mathematics

Article

Compression Sensitivity of the Burrows–Wheeler Transform and Its Bijective Variant

Hyodam Jeon and Dominik Köppl *

Department of Computer Science and Engineering, University of Yamanashi, Kōfu 400-8511, Japan
* Correspondence: dkppl@yamanashi.ac.jp

Abstract: The Burrows–Wheeler Transform (BWT) is a widely used reversible data compression method, forming the foundation of various compression algorithms and indexing structures. Prior research has analyzed the sensitivity of compression methods and repetitiveness measures to single-character edits, particularly in binary alphabets. However, the impact of such modifications on the compression efficiency of the bijective variant of BWT (BBWT) remains largely unexplored. This study extends previous work by examining the compression sensitivity of both BWT and BBWT when applied to larger alphabets, including alphabet reordering. We establish theoretical bounds on the increase in compression size due to character modifications in structured sequences such as Fibonacci words. Our devised lower bounds put the sensitivity of BBWT on the same scale as of BWT, with compression size changes exhibiting logarithmic multiplicative growth and square-root additive growth patterns depending on the edit type and the input data. These findings contribute to a deeper understanding of repetitiveness measures.

Keywords: lossless data compression; Burrows–Wheeler Transform (BWT); bijective BWT (BBWT); compression sensitivity; string transformations; Fibonacci words; Lyndon factorization; compression efficiency analysis

MSC: 68P30

check for updates

Academic Editor: Iliya Bouyukliev

Received: 27 February 2025
Revised: 21 March 2025
Accepted: 24 March 2025
Published: 25 March 2025

Citation: Jeon, H.; Köppl, D. Compression Sensitivity of the Burrows–Wheeler Transform and Its Bijective Variant. *Mathematics* **2025**, *13*, 1070. https://doi.org/10.3390/math13071070

1. Introduction

The Burrows–Wheeler transform (BWT) [1] has attracted great attention in interdisciplinary fields such as lossless data compression and text indexing. It lies at the heart of compression algorithms like bzip2 and text indexing data structures such as the FM-index [2]. By compressing single character runs of the BWT, we obtain a compressed but reversible transformation, which can be augmented with techniques akin to the FM-index to give rise to compressed text indices [3–8]. Because of its reversible nature, the BWT is also used in bioinformatics applications such as sequence alignment and genome assembly [9,10]. Workshops (e.g., [11,12]) and books (e.g., [13]) have been dedicated exclusively to the BWT and its applications.

Given word T of length n, the BWT of T is a permutation of its characters. In detail, we sort all cyclic conjugates of T lexicographically and concatenate the last characters of these conjugates to form the BWT of T. The BWT is a reversible transformation by application of the so-called Gessel–Reutenauer transform [14].

Among various variants of the BWT (e.g., [15–19]), the *bijective BWT* (BBWT) [20] can be considered as one of the well-perceived ones that is a word isomorphism. A word isomorphism maps a word to another word *injectively*, and each word is a unique image of another word. For instance, this is not the case for the BWT, whether we add additional information such as an artificial delimiter (known as the $ character) or a starting position, cf. [21–23].

In this article, we focus on the *run-length compression* of the BWT and the BBWT: run-length compression is usually the first step in the compression pipeline of the BWT and its variants. In addition, compressed text indices such as the r-index [4] store the BWT in a run-length compressed form. The run-length compression of word T is the number of maximal runs of equal characters in T. For instance, the word $\texttt{mississippi}$ can be written in an exponential notation as $\texttt{m}^1\texttt{i}^1\texttt{s}^2\texttt{i}^1\texttt{s}^2\texttt{i}^1\texttt{p}^2\texttt{i}^1$ and therefore has eight runs. We denote the run-length compression of word T by $\mathsf{runs}(T)$. Given word T, we define the following two repetitiveness measures:

- $r = r(T) = \mathsf{runs}(\mathrm{BWT}(T))$ and
- $\rho = \rho(T) = \mathsf{runs}(\mathrm{BBWT}(T))$.

In this article, we investigate the sensitivity of the BWT and the BBWT to single-character edits. This means that we analyze how the run-length compression of the BWT and the BBWT changes when we modify a single character of the input word. Previous research has shown that the run-length compression of the BWT is sensitive to single-character edits in binary alphabets [24]. Here, we extend this research to larger alphabets and analyze the sensitivity of the BBWT to single-character edits. Research on compression sensitivity is not a new topic, of which we are aware. We present following related work.

2. Related Work and Contribution

The *sensitivity* [25] of a repetitiveness measure m is the maximum difference in the sizes of $\mathtt{m}(T)$ for word T and for a single-character edited word T'. Sensitivity measures the robustness of a repetitiveness measure against small changes in the input word introduced by various sources of input (source code changes, biological sequencing errors, typos, etc.). Akagi et al. [25] reviewed known results that directly imply a sensitivity for repetitiveness measures such as for Lempel–Ziv 78 [26] or the BWT [24]. Additionally, they offered and improved upper and lower bounds on the multiplicative sensitivity of various compressors and measures including the Lempel–Ziv dictionary compressors [27,28] and the smallest string attractors [29].

In detail, for two words W_1 and W_2, we let $\mathsf{ed}(W_1, W_2)$ denote the edit distance between W_1 and W_2. We define the *additive sensitivity* $\mathsf{AS}_\mathtt{m}$ and *multiplicative sensitivity* $\mathsf{MS}_\mathtt{m}$ of a repetitiveness measure $\mathtt{m}$ by

- $\mathsf{AS}_\mathtt{m}(n) = \max_{W_1 \in \Sigma^n}\{\mathtt{m}(W_2) - \mathtt{m}(W_1) \mid W_2 \in \Sigma^* : \mathsf{ed}(W_1, W_2) = 1\}$, and
- $\mathsf{MS}_\mathtt{m}(n) = \max_{W_1 \in \Sigma^n}\left\{\frac{\mathtt{m}(W_2)}{\mathtt{m}(W_1)} \mid W_2 \in \Sigma^* : \mathsf{ed}(W_1, W_2) = 1\right\}$.

The sensitivity has been studied for lexparse [30] by Nakashima et al. [31] and for the size of the compact directed acyclic word graph [32] by Fujimaru et al. [33]. In particular, Giuliani et al. [24] showed that $\mathsf{MS}_r(n) = \Omega(\log n)$ and $\mathsf{AS}_r(n) = \Omega(\sqrt{\log n})$.

Our contribution. In this article, we show identical results for the BBWT, confirming that it is also sensitive to single-character edits. Concretely, we establish that $\mathsf{MS}_\rho(n) = \Omega(\log n)$ with Theorem 5 and $\mathsf{AS}_\rho(n) = \Omega(\sqrt{\log n})$ with Lemma 47. In detail, we obtain the asymptotically same results regarding $\mathsf{MS}_\rho(n)$:

- in Theorem 5 for deletion,
- in Theorem 6 and Theorem 7 for substituting a character with a smaller or larger one, respectively, and
- in Theorem 8 and Theorem 9 for insertion of $\texttt{a}$ or a strictly smaller character #, respectively.

We also obtain the asymptotically same results regarding $\mathsf{AS}_\rho(n)$:

- in Theorem 10 for deletion,
- in Theorem 12 for inserting a large character, and
- in Theorem 11 and Theorem 13 for substituting a character with a smaller or larger one, respectively.

Additionally, we broaden the study of the sensitivity of the BWT by allowing larger alphabets (Theorem 2) and alphabet reordering (Theorem 4), obtaining the same asymptotic complexities as reported by Giuliani et al. [24].

Since our major contribution is on the BBWT, we also briefly review known results related to it.

BBWT. Since its inception [20], the BBWT has been studied under various aspects. We are aware of construction algorithms (cf. [34] or [35] and the references therein), indexes [35] based on the BBWT, studies about the relationship of ρ and r [36], $\rho(T)$ and ρ of the reverse of T [37].

3. Preliminaries

In this section, we provide the necessary definitions and terminology used throughout the paper. A list of symbols is given in Table 1.

Words. We let Σ be a finite and ordered alphabet with cardinality σ. The elements of Σ are called *characters*. A *word* over Σ is a finite sequence $W = W[0]W[1]\cdots W[n-1] = W[0..n-1]$ of characters from Σ. The order of the alphabet induces the lexicographic order on words, which we also denote by $\prec_{\text{lex}}$.

