{1\leqslant j<j'\leqslant k\\ j,j'\neq i}} \frac{(r_{j'}+1)(r_j+1)}{\binom{n}{2}} \\
&= \frac{2}{n(n-1)}\Bigg(\sum_{\substack{1\leqslant j\leqslant k\\ j\neq i}} \frac{(r_j+2)(r_j+1)}{2} \\
&\quad + \frac{1}{2}\sum_{\substack{1\leqslant j\leqslant k\\ j\neq i}}\sum_{\substack{1\leqslant j'\leqslant k\\ j'\neq i}} (r_j+1)(r_{j'}+1) - \frac{1}{2}\sum_{\substack{1\leqslant j\leqslant k\\ j\neq i}} (r_j+1)^2\Bigg) \\
&= \frac{2}{n(n-1)}\cdot\Bigg(\frac{1}{2}\sum_{\substack{1\leqslant j\leqslant k\\ j\neq i}} (r_j^2+3r_j+2) \\
&\quad + \frac{1}{2}(n-r_i+k-3)^2 - \frac{1}{2}\sum_{\substack{1\leqslant j\leqslant k\\ j\neq i}} r_j^2 - \sum_{\substack{1\leqslant j\leqslant k\\ j\neq i}} r_j - \frac{k-1}{2}\Bigg) \\
&= \frac{(n-r_i+k-3)(n-r_i+k-2)}{n(n-1)}.
\end{aligned}$$

Thus,

$$\begin{aligned}\mathcal{S}um_3 &= \sum_{\substack{\boldsymbol{c}\in I_2(\boldsymbol{y})\\ d_L(\mathcal{D}_{EN}^{n-1}(\boldsymbol{y}),\boldsymbol{c})=3}} p(\boldsymbol{y}|\boldsymbol{c})\\ &= \sum_{\mathbf{c}\in\mathcal{C}_1} p(\boldsymbol{y}|\boldsymbol{c}) + \sum_{\mathbf{c}\in\mathcal{C}_2} p(\boldsymbol{y}|\boldsymbol{c}) + \sum_{\mathbf{c}\in\mathcal{C}_3} p(\boldsymbol{y}|\boldsymbol{c})\\ &= \frac{(n-k)(n-k+1)}{n(n-1)}\\ &\quad + \frac{(n-r_i+k-3)}{n(n-1)}\cdot(3n-k-r_i)\\ &= \frac{1}{n(n-1)}\cdot(4n^2-4nr_i-8n+r_i^2+3r_i+2k).\end{aligned}$$

It holds that $\mathcal{S}um_3 - \mathcal{S}um_1 \geqslant 0$ if and only if

$$4n^2-4nr_i-8n+r_i^2+3r_i+2k \geqslant 4n(r_i+1)-r_i^2-3r_i-2$$
$$2n^2-4nr_i-6n+r_i^2+3r_i+k+1 \geqslant 0.$$

■

Using this result we can explicitly define the ML* decoder $\mathcal{D}_{\mathsf{ML}^*}$. This decoder works as follows. For each word $\boldsymbol{y}$ it calculates the number of runs t and the length of the longest run r_i and then checks if

$$2n^2-4nr_i-6n+r_i^2+3r_i+k+1 \geqslant 0. \quad (5)$$

If this condition holds, the decoder works as the lazy decoder and simply returns the word $\boldsymbol{y}$. Otherwise, it acts like the embedding number decoder of length $n-1$ and prolongs the first longest run by one. This result is summarized in the following theorem.

Theorem 25. *The ML* decoder $\mathcal{D}_{\mathsf{ML}^*}$ for a single 2-deletion channel is a decoder that performs as the lazy decoder if inequality (5) holds and otherwise it acts like the embedding number decoder of length $n-1$. i.e.,*

$$\mathcal{D}_{\mathsf{ML}^*}(\boldsymbol{y}) = \begin{cases}\mathcal{D}_{Lazy}(\boldsymbol{y}) & \textit{inequality (5) holds ,}\\ \mathcal{D}_{EN}^{n-1}(\boldsymbol{y}) & \textit{otherwise.}\end{cases}$$

Proof: Using the previous lemmas, one can verify that $\mathcal{D}_{\mathsf{ML}^*}$ minimizes the average decoding error probability for any possible channel output $\boldsymbol{y}$ and hence it is the ML* decoder. ■

The result of Theorem 25 implies that if the ML* decoder chooses the same output as the decoder $\mathcal{D}_{EN}^{n-1}$ then inequality (5) does not hold. It can be shown that this implies that $r_i \geqslant (2-\sqrt{2})n$ and thus, by Claim 7, in almost all cases the output of the ML* decoder is the lazy decoder's output.

Appendix A

Claim 7. For all $n \geqslant 1$ it holds that $\tau((\Sigma_2)^n) \leqslant 2\log(n)$.

Proof: For $1 \leqslant r \leqslant n$, let $N(r)$ denote the number of words in Σ_2^n which the length of their maximal run is r. Note that $N(r) \leqslant n2^{n-r-1}$. This holds since we can set the location of the maximal run to start at some index i, which has less than n options. There are two options for the bit value in the maximal run, the two bits before and after the run are fixed and have to opposite to the bit value in the run, and the rest of the bits can be arbitrary. Then, it holds that

$$\begin{aligned}\tau((\Sigma_2)^n) &= \frac{\sum_{r=1}^n rN(r)}{2^n} = \frac{\sum_{r=1}^{\ell(n)} rN(r)}{2^n} + \frac{\sum_{r=\ell(n)+1}^n rN(r)}{2^n}\\ &\leqslant \frac{\sum_{r=1}^{\ell(n)} \ell(n)N(r)}{2^n} + \frac{\sum_{r=\ell(n)+1}^n rn2^{n-r-1}}{2^n}\\ &= \frac{\ell(n)\sum_{r=1}^{\ell(n)} N(r)}{2^n} + \frac{n2^{n-1}\sum_{r=\ell(n)+1}^n r2^{-r}}{2^n}\\ &\leqslant \frac{\ell(n)2^n}{2^n} + \frac{n2^{n-1}\cdot n2^{-\ell(n)-1}}{2^n} = \ell(n) + \frac{n^2}{2^{\ell(n)+2}}.\end{aligned}$$

Finally, by setting $\ell(n) = \lceil 2\log(n)\rceil - 2$ we get that

$$\begin{aligned}\tau((\Sigma_2)^n) &\leqslant \lceil 2\log(n)\rceil - 2 + \frac{n^2}{2^{\lceil 2\log(n)\rceil}}\\ &\leqslant \lceil 2\log(n)\rceil - 1 \leqslant 2\log(n).\end{aligned}$$

■

References

[1] M. Abroshan, R. Venkataramanan, L. Dolecek, and A. G. i Fàbregas. "Coding for deletion channels with multiple traces," *IEEE International Symposium on Information Theory*, pp. 1372-1376, 2019.

[2] A. Atashpendar, M. Beunardeau, A. Connolly, R. Géraud, D. Mestel, A. W. Roscoe, and P. Y. A. Ryan. "From clustering supersequences to entropy minimizing subsequences for single and double deletions," *arXiv preprint:1802.00703*, 2018.

[3] M. Blawat, K. Gaedke, I. Hütter, X.-M. Chen, B. Turczyk, S. Inverso, B.W. Pruitt, and G.M. Church, "Forward error correction for DNA data storage," *Int. Conf. on Computational Science*, vol. 80, pp. 1011–1022, 2016.

[4] J. Brakensiek, V. Guruswami, and S. Zbarsky, "Efficient low-redundancy codes for correcting multiple deletions," *Proc. of the Twenty-seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1884–1892, Philadelphia, PA, USA, 2016.

[5] J.A. Briffa, V. Buttigieg, and S. Wesemeyer, "Time-varying block codes for synchronization errors: MAP decoder and practical issues," arXiv:1802.08595, Feb. 2018.

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

[7] J. Castiglione and A. Kavcic, "Trellis based lower bounds on capacities of channels with synchronization errors, *Information Theory Workshop*, Jeju, South Korea, pp. 24–28, 2015.

[8] M. Cheraghchi, "Capacity upper bounds for deletion-type channels," *Journal of the ACM*, vol. 66, no. 2, p. 9, 2019.

[9] R. Con and A. Shpilka, "Explicit and efficient constructions of coding schemes for the binary deletion channel and the Poisson repeat channel," arXiv:1909.10177, Sep 2019.

[10] M. Dalai, "A new bound on the capacity of the binary deletion channel with high deletion probabilities," *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, pp. 499–502, Jun. 2011.

[11] S. Diggavi and M. Grossglauser, "On information transmission over a finite buffer channel," IEEE Transactions on Information Theory, vol. 52, no. 3, pp. 1226–1237, March 2006.

[12] L. Dolecek and V. Anantharam, "Using Reed?Muller RM $(1, m)$ codes over channels with synchronization and substitution errors,"' *IEEE Trans. on Inform. Theory*, vol. 53, no. 4, pp. 1430–1443, Apr. 2007.

[13] E. Drinea and M. Mitzenmacher, "Improved lower bounds for the capacity of iid deletion and duplication channels," IEEE Transactions on Information Theory, vol. 53, no. 8, pp. 2693–2714, 2007.

[14] C. Elzinga, S. Rahmann, and H. Wang. Algorithms for subsequence combinatorics. *Theoretical Computer Science*, 409(3):394–404, 2008.

[15] D. Fertonani and T. M. Duman, "Novel bounds on the capacity of the binary deletion channel," IEEE Transactions on Information Theory, vol. 56, no. 6, pp. 2753–2765, 2010.

[16] R. Gabrys and F. Sala, "Codes correcting two deletions," *IEEE Trans. on Inform. Theory*, vol. 65, no. 2, pp. 965–974, Feb. 2018.