We denote the length of W by $|W|$, with ε being the unique word of length 0. We denote the set of words of length n by Σ^n, and represent the set of all words on Σ by $\Sigma^* = \bigcup_{n\geq 0}\Sigma^n$. Given word $W = W[0..n-1]$, we define its *reverse* by $\text{rev}(W) = W[n-1]W[n-2]\cdots W[0]$. If $W = XYZ$ for words W, X, Y, Z, then X, Y, Z are, respectively, a prefix, a subword, and a suffix of W. We call word W' a *conjugate* of W if and only if there is integer $i \in [0..|W|-1]$ such that $W' = W[i..|W|]W[0..i-1]$. In this case, we write $W' = \text{conj}_i(W)$. In particular, $W = \text{conj}_0(W)$. We call word U a *circular factor* of word W if it is a prefix of $\text{conj}_i(W)$ for some $i \in [0..|W|-1]$; in this case, we call i (the starting position of) an *occurrence* of U. If we can express word W as $W = V^k$ for word V and integer $k \geq 2$, then we call W a *power*, otherwise we call W *primitive*. Finally, W is primitive if and only if it has $|W|$ distinct conjugates.

Given two words V, W, the *longest common prefix* of V and W, denoted $\text{lcp}(V, W)$, is the unique word U such that U is a prefix of both V and W, and $V[|U|] \neq W[|U|]$ if neither of the two words is a prefix of the other.

The Burrows–Wheeler Transform (BWT). We define the BWT of word W based on its conjugates. For that, we define two concepts, an order and a list of conjugates sorted in that order. First, the omega-order [16] of two words T and S as follows: $T \prec_\omega S$ if either $T^\omega \prec S^\omega$ or $T^\omega = S^\omega$ and $|T| < |S|$. Here, S^ω denotes the infinite word obtained by concatenating word S an infinite number of times. The omega-order coincides with the lexicographic order if neither of two words is a proper prefix of the other but may differ otherwise. Second, we let $\mathcal{M}(W)$ be the list of sorted conjugates of word W in omega-order.

Now, we can define the *Burrows–Wheeler Transform* (BWT) [1] of the word W, denoted by $\text{BWT}(W)$, as the word obtained by reading the last character of each conjugate in $\mathcal{M}(W)$.

For instance, the BWT of word `mississippi` is `pssmipissii`. By construction, it follows that W and W' are conjugates if and only if $\text{BWT}(W) = \text{BWT}(W')$. We denote by $r(W) = \mathsf{runs}(\mathrm{BWT}(W))$ the number of runs in the BWT of word W. For example, $r(\mathtt{mississippi}) = \mathsf{runs}(\mathtt{pssmipissii}) = 8$.

Table 1. Definitions of symbols introduced in this article.

Symbol	Meaning
r	run length of the BWT
ρ	run length of the bijective BWT
n	length
k	index
$\mathtt{\#}$	a character lexicographically smaller than $\mathtt{a}$
$\mathtt{c}$	a character lexicographically larger than $\mathtt{b}$
F_k	kth Fibonacci word
f_k	kth Fibonacci number
X_k	kth central word
L_k	kth Lyndon Fibonacci word
$F_k^\flat$	kth Fibonacci word deleting its last character
$L_k^\flat$	kth Fibonacci Lyndon word deleting its last character
P_k	$\mathtt{ab}^k\mathtt{aa}$
E_k	$\mathtt{ab}^k\mathtt{aba}^{k-2}$
Q_k	$\mathtt{ab}^k\mathtt{a}$
$Q_k^\flat$	$\mathtt{ab}^k$
W_k	$\left(\prod_{i=2}^{k-1} P_i E_i\right) Q_k$
$W_{2k}^\flat$	W_{2k} deleting its last character
$\overline{P_k}$	$\mathtt{ba}^k\mathtt{bb}$
$\overline{E_k}$	$\mathtt{ba}^k\mathtt{bab}^{k-2}$
$\overline{Q_k}$	$\mathtt{ba}^k\mathtt{b}$
$\overline{Q_k^\flat}$	$\mathtt{ba}^k$
$\overline{W_k}$	$\left(\prod_{i=2}^{k-1} \overline{P_i E_i}\right) \overline{Q_k}$
$\overline{W_k^\flat}$	W_k' deleting the last character
C_k	Lyndon word of W_k
$C_k^\flat$	C_k deleting its last character $\mathtt{b}$
D_k	$C_k^\flat$ deleting its last character $\mathtt{a}$
H_{k-1}	E_{k-1} changed into $\mathtt{ab}^{k-1}\mathtt{a}^{k-3}$
S_{k-1}	E_{k-1} changed into $\mathtt{ab}^{k-1}\mathtt{abca}^{k-3}$
R_{k-1}	E_{k-1} changed into $\mathtt{ab}^{k-1}\mathtt{aca}^{k-3}$
$\beta(W)$	subword of BWT(W) corresponding to the range of contiguous conjugates prefixed by W
$\beta'(W)$	subword of BWT(W) applied to a specific edit operation
α	$f_{2k-3} + f_{2k-5} + \cdots + f_3 + f_1$
$\mathcal{M}(W)$	the list of lexicographically sorted conjugates of word W

Lyndon Words. A word is called a Lyndon word if it is lexicographically strictly smaller than all of its conjugates [38]. In particular, a Lyndon word must be primitive. Each primitive word S has exactly one conjugate that is Lyndon. We denote this conjugate by $\mathsf{LynConj}(S)$ and call it the *Lyndon conjugate* of S. The *Lyndon factorization* [39] of word W is a unique factorization of W into Lyndon words. In detail, it decomposes word W into a list of Lyndon words $S_1^{e_1}, S_2^{e_2}, \ldots, S_m^{e_m}$ such that $W = S_1^{e_1} S_2^{e_2} \cdots S_m^{e_m}$, where $S_m \prec_{\text{lex}} S_{m-1} \prec_{\text{lex}} \cdots \prec_{\text{lex}} S_1$ and $e_i \geq 1$. By construction, word S is Lyndon if and only if its Lyndon factorization consists of only one factor, i.e., S itself. We denote the multiset of Lyndon factors in the Lyndon factorization of S by $\mathcal{L}(S)$. As an example, we consider $\mathsf{LynConj}(\mathtt{mississippi}) = \mathtt{imississipp}$. The Lyndon factorization of $\mathtt{mississippi}$ is $\mathtt{m} \cdot \mathtt{iss}^2 \cdot \mathtt{ipp} \cdot \mathtt{i}$. We have $\mathcal{L}(\mathtt{mississippi}) = \{\mathtt{m}, \mathtt{iss}, \mathtt{iss}, \mathtt{ipp}, \mathtt{i}\}$.

Bijective BWT (BBWT). The *Bijective BWT (BBWT)* [20] of word T is the word obtained by sorting all conjugates of the Lyndon factors in the multiset $\mathcal{L}(T)$ in ω-order and then concatenating the last character of each sorted conjugate. For example, the BBWT of the word `mississippi` is `ipssmpissii`. In this article, we denote $\rho(W)$ as the compression ratio of BBWT, which means $\rho(W) = \mathsf{runs}(\text{BBWT}(W))$. For instance, $\rho(\texttt{mississippi}) = \mathsf{runs}(\texttt{ipssmpissii}) = 8$.

Fibonacci Words. *Fibonacci* words are so-called *standard words* ([40], Section 10.1), which are defined as follows. $F_0 = \texttt{b}$, $F_1 = \texttt{a}$, $F_{k+1} = F_k F_{k-1}$, for every $k \geq 1$. For all $k \geq 0$, $|F_k| = f_k$, where $\{f_k\}_{k \geq 0}$ are the Fibonacci numbers $1, 1, 2, 3, 5, 8, 13, 21, \ldots$, defined by the recurrence $f_0 = f_1 = 1$, $f_{k+1} = f_k + f_{k-1}$, for $k \geq 1$. Since Fibonacci numbers grow exponentially in k, we have $k = \Theta(\log |F_k|)$. We also introduce so-called *central words* [41] X_k for $k \geq 2$, which are palindromes defined by equation $F_{2k} = X_{2k}\texttt{ab}$, $\quad F_{2k+1} = X_{2k+1}\texttt{ba}$ for all $k \geq 1$. The central words X_{2k} and X_{2k+1} are palindromes. In particular, $X_2 = \varepsilon$. The recursive structure of words X_{2k} and X_{2k+1} is also known [42]:

- $X_{2k} = X_{2k-1}\texttt{ba}X_{2k-2} = X_{2k-2}\texttt{ab}X_{2k-1}$ and
- $X_{2k+1} = X_{2k}\texttt{ab}x_{2k-1} = X_{2k-1}\texttt{ba}X_{2k}$.