[17] B. Graham. A Binary Deletion Channel With a Fixed Number of Deletions. Combinatorics, Probability and Computing, 24(3), 486-489.2018.

[18] V. Guruswami, B. Haeupler, and A. Shahrasbi, "Optimally resilient codes for list-decoding from insertions and deletions," *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 524–537, 2020.

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

[20] B. Haeupler, A. Shahrasbi, and M. Sudan, "Synchronization strings: List decoding for insertions and deletions," https://arxiv.org/abs/1802.08663, 2018.

[21] T. Hayashi and K. Yasunaga, "On the list decodability of insertions and deletions," *Int. Symp. Inform. Theory*, pp. 86–90, 2018.

[22] R. Heckel, G. Mikutis, and R.N. Grass, "A characterization of the DNA data storage channel," arxiv.org/pdf/1803.03322.pdf, 2018.

[23] S. Kas Hanna and S. El Rouayheb, "List decoding of deletions using guess & check codes," *Int. Symp. Inform. Theory*, pp. 2374–2378, 2019.

[24] A. Kirsch and E. Drinea, "Directly lower bounding the information capacity for channels with i.i.d. deletions and duplications," IEEE Transactions on Information Theory, vol. 56, no. 1, pp. 86–102, Jan. 2010.

[25] S. Kosuri and G.M. Church, "Large-scale de novo DNA synthesis: technologies and applications," *Nature Methods*, vol. 11, no. 5, pp. 499–507, May 2014.

[26] V. I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, 10(8):707–710, 1966.

[27] S. Liu, I. Tjuawinata, and C. Xing, "On list decoding of insertion and deletion errors," https://arxiv.org/abs/1906.09705, 2019.

[28] M. Mitzenmacher, "A survey of results for deletion channels and related synchronization channels," *Probability Surveys*, vol. 6, pp. 1–33, 2009.

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

[30] L. Organick, S.D. Ang, Y.-J. Chen, R. Lopez, S. Yekhanin, K. Makarychev, M.Z. Racz, G. Kamath, P. Gopalan, B. Nguyen, C. Takahashi, S. Newman, H.-Y. Parker, C. Rashtchian, K. Stewart, G. Gupta, R. Carlson, J. Mulligan, D. Carmean, G. Seelig, L. Ceze, and K. Strauss, "Scaling up DNA data storage and random access retrieval," *bioRxiv*, Mar. 2017.

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

[32] M.G. Ross, C. Russ, M. Costello, A. Hollinger, N.J. Lennon, R. Hegarty, N. Nusbaum, and D.B. Jaffe, "Characterizing and measuring bias in sequence data," *Genome Biol.*, vol. 14, no. 5, R51, 2013.

[33] O. Sabary, Y. Orlev, R. Shafir L. Anavy, E. Yaakobi, Z.Yakhini, "SOLQC: Synthetic oligo library quality control Tool," *bioRxiv*, Nov. 2019.

[34] F. Sala, C. Schoeny, N. Bitouzé, and L. Dolecek, "Synchronizing files from a large number of insertions and deletions," *IEEE Trans. on Comm.*, vol. 64, no. 6, pp. 2258–2273, June 2016.

[35] J. Sima and J. Bruck, "Optimal k-deletion correcting codes," *Proc. IEEE Int. Symp. Inf. Theory*, Paris, France, pp. 847–851, Jul. 2019.

[36] J. Sima, N. Raviv, and J. Bruck, "On coding over sliced information," http://arxiv.org/abs/1809.02716, 2018.

[37] S. R. Srinivasavaradhan, M. Du, S. Diggavi, and C. Fragouli. On maximum likelihood reconstruction over multiple deletion channels. In *IEEE International Symposium on Information Theory (ISIT)*, pages 436–440, 2018.

[38] I. Tal, H. D. Pfister, A. Fazeli and A. Vardy, "Polar codes for the deletion channel: weak and strong polarization," https://arxiv.org/abs/1904.13385.

[39] R. R. Varshamov and G. M. Tenenholtz. A code for correcting a single asymmetric error. *Automatica i Telemekhanika*, 26(2):288–292, 1965.

[40] R. Venkataramanan, S. Tatikonda, and K. Ramchandran, "Achievable rates for channels with deletions and insertions," IEEE Transactions on Information Theory, vol. 59, no. 11, pp. 6990–7013, 2013.

[41] A. Wachter-Zeh, "List decoding of insertions and deletions," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6297–6304, 2017.

[42] S. H. T. Yazdi, R. Gabrys, and O. Milenkovic, "Portable and error-free DNA-based data storage," *Scientific Reports*, vol. 7, sp. 5011, Jul. 2017.

[43] A.K.-Y. Yim, A.C.-S. Yu, J.-W. Li, A.I.-C. Wong, J.F.C. Loo, K.M. Chan, S.K. Kong, and T.-F. Chan, "The Eesential component in DNA-based information storage system: Robust error-uolerating module," *Frontiers in Bioengineering and Biotechnology* vol. 2, pp. 1–5, Nov. 2014.