We study Fibonacci words in this article because they have the minimal number of BWT runs among binary words. This is because Mantaci et al. [43] have shown that the BWT of a binary word has exactly two runs if and only if it is a conjugate of a standard word or a conjugate of a power of a standard word. Further, there is rich literature (e.g., [44–46]) about Fibonacci words and their rotations.

4. Multiplicative Sensitivity of r by $\Omega(\log n)$

As a startup, we follow the steps of (Giuliani et al. [24], Section 3), who studied a family of Fibonacci word-related words for which they could observe a multiplicative sensitivity of $\Theta(\log n)$ for the number of character runs in the BWT. We here show a similar result, but use a new character (#) instead of one already appearing in the binary word. To facilitate notation, we write $<$ for $<_\omega$ when sorting conjugates. We build our proofs on the insights from the following results from the literature.

Lemma 1 (Remark 11 from [16])**.** *All conjugates of a word have the same BWT.*

Lemma 2 (Proposition 4 of [24])**.** *We let $F_{2k}^\flat$ be a word that removes the last character of F_{2k}, then $r(F_{2k}^\flat) = 2k$.*

Lemma 3 (Lemma 7 of [24])**.** *$conj_{n-3}(F_{2k}^\flat)$ is the smallest conjugate in $\mathcal{M}(F_{2k}^\flat)$.*

Lemma 4. *We let $v \in \Sigma^*$ be a Lyndon word of $F_{2k}^\flat$ that contains at least two distinct characters and let # be a character that does not occur in v. Then, $r(v) \leq r(\#v) = r(v\#) \leq r(v) + 2$.*

Proof. We refer to $\mathrm{conj}_{n-3}(F_{2k}^\flat)$ from Lemma 3 as v here if only $0 \leq i, j \leq f_{2k} - 1$. The conjugates of v with index i and j are $\mathrm{conj}_i(v)$, $\mathrm{conj}_j(v)$, respectively. Also, we set the lexicographic order between two conjugates as $\mathrm{conj}_i(v) < \mathrm{conj}_j(v)$; thus, $v[i..|v|-1]v[0..i-1] < v[j..|v|-1]v[0..j-1]$. We prove this separately in two cases, where Figure 1 sketches the setting.

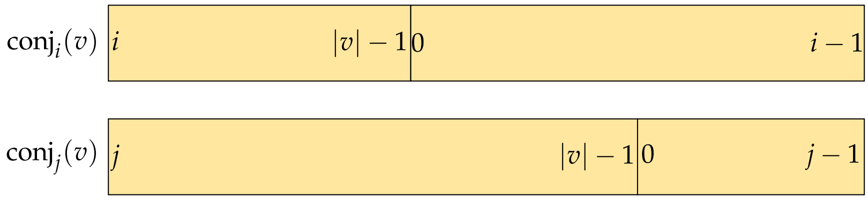

Figure 1. Sketch of the setting $\text{conj}_i(v) < \text{conj}_j(v)$ considered in the proof of Lemma 4.

Case 1: $|\text{lcp}(\text{conj}_i(v), \text{conj}_j(v))| < \min(|v| - i + 1, |v| - j + 1)$;
Case 2: $|\text{lcp}(\text{conj}_i(v), \text{conj}_j(v))| > \min(|v| - i + 1, |v| - j + 1)$.

The red rectangle in Figure 2 is an example of a common prefix of $\text{conj}_i(v)$ and $conj_j(v)$. In Case 1, it is $\text{conj}_i(v) < \text{conj}_j(v)$, meaning that the character of $\text{conj}_i(v)$ in position $|\text{lcp}(\text{conj}_i(v), \text{conj}_j(v))| + 1$ is smaller than in the one in the same position in $\text{conj}_j(v)$. Thus, inserting # in position $|v|$ does not change the lexicographic order between $\text{conj}_i(v)$ and $\text{conj}_j(v)$. The order is preserved.

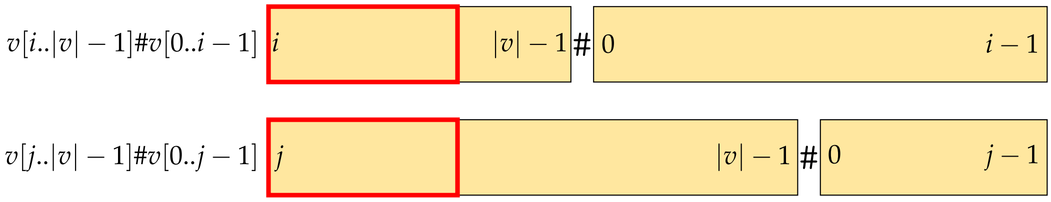

Figure 2. Illustration of the first case in Lemma 4. Inserting # does not change the lexicographic order between $\text{conj}_i(v)$ and $\text{conj}_j(v)$.

The red rectangle in Figure 3 depicting the longest common prefix of the two strings in question is longer than $|\text{lcp}(\text{conj}_i(v), \text{conj}_j(v))|$. In Case 2, it must be $i > j$, which means $|v[i..|v| - 1]| < |v[j..|v| - 1]|$. When it is $i < j$, then $|v[j..|v| - 1]| < |v[i..|v| - 1]|$, meaning that # appears first in $\text{conj}_j(v)$. As a result, $\text{conj}_j(v) < \text{conj}_i(v)$, which contradicts $\text{conj}_i(v) < \text{conj}_j(v)$. Thus, in Case 2, we only consider when it is $i > j$, as illustrated in Figure 3.

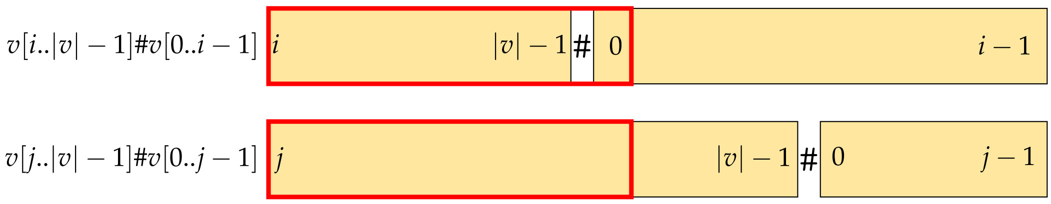

Figure 3. Illustration of the second case in Lemma 4.

Furthermore, we distinguish the second case between two subcases: We let u be unique circular factor which is smaller than all the other circular factors having the same length in v.

Case 2 (a): when u is a prefix of $v[i..|v| - 1]$;
Case 2 (b): when $v[0..i - 1]$ is a prefix of u.

When it is Case 2 (a), u appears only in the prefix of $v[0..i - 1]$. Thus, the first difference between $\text{conj}_i(v)$ and $\text{conj}_j(v)$ lies within the unique occurrence of u. The situation is depicted at Figure 4. After inserting the #, $\text{conj}_i(v)$ becomes $v[i..|v| - 1]\#v[0..i - 1]$, creating factor $\#u$ at position $|v| - i + 1$, which is not only unique but also smallest among other factors of length $|\#u|$ in v. Any factor that appears in the same position in $v[j..|v| - 1]\#u$ is greater than $\#u$. Thus, the order is preserved.

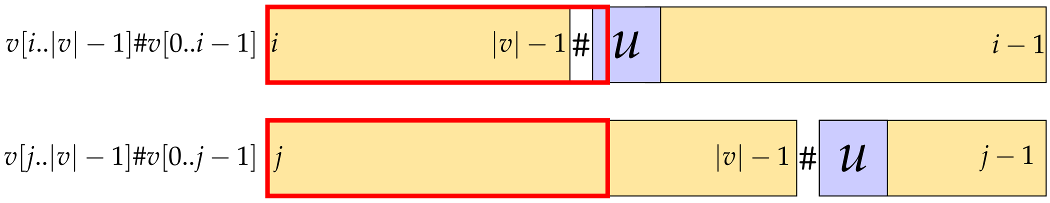

Figure 4. Illustration of Case 2 (a) in Lemma 4. Inserting # in does not affect lexicographic order between $\text{conj}_i(v)$ and $\text{conj}_j(v)$.

In Case 2 (a), u is the smallest prefix which appears only once in $v[0..i-1]$. v is a Lyndon word; thus, $v[0..i-1]$, it is also the smallest factor in v. However, in Case 2 (b), u is longer than $v[0..i-1]$. We sketch the situation in Figure 5, where we visualize u with a purple rectangle. Therefore, $v[0..i-1]$ must appear more than twice in u. If $v[0..i-1]$ appears only once, u is analogous with $v[0..i-1]$. Also, from $\text{conj}_i(v) \neq \text{conj}_j(v)$, there must be a difference in $v[0..i-1]$. Moreover, since v is primitive, v cannot be expressed in the form Z^k for word Z and a integer $k \geq 2$. The first distinct character between $\text{conj}_i(v)$ and $\text{conj}_j(v)$ is within $\text{conj}_i(v)[|v|-i+1..|v|-1]$. We assume otherwise that there is no mismatching character pair with $v[0..i-1]$ and the prefix of $\text{conj}_i(v)$, which is $v[i..2i-1]$. Since $v[0..i-1] = v[i..[2i-1]$, $\text{conj}_i(v)$ also has a smallest prefix and it contradicts with v, which is one and only Lyndon word. Moreover, $\text{conj}_i(v)$ becomes $v[0..i-1]v[i..2i-1]\ldots = v[0..i-1]^2\ldots$, thus contradicting its primitivity.

In this way, after inserting a #, the analogous behavior of Case 2 (a) is observed.

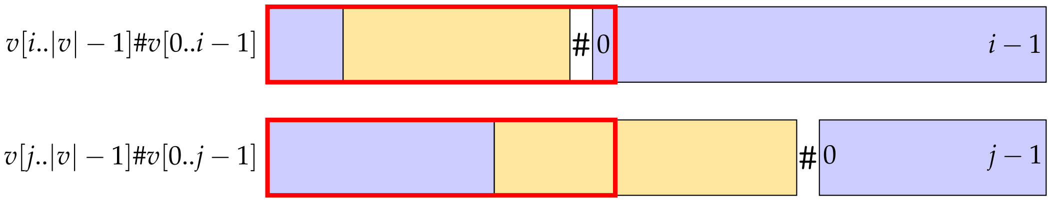

Figure 5. Illustration of Case 2 (b) in Lemma 4.

The order of original conjugates of v is preserved with respect to the original BWT according to the cases above. Thus, the only difference in inserting # in v occurs in conjugates of $\#v$ and $v\#$. On the one hand, we observe that $\#v$ is now the smallest among all conjugates of $\mathcal{M}(\#v)$, and it ends with the last character of v. On the other hand, $v\#$ becomes the second smallest conjugate and ends with #. Hence, we have $\text{BWT}(\#v) = \text{BWT}(v)[0] \cdot \# \cdot \text{BWT}(v)[1..|v|-1]$, which concludes the proof. □

Theorem 1. *We let F_{2k} be the Fibonacci word of even order $2k > 4$, and $f_{2k} = |F_{2k}|$. We let $F^{\flat}_{2k}$ be the word that results from substituting a* $\texttt{b}$ *by a # at position $f_{2k} - 1$. Then, $r(F^{\flat}_{2k}\#) = 2k + 2$.*

Proof. We let $S = F^{\flat}_{2k}\#$. From Lemma 2, $r(F^{\flat}_{2k}) = 2k$. And by Lemma 3, we know that $\text{conj}_{n-3}(F^{\flat}_{2k})$ is the smallest conjugate among $\mathcal{M}(F^{\flat}_{2k})$. By Lemma 4, we have $2k \leq r(\#\text{conj}_{n-3}(F^{\flat}_{2k})) \leq 2k+2$. More precisely, it is $2k+2$ since $\#\text{conj}_{n-3}(F^{\flat}_{2k})$ is the smallest conjugate in $\mathcal{M}(F^{\flat}_{2k}\#)$ and $\text{conj}_{n-3}(F^{\flat}_{2k})\#$ is the second smallest conjugate. The relative order among the conjugates of $\#\text{conj}_{n-3}(F^{\flat}_{2k})$ coincides with that of the conjugates of $F^{\flat}_{2k}$, using the same argument as in the proof of Lemma 4. This means that to obtain $\text{BWT}(S)$, it suffices to insert a # between the first two $\texttt{b}$s in $\text{BWT}(F^{\flat}_{2k})$. Since $r(\#\text{conj}_{n-3}(F^{\flat}_{2k})) = r(\#F^{\flat}_{2k}) = r(F^{\flat}_{2k}\#)$, we obtain the claim. □

5. Additive Sensitivity of r by $\Omega(\sqrt{n})$

In Section 4, we presented a word such that substituting one of its characters by `#`, which is strictly lexicographically smaller than all its characters, resulted in a logarithmic multiplicative increase in the number of runs r in the BWT. We now follow (Giuliani et al. [24], Section 4), who presented a family of words where a single edit can produce an additive increase of $\Theta(\sqrt{n})$ in r. Like before, we want to study the sensitivity when introducing a new character (#) in Section 5.1 or additionally when inverting the order of the alphabet in Section 5.2.

Definition 1. *For any $k > 5$, we let $P_k = \mathtt{ab}^k\mathtt{aa}$ and $E_k = \mathtt{ab}^k\mathtt{aba}^{k-2}$ for all $i \in [2..k-1]$, and $Q_k = \mathtt{ab}^k\mathtt{a}$. Then,*

$$W_k = \left(\prod_{i=2}^{k-1} P_i E_i\right) Q_k = \left(\prod_{i=2}^{k-1} \mathtt{ab}^i\mathtt{aaab}^i\mathtt{aba}^{i-2}\right)\mathtt{ab}^k\mathtt{a}. \tag{1}$$

The length of these words is

$$n = \sum_{i=2}^{k-1}(3i+4) + (k+2) = \frac{3k^2+7k}{2} - 9. \tag{2}$$

Thus, $k = \Theta(\sqrt{n})$. $W_k^\flat$ is

$$W_k^\flat = \left(\prod_{i=2}^{k-1} P_i E_i\right)\mathtt{ab}^k = \left(\prod_{i=2}^{k-1} \mathtt{ab}^i\mathtt{aaab}^i\mathtt{aba}^{i-2}\right)\mathtt{ab}^k. \tag{3}$$

We append `#`, which is lexicographically smaller than character `a` at the last part of $W_k^\flat$ and name the resulting word $W_k^\flat\#$.

Also, W_k, with its characters `a` and `b` swapped, is defined as $\overline{W_k}$, which is

$$\overline{W_k} = \left(\prod_{i=2}^{k-1} \overline{P_i E_i}\right)\overline{Q_k} = \left(\prod_{i=2}^{k-1} \mathtt{ba}^i\mathtt{bbba}^i\mathtt{bab}^{i-2}\right)\mathtt{ba}^k\mathtt{b}. \tag{4}$$

To characterize the BWT of words $W_k^\flat\#$, $\overline{W_k}$ and $\overline{W_k^\flat}\mathtt{c}$, we partition each of the BWT conjugates $\mathcal{M}(W_k^\flat\#)$, $\mathcal{M}(\overline{W_k})$, $\mathcal{M}(\overline{W_k^\flat}\mathtt{c})$ into distinct groups of consecutive conjugates having identical prefixes and define the subword of BWT(W_k) corresponding to each of these prefixes.

Given $X \in \Sigma^*$, we denote by $\beta(X, W_k)$ the subword of BWT(W_k) corresponding to the range of contiguous conjugates prefixed by X. We omit the second parameter of $\beta(X, W_k)$ when it is clear from the context. $\beta(X)$ is the concatenation of the last characters of conjugates with prefix X. For example, when X is `banana`, there are two conjugates starting with the prefix `an` which are `ananab` and `anaban`; thus, $\beta(\mathtt{an})$ of `banana` is `bn`.

Lemma 5. *In Proposition 28 of [24], it is already known that $r(W_k) = 6k - 12$.*

5.1. BWT of W_k After Substituting a Character

The lemmas presented below characterize the BWT of W_k after certain modifications have been applied. Rather than deriving the entire structure of the BWT from scratch, we analyze how replacing a character affects either the relative order or the final character of the conjugates of W_k. We let $\mathcal{M}(W_k^\flat\#)$ be the list of lexicographically sorted conjugates of the word $W_k^\flat\#$.

Lemma 6. $\beta(\texttt{\#}, W_k^\flat\texttt{\#}) = \texttt{b}$.

Proof. The first conjugate in $\mathcal{M}(W_k^\flat\texttt{\#})$ is $\texttt{\#}P_2\cdots\texttt{b}$. Since the lexicographic order of # is smaller than all other characters, a conjugate starting with # is smaller than every conjugate starting with a. # can be obtained by the last character of $W_k^\flat\texttt{\#}$, which is preceded by a b. □

Lemma 7. $\beta(\texttt{a}^i\texttt{b}, W_k^\flat\texttt{\#}) = \texttt{ba}^{k-i-2}$ *for all* $i \in [4..k-2]$.

Proof. Given integer $i \in [4..k-2]$, the conjugates of $\mathcal{M}(W_k^\flat\texttt{\#})$ starting with $\texttt{a}^i\texttt{b}$ are

$$\texttt{a}^{i-1}P_{i+2}\cdots\texttt{b} < \texttt{a}^{i-1}P_{i+3}\cdots\texttt{a} < \cdots < \texttt{a}^{i-1}P_{k-1}\cdots\texttt{a} < \texttt{a}^{i-1}Q_k^\flat\texttt{\#}\cdots\texttt{a}.$$

In $\mathcal{M}(W_k^\flat\texttt{\#})$, a prefix $\texttt{a}^i\texttt{b}$ can only be obtained by concatenation of the suffix $\texttt{a}^{i-2}$ of E_i, with the prefix ab of P_{i+1} or the prefix of ab of $Q_k^\flat\texttt{\#}$ if $i = k$. Note that all these conjugates end with an a, with the exception of the conjugate starting with $\texttt{a}^{i-1}P_{i+1}$, since this is where the unique occurrence of $\texttt{ba}^{i-1}\texttt{b}$ can be found. □

Lemma 8. $\beta(\texttt{aaab}, W_k^\flat\texttt{\#}) = \texttt{b}^5(\texttt{ab})^{k-6}\texttt{a}$.

Proof. The conjugates in $\mathcal{M}(W_k^\flat\texttt{\#})$ starting with aaab are

$$\begin{aligned}\texttt{aa}E_2\cdots\texttt{b} < \texttt{aa}E_3\cdots\texttt{b} < \texttt{aa}E_4\cdots\texttt{b} < \texttt{aa}P_5\cdots\texttt{b} < \texttt{aa}E_5\cdots\texttt{b}\\ < \texttt{aa}P_6\cdots\texttt{a} < \texttt{aa}E_6\cdots\texttt{b} < \cdots < \texttt{aa}P_{k-1}\cdots\texttt{a} < \texttt{aa}E_{k-1}\cdots\texttt{b}\\ < \texttt{aa}Q_k^\flat\texttt{\#}\cdots\texttt{a}.\end{aligned}$$

In $\mathcal{M}(W_k^\flat\texttt{\#})$, the conjugates that start with aaab can be obtained for all $i \in [4..k-1]$ from the concatenation of the suffix aa from E_i with P_{i+1} or with $Q_k^\flat\texttt{\#}$ if $i = k$. If $i \in [2..k-1]$, concatenation of the suffix aa of P_i with the prefix ab of E_i also makes aaab. Also, we can sort the conjugates with following order: $\bigcup_{i=2}^{4}\{\texttt{aa}E_i\} \cup \bigcup_{i=5}^{k-1}\{\texttt{aa}P_i\texttt{aa}E_i\} \cup \{\texttt{aa}Q_k^\flat\texttt{\#}\}$. All conjugates of $\texttt{aa}E_i$ end with a b and if $i \in [5..k-2]$, aa of E_i concatenated with P_{i+1} or $Q_k^\flat\texttt{\#}$ if $i = j$ also ends with a. On the other hand, $\texttt{aa}P_5$ ends with b. □

Lemma 9. $\beta(\texttt{aab}, W_k^\flat\texttt{\#}) = \texttt{aaba}^{2k-8}$.

Proof. The conjugates in $\mathcal{M}(W_k^\flat\texttt{\#})$ starting with aab are

$$\begin{aligned}\texttt{a}E_2\cdots\texttt{a} < \texttt{a}E_3\cdots\texttt{a} < \texttt{a}P_4\cdots\texttt{b}\\ < \texttt{a}E_4\cdots\texttt{a} < \texttt{a}P_5\cdots\texttt{a} < \texttt{a}E_5\cdots\texttt{a} < \cdots < \texttt{a}P_{k-1}\cdots\texttt{a} < \texttt{a}E_{k-1}\cdots\texttt{a} < \texttt{a}Q_k^\flat\texttt{\#}\cdots\texttt{a}.\end{aligned}$$

Each of the conjugates starting with aaab from Lemma 8 induces a conjugate starting with aab, obtained by shifting on the left one character a . It follows that all of these conjugates end with an a. The other conjugates that start with aab are those obtained by concatenating the suffix a of E_3 with the prefix ab of P_4 which ends with b. □

Lemma 10. $\beta(\texttt{ab}, W_k^\flat\texttt{\#}) = \texttt{b}^{k-2}\texttt{\#aba}^{2k-6}$.

Proof. The conjugates in $\mathcal{M}(W_k^\flat\texttt{\#})$ starting with the prefix ab are

$$\begin{aligned}\texttt{aba}^{k-3}Q_k^\flat\texttt{\#}\cdots\texttt{b} < \texttt{aba}^{k-4}P_{k-1}\cdots\texttt{b} < \cdots < \texttt{ab}P_3\cdots\texttt{b}\\ < P_2\cdots\texttt{\#} < E_2\cdots\texttt{a} < P_3\cdots\texttt{b}\\ < E_3\cdots\texttt{a} < P_4\cdots\texttt{a} < E_4\cdots\texttt{a} < \cdots < P_{k-1}\cdots\texttt{a} < E_{k-1}\cdots\texttt{a} < Q_k^\flat\texttt{\#}\cdots\texttt{a}.\end{aligned}$$

For all two distinct integers i, i' with $i > i' \geq 0$, we have $\mathtt{aba}^i\mathtt{b} < \mathtt{aba}^{i'}\mathtt{b}$. Thus, the first conjugate in the lexicographic order starting with ab is the one followed by the longest a. The smallest of these conjugates can be found from the suffix $\mathtt{aba}^{k-3}\mathtt{b}$ of E_{k-1}, followed by the suffix $\mathtt{aba}^{i-2}\mathtt{b}$ of E_i for all $2 \leq i \leq k-2$ taken in decreasing order.

By construction of E_i, for all $2 \leq i \leq k-1$, these conjugates must end in a b. The remaining conjugates starting with ab are exactly those of either P_i or E_i, for all $2 \leq i \leq k-1$, or $Q_k^\flat$#. The conjugates can be obtained by shifting on the left one character a from the conjugates starting with aab from Lemma 9, with the exception of one starting with P_3 since it ends with a b, and the other starting with P_2 which ends with #, while the other conjugates end with an a. □

Lemma 11. *$\beta(\mathtt{b}^i\#, W_k^\flat\#) = \mathtt{b}$ for all $1 \leq i \leq k-1$.*

Proof. The conjugate in $\mathcal{M}(W_k^\flat\#)$ starting with $\mathtt{b}^i\#$ for all $1 \leq i \leq k-1$ is $\mathtt{b}^i\#P_2\cdots\mathtt{b}$. This conjugate can be obtained by a suffix of $Q_k^\flat$#, and is always preceded by a b. □

Lemma 12. $\beta(\mathtt{ba}, W_k^\flat\#) = \mathtt{a}^{k-5}\mathtt{bbbab}^{k-5}\mathtt{ab}^{k-2}\mathtt{a}$.

Proof. The conjugates in $\mathcal{M}(W_k^\flat\#)$ starting with ba are

$$\begin{aligned}\mathtt{ba}^{k-3}Q_k^\flat\#\cdots\mathtt{a} < \mathtt{ba}^{k-4}P_{k-1}\cdots\mathtt{a} < \cdots < \mathtt{ba}^3P_6\cdots\mathtt{a}&\\ < \mathtt{baa}E_2\cdots\mathtt{b} < \mathtt{baa}E_3\cdots\mathtt{b} < \mathtt{baa}E_4\cdots\mathtt{b} < \mathtt{baa}P_5\cdots\mathtt{a}&\\ < \mathtt{baa}E_5\cdots\mathtt{b} < \mathtt{baa}E_6\cdots\mathtt{b} < \cdots < \mathtt{baa}E_{k-1}\cdots\mathtt{b} < \mathtt{ba}P_4\cdots\mathtt{a}&\\ < \mathtt{baba}^{k-3}Q_k^\flat\#\cdots\mathtt{b} < \mathtt{baba}^{k-4}P_{k-1}\cdots\mathtt{b} < \cdots < \mathtt{bab}P_3\cdots\mathtt{b} < \mathtt{babbbaa}\cdots\mathtt{a}.&\end{aligned}$$

We have as many circular occurrences of ba as the number of maximal character runs of b in $W_k^\flat$#. Then, for all $2 \leq i \leq k-1$,

Case 1: one run of b in P_i and
Case 2: two runs in E_i .

For Case 1, we have one conjugate starting with $\mathtt{baa}E_i$ for each $i \in [2..k-1]$. Since each run of b within each word of P_i is of length of at least 2, all conjugates in (1) end with b. For Case 2, for all $i \in [2..k-1]$ we can distinguish between two subcases, based on where ba starts:

Case 2 (a): from the first run of b in E_i, which is $\mathtt{baba}^{i-2}P_{i+1}$ when $i \in [2..k-2]$ or $\mathtt{baba}^{k-3}Q_k^\flat\#$ if $i = k-1$. Since b has at least 2 runs, conjugates with prefix (2.1) always end with b.

Case 2 (b): from run $\mathtt{ba}^{i-3}P_{i+1}$ for all $i \in [2..k-2]$, and $\mathtt{ba}^{k-3}Q_k^\flat\#$. Each conjugate of Case 2 (b) is obtained by shifting two characters to the right in each conjugate in Case 2 (a). Therefore, these conjugates end with an a.

Observe that only for Case 2 (b) we have conjugates starting with baaaa. Hence, the first conjugate in the lexicographic order is the one starting with $\mathtt{ba}^{k-3}Q_k^\flat\#$, followed by those starting with $\mathtt{ba}^{k-4}P_{k-1} < \mathtt{ba}^{k-5}P_{k-2} < \cdots < \mathtt{baaa}P_6$.

Among the remaining conjugates, those that have the prefix baaa start with $\mathtt{baa}P_5$ from Case 2 (b) or $\mathtt{baa}E_i$ from Case 2 (a). Thus, we can sort them according to lexicographic order. Then, the remaining conjugates, which start with baa, are obtained by $\mathtt{ba}P_3$ only. Finally, let us focus on the conjugates from Case 2 (a), which start with ba. These conjugates are sorted according to the length of the runs of as following the common prefix bab, similarly to the sorting of conjugates from Case 2 (b). The last conjugate left is the one starting with $\mathtt{b}P_3$ from Case 2 (b). Since $\mathtt{b}P_3$ is lexicographically greater than all other cases, this is the greatest conjugate of $W_k^\flat$# starting with ba and we can conclude our claim. □

Lemma 13. $\beta(\mathtt{b}^j\mathtt{a}, W_k^\flat\#) = \mathtt{ab}^{2k-2j-2}\mathtt{a}$ *for all* $2 \le j \le k-2$.

Proof. The conjugates starting with $\mathtt{ba}^i$ with integer $2 \le j \le k-2$ in $\mathcal{M}(W_k^\flat\#)$ are

$$\begin{aligned} \mathtt{b}^i\mathtt{aa}E_i\cdots\mathtt{a} &< \mathtt{b}^i\mathtt{aa}E_{i+1}\cdots\mathtt{b} < \cdots < \mathtt{b}^i\mathtt{aa}E_{k-1}\cdots\mathtt{b} \\ &< \mathtt{b}^i\mathtt{aba}^{k-3}Q_k^\flat\#\cdots\mathtt{b} < \mathtt{b}^i\mathtt{aba}^{k-4}P_{k-1}\cdots\mathtt{b} < \cdots < \mathtt{b}^i\mathtt{aba}^{i-1}P_{i+2}\cdots\mathtt{b} \\ &< \mathtt{b}^i\mathtt{aba}^{i-2}P_{i+1}\cdots\mathtt{a}. \end{aligned}$$

All runs of b of length at least $2 \le i \le k-2$ appear in either

Case 1: P_i or
Case 2: E_i for all $i \le j \le k-1$.

Let us consider these two cases separately. For all $i \le j \le k-1$, the conjugate starting within P_j has prefix $\mathtt{b}^i\mathtt{aa}E_j$. For all $i \le j \le k-2$, the conjugate starting within E_j has prefix $\mathtt{b}^i\mathtt{aba}^{j-2}P_{j+1}$, and for $j = k-1$, we have the conjugate with prefix $\mathtt{b}^i\mathtt{aba}^{k-3}Q_k^\flat\#$. By construction, we have all the conjugates from Case 1 sorted according to the lexicographic order of the words with respect to the length of the run by b obtained by E_j.

The conjugates covered by Case 2 are sorted according to the decreasing length of the run of a, following the common prefix $\mathtt{b}^i\mathtt{ab}$. Only when the run of b is exactly i long, its conjugate ends with a. Thus, the conjugates ending with an a are those starting with P_i and E_i, which have prefixes $\mathtt{b}^i\mathtt{aa}E_i$ and $\mathtt{b}^i\mathtt{aba}^{i-2}P_{i+1}$. □

Lemma 14. $\beta(\mathtt{b}^{k-1}\mathtt{a}, W_k^\flat\#) = \mathtt{aa}$.

Proof. The *two* conjugates in $\mathcal{M}(W_k^\flat\#)$ which start with $\mathtt{ba}^{k-1}\mathtt{a}$ are

$$\mathtt{b}^{k-1}\mathtt{aa}E_{k-1}\cdots\mathtt{a} < \mathtt{b}^{k-1}\mathtt{aba}^{k-3}Q_k^\flat\#\cdots\mathtt{a}.$$

The conjugates with the prefix $\mathtt{b}^{k-1}\mathtt{a}$ start with E_{k-1} or $Q_k^\flat\#$. These conjugates have prefixes of $\mathtt{b}^{k-1}\mathtt{aa}E_{k-1}$ and $\mathtt{b}^{k-1}\mathtt{aba}Q_k^\flat\#$, respectively. One can see that these conjugates taken in this order are already sorted, and both conjugates end with a. □

Lemma 15. $\beta(\mathtt{b}^k\#, W_k^\flat\#) = \mathtt{a}$.

Proof. The last conjugate in $\mathcal{M}(W_k^\flat\#)$ is $\mathtt{b}^k\#P_2\cdots\mathtt{a}$. The last conjugate in lexicographic order starts with $\mathtt{b}^k\#P_2$, and since the run of b is maximal, it ends with a, and the claim follows. □

In conclusion, we define the above theorem.

Theorem 2. $r(W_k^\flat\#) - r(W_k) = 2k-5$ *for every* $k \ge 6$.

Proof. The BWT of the $W_b^k\#$ is $\mathrm{BWT}(W_b^k\#) = \beta(\#)\prod_{i=2}^{k-1}\beta(\mathtt{a}^{k-i}\mathtt{b})\cdot\prod_{i=1}^{k-1}\beta(\mathtt{b}^i\#)\beta(\mathtt{b}^i\mathtt{a})\cdot\beta(\mathtt{b}^k\mathtt{a})$. We refer to Table 2. Moreover, $r(W_k^\flat\#) = 8k-17$ which has $2k-5$ more runs than $r(W_k) = 6k-12$, cf. Lemma 5.

The lexicographic order of # is lower than an a, and a conjugate starting with # is smaller than any conjugate starting with a. Moreover, every conjugate in $\beta(\mathtt{a}^i\mathtt{b})$ is smaller than every one in $\beta(\mathtt{a}^{i'}\mathtt{b})$, for every $1 \le i' \le i \le k-2$. In addition, every conjugate contributing a character to $\beta(\mathtt{b}^j\mathtt{a})$ is smaller than a conjugate contributing a character to $\beta(\mathtt{b}^{j'}\mathtt{a})$ for every $1 \le j \le j' \le k-1$. And with a conjugate starting with $b^i\#$, the number is smaller than that of $\mathtt{b}^i\mathtt{a}$. Since we considered all the disjoint ranges of conjugates of $W_k^\flat\#$ based on their common prefix, the word $\mathrm{BWT}(W_k^\flat\#)$ is

$\beta(\#)\prod_{i=2}^{k-1}\beta(\mathtt{a}^{k-i}\mathtt{b})\cdot\prod_{i=1}^{k-1}\beta(\mathtt{b}^{i}\#)\beta(\mathtt{b}^{i}\mathtt{a})\cdot\beta(\mathtt{b}^{k}\mathtt{a})$. With the structure of BWT($W_k^\flat\#$), we can derive its number of runs. The words $\beta(\#)$ and $\prod_{i=2}^{k-4}\beta(\mathtt{a}^{k-i}\mathtt{b})$ have $2(k-6)$ runs: we start with 1 run from $\beta(\#) = \mathtt{b}$ which is merged by $\beta(\mathtt{a}^{k-2}\mathtt{b})\beta(\mathtt{a}^{k-3}\mathtt{b}) = \mathtt{bba}$. And concatenating them $\beta(\mathtt{a}^{i}\mathtt{b})$ up to $\beta(\mathtt{a}^{4}\mathtt{b})$ adds 2 new runs each. $\beta(\mathtt{aaab})$, $\beta(\mathtt{aab})$, $\beta(\mathtt{ab})$ have $2(k-5)$, 3, 5 runs, respectively. However, the boundaries between $\beta(\mathtt{aaab})$ and $\beta(\mathtt{aab})$ are merged by an $\mathtt{a}$; therefore, $\beta(\mathtt{aab})$ has 2 runs. $\beta(\mathtt{b}\#)$ has 1 run, followed by $\beta(\mathtt{ba})$ which makes 7 runs. Then, $\beta(\mathtt{b}^{i}\#)$ and $\beta(\mathtt{b}^{i}\mathtt{a})$ repeat, making 1 and 3 runs until $i = k-2$ thus makes $4(k-3)$ runs. $\beta(\mathtt{b}^{k-1}\#)$ adds 1 run. Also, $\beta(\mathtt{b}^{k-1}\mathtt{a})$ adds 1 run and is the last run since $\beta(\mathtt{b}^{k}\#)$ does not add new runs, since it consists only of a $\mathtt{a}$ that merges with the previous one. Altogether, we have $2(k-6)+2(k-5)+2+5+1+7+4(k-3)+1+1 = 8k-17$, and the claim holds. The main difference in the runs of $W_k^\flat\#$ and W_k occurs from the prefix beginning with $\mathtt{b}^{i}\#$ that concatenates with $\mathtt{b}^{i}\mathtt{a}$, repeating $\mathtt{baba}$ for $i \in [2..k-1]$, while W_k repeats only $\mathtt{ba}$. Thus, it makes additive runs of $2k-5 = \Theta(k) = \Theta(\sqrt{n})$.

Table 2. Classification of the number of runs obtain in Theorem 2. The total number of runs is $8k-17$.

BWT of $W_k^\flat\#$	**Runs**
$\beta(\#) = \mathtt{b}$	1
$\beta(\mathtt{a}^{i}\mathtt{b}) = \mathtt{ba}^{k-i-2}$ for all $4 \le i \le k-2$	$2k-11$ but, when merged, $2k-12$
$\beta(\mathtt{aaab}) = \mathtt{b}^{5}(\mathtt{ab})^{k-6}\mathtt{a}$	$2k-10$ but, when merged, $2k-11$
$\beta(\mathtt{aab}) = \mathtt{aaba}^{2k-8}$	3 but, when merged, 2
$\beta(\mathtt{ab}) = \mathtt{b}^{k-2}\#\mathtt{aba}^{2k-6}$	5
$\beta(\mathtt{b}^{i}\#) = \mathtt{b}$ for all $i \in [1..k-1]$	$k-1$
$\beta(\mathtt{ba}) = \mathtt{a}^{k-5}\mathtt{bbbab}^{k-5}\mathtt{ab}^{k-2}\mathtt{a}$	7
$\beta(\mathtt{b}^{j}\mathtt{a}) = \mathtt{ab}^{2(k-j-1)}\mathtt{a}$ for all $j \in [2..k-2]$	$3(k-3)$
$\beta(\mathtt{b}^{k-1}\mathtt{a}) = \mathtt{aa}$	1
$\beta(\mathtt{b}^{k}\#) = \mathtt{a}$	1 but, when merged, 0

Tables 3–7 depict $\mathcal{M}(W_k^\flat\#)$. The first column partitions conjugates by common prefixes and names the common prefix shared by all conjugates in a partition. The second column shows the remaining part of the respective conjugate followed by the prefix of its partition. The remaining part of a conjugate decides its relative order inside its partition. The BWT column shows the last character of each conjugate. □

Table 3. Lexicographically sorted conjugates of $W_k^\flat\#$ studied in Theorem 2, Part 1.

Prefix	**Remaining Part**	**BWT**
#	P_2	b
$\mathtt{a}^{k-2}\mathtt{b}$	$\mathtt{b}^{k-1}\#$	b
$\mathtt{a}^{k-3}\mathtt{b}$	$\mathtt{b}^{k-2}\mathtt{aa}$ $\mathtt{b}^{k-1}\#$	b a
$\mathtt{a}^{k-4}\mathtt{b}$	$\mathtt{b}^{k-3}\mathtt{aa}$ $\mathtt{b}^{k-2}\mathtt{aa}$ $\mathtt{b}^{k-1}\#$	b a a
...	...	...

Table 4. Lexicographically sorted conjugates of $W_k^\flat$# studied in Theorem 2, Part 2.

Prefix	Remaining Part	BWT
a^3b	bab	b
	bbaba	b
	bbbabaa	b
	b^4aa	b
	b^4aba^3	b
	b^5aa	a
	b^5aba^4	b
	b^6aa	a
	b^6aba^5	b
	...	...
	$b^{k-2}aa$	a
	$b^{k-2}aba^{k-3}$	b
	b^{k-1}#	a
a^2b	bab	a
	bbaba	a
	bbbaa	b
	bbbabaa	a
	b^4aa	a
	b^4aba^3	a
	...	...
	$b^{k-2}aa$	a
	$b^{k-2}aba^{k-3}$	a
	b^{k-1}#	a

Table 5. Lexicographically sorted conjugates of $W_k^\flat$# studied in Theorem 2, Part 3.

Prefix	Remaining Part	BWT
ab	$a^{k-3}Q_k^\flat$#	b
	$a^{k-4}P_{k-1}$	b
	...	...
	P_3	b
	baa	#
	bab	a
	bbaa	b
	bbaba	a
	b^3aa	a
	b^3aba^2	a
	...	...
	$b^{k-2}aa$	a
	$b^{k-2}aba^{k-3}$	a
	b^{k-1}#	a
b#	P_2	b
ba	$a^{k-4}Q_k^\flat$#	a
	$a^{k-5}P_{k-1}$	a
	...	...
	a^2P_6	a
	b^2ab	b
	b^3aba	b
	b^4abaa	b
	aab^5aa	a
	aab^5aba^3	b
	...	...
	$aab^{k-1}aba^{k-3}$	b
	P_4	a
	bP_3	b
	...	...
	$ba^{k-3}Q_k^\flat$#	b
	b^3aa	a

Table 6. Lexicographically sorted conjugates of $W_k^\flat\#$ studied in Theorem 2, Part 4.

Prefix	Remaining Part	BWT
bb#	P_2	b
$\mathtt{b}^2\mathtt{a}$	$\mathtt{aab}^2\mathtt{ab}$	a
	$\mathtt{aab}^3\mathtt{aba}$	b
	...	...
	$\mathtt{aab}^{k-1}\mathtt{aba}^{k-3}$	b
	$\mathtt{ba}^{k-3}Q_k^\flat\#$	b
	...	...
	$\mathtt{ba}P_4$	b
	$\mathtt{b}P_3$	a
bbb#	P_2	b
$\mathtt{b}^3\mathtt{a}$	$\mathtt{aab}^3\mathtt{aba}$	a
	$\mathtt{aab}^4\mathtt{abaa}$	b
	...	...
	$\mathtt{aab}^{k-1}\mathtt{aba}^{k-3}$	b
	$\mathtt{ba}^{k-3}Q_k^\flat\#$	b
	...	...
	$\mathtt{baa}P_5$	b
	$\mathtt{ba}P_4$	a
bbbb#	P_2	b
...	...	...

Table 7. Lexicographically sorted conjugates of $W_k^\flat\#$ studied in Theorem 2, Part 5.

Prefix	Remaining Part	BWT
$\mathtt{b}^{k-2}\#$	P_2	b
$\mathtt{b}^{k-2}\mathtt{a}$	$\mathtt{aab}^{k-2}\mathtt{aba}^{k-4}$	a
	$\mathtt{aab}^{k-1}\mathtt{aba}^{k-3}$	b
	$\mathtt{ba}^{k-3}Q_k^\flat\#$	b
	$\mathtt{ba}^{k-4}P_{k-1}$	a
$\mathtt{b}^{k-1}\#$	P_2	b
$\mathtt{b}^{k-2}\mathtt{a}$	$\mathtt{aab}^{k-1}\mathtt{aba}^{k-3}$	a
	$\mathtt{ba}^{k-3}Q_k^\flat\#$	a
$\mathtt{b}^{k}\#$	P_2	a

5.2. BWT of $\overline{W_k}$ After Substituting a Character

In this subsection, we consider the word $\overline{W_k} = \left(\prod_{i=2}^{k-1} \overline{P_i E_i}\right)\overline{Q_k} = \left(\prod_{i=2}^{k-1} \mathtt{ba}^i\mathtt{bbba}^i\mathtt{bab}^{i-2}\right)\mathtt{ba}^i\mathtt{b}$, where we swapped a with b in W_k. The following series of lemmas characterize the subword of BWT($\overline{W_k}$) using $\mathcal{M}(\overline{W_k})$ for each range we consider.

Lemma 16. $\beta(\mathtt{a}^k\mathtt{b}, \overline{W_k}) = \mathtt{b}$.

Proof. The first conjugate in $\mathcal{M}(\overline{W_k})$ is $\mathtt{a}^k\mathtt{b}\cdots\mathtt{b}$. The first conjugate in lexicographic order must start with the longest run of as. By the definition of $\overline{W_k}$, the longest run of a has length k, which is obtained by $\mathtt{a}^k$ of $\overline{Q_k}$, which is preceded by a b. □

Lemma 17. $\beta(\mathtt{a}^i\mathtt{b}, \overline{W_k}) = \mathtt{ba}^{2k-2i-1}\mathtt{b}$ *for all* $i \in [2..k-1]$.

Proof. With integer $i \in [2..k-1]$, the conjugates starting with $\mathtt{a}^i\mathtt{b}$ in $\mathcal{M}(\overline{W_k})$ are

$$\begin{aligned} \mathtt{a}^i\mathtt{bab}^{i-2}\cdots\mathtt{b} &< \mathtt{a}^i\mathtt{bab}^{i-1}\cdots\mathtt{a} < \cdots < \mathtt{a}^i\mathtt{bab}^{k-3}\cdots\mathtt{a} < \mathtt{a}^i\mathtt{b}\overline{P_2}\cdots\mathtt{a} \\ &< \mathtt{a}^i\mathtt{bbba}^{k-1}\cdots\mathtt{a} < \cdots < \mathtt{a}^i\mathtt{bbba}^{i+1}\cdots\mathtt{a} \\ &< \mathtt{a}^i\mathtt{bbba}^{i}\cdots\mathtt{b}. \end{aligned}$$

For all $i \in [2..k-1]$, the factor of $\mathtt{a}^i\mathtt{b}$ can only be obtained for all $j \in [i..k-1]$, from $\mathtt{a}^i\mathtt{bab}^{j-2}$ from $\overline{E_j}$, or $\mathtt{a}^i\mathtt{bb}$ from $\overline{P_j}$, and if $j = k$, $\mathtt{a}^i\mathtt{b}$ from $\overline{Q_k}$. We can sort the conjugate according to the lexicographic order. Note that all these conjugates end with $\mathtt{b}$, with the exception of the conjugate starting with $\mathtt{a}^i\mathtt{b}$ obtained by $\overline{E_i}$ and $\overline{P_i}$ ending with $\mathtt{b}$. □

Lemma 18. $\beta(\mathtt{ab}, \overline{W_k}) = \mathtt{ba}^{k-2}\mathtt{baa}^{k-5}\mathtt{baaab}^{k-5}$.

Proof. In $\mathcal{M}(\overline{W_k})$, the conjugates starting with $\mathtt{ab}$ are

$$\begin{aligned} \mathtt{abaaabb}\overline{E_2}\cdots\mathtt{b} &< \mathtt{aba}\overline{P_3}\cdots\mathtt{a} < \mathtt{abab}\overline{P_4}\cdots\mathtt{a} < \cdots < \mathtt{abab}^{k-3}\overline{Q_k}\cdots\mathtt{a} \\ &< \mathtt{ab}\overline{P_4}\cdots\mathtt{b} < \mathtt{ab}\overline{P_2}\cdots\mathtt{a} < \mathtt{abb}\overline{E_{k-1}}\cdots\mathtt{a} < \cdots < \mathtt{abb}\overline{E_5}\cdots\mathtt{a} \\ &< \mathtt{abb}\overline{P_5} < \mathtt{abb}\overline{E_4}\cdots\mathtt{a} < \mathtt{abb}\overline{E_3}\cdots\mathtt{a} < \mathtt{abb}\overline{E_2}\cdots\mathtt{a} \\ &< \mathtt{abbb}\overline{P_6}\cdots\mathtt{b} < \cdots < \mathtt{ab}^{k-3}\overline{Q_k}\cdots\mathtt{b}. \end{aligned}$$

We have as many circular occurrences of $\mathtt{ab}$ as the number of maximal (circular) runs of $\mathtt{b}$ in $\overline{W_k}$. Then, for all $i \in [2..k-1]$, we have three cases.

Case 1: one run of $\mathtt{ab}$ in $\overline{P_i}$,
Case 2: two runs in $\mathtt{ab}$ in $\overline{E_i}$,
Case 3: one run $\mathtt{ab}$ in $\overline{Q_k}$.

For Case 1, we have one conjugate starting with $\mathtt{abb}\overline{E_i}$, for each $i \in [2..k-1]$. Since each run of $\mathtt{a}$ within each word of $\overline{P_i}$ is of length at least 2, all conjugates in Case 1 end in $\mathtt{a}$. For Case 2, for all $i \in [2..k-1]$, we can distinguish between two sub-cases based on where $\mathtt{ab}$ starts.

Case 2 (a): from the first run of $\mathtt{a}$ in $\overline{E_i}$, starting with $\mathtt{abab}^{i-2}\overline{P_{i+1}}$, if $i \in [2..k-2]$, or $\mathtt{abab}^{k-3}\overline{Q_k}$,
Case 2 (b): from the second run in $\overline{E_i}$, starting with $\mathtt{ab}^{i-2}\overline{P_{i+1}}$, if $i \in [2..k-2]$, or $\mathtt{ab}^{k-3}\overline{Q_k}$.

Similarly to Case 1, each conjugate for Case 2 (a) ends with $\mathtt{a}$. Each conjugate in Case 2 (b) is obtained by shifting two characters on the right in each conjugate in Case 2 (a). Therefore, all these conjugates end with $\mathtt{b}$.

For Case 3, the conjugate starting with $\mathtt{ab}$ in $\overline{Q_k}$ has $\mathtt{ab}\overline{P_2}$ as a prefix and is preceded by $\mathtt{a}$. Observe that only for Case 2 (b), we have one conjugate that starts with $\mathtt{abaaa}$ obtained by $\mathtt{a}\overline{P_3}$ and it is the first conjugate in the lexicographic order of $\overline{W_k}$. Then, the conjugates start with $\mathtt{abab}$ followed by $\mathtt{aba}\overline{P_3} < \mathtt{abab}\overline{P_4} < \cdots < \mathtt{abab}^{k-3}\overline{Q_k}$ from Case 2 (a).

Among the remaining conjugates, those with the prefix $\mathtt{abb}$ start with $\mathtt{ab}\overline{P_4}$ from Case 2 (b) or $\mathtt{ab}\overline{P_2}$ from Case 3. Then, among the left conjugates, the conjugate with the prefix $\mathtt{abbb}$ from Case 2 (a), for all $i \in [2..k-1]$, or $\mathtt{abb}\overline{P_5}$ from Case 2 (b) follows. The last remaining conjugates have the prefix $\mathtt{ab}^{i-2}$ for $i \in [6..k-1]$ or $\mathtt{ab}^{k-3}\overline{Q_k}$, which can be obtained by Case 2 (b). Since $\mathtt{ab}^{k-3}\overline{Q_k}$ is greater than all other conjugates, it is the greatest conjugate of $\overline{W_k}$ starting with $\mathtt{ab}$ and we conclude this proof. □

Lemma 19. $\beta(\mathtt{ba}, \overline{W_k}) = \mathtt{bb}^{2k-8}\mathtt{babba}^{k-2}$.

Proof. The conjugates in $\mathcal{M}(\overline{W_k})$ that start with $\mathtt{ba}$ are

$$\begin{aligned}\mathtt{ba}^k\mathtt{b}\overline{P_2}\cdots\mathtt{b} &< \mathtt{ba}^{k-1}\mathtt{ab}^{k-3}\overline{Q_k}\cdots\mathtt{b} < \mathtt{ba}^{k-1}\mathtt{bb}\overline{E_{k-1}}\cdots\mathtt{b} < \mathtt{ba}^{k-2}\mathtt{bab}^{k-4}\overline{P_{k-1}}\cdots\mathtt{b}\\ &< \mathtt{ba}^{k-2}\mathtt{bb}\overline{E_{k-2}}\cdots\mathtt{b} < \cdots < \mathtt{ba}^4\mathtt{babb}\overline{P_5}\cdots\mathtt{b} < \mathtt{ba}^4\mathtt{bb}\overline{E_4}\cdots\mathtt{b}\\ &< \mathtt{baaabab}\overline{P_4}\cdots\mathtt{b} < \mathtt{baaabb}\overline{E_3}\cdots\mathtt{a} < \mathtt{baaba}\overline{P_3}\cdots\mathtt{b} < \mathtt{baabb}\overline{E_2}\cdots\mathtt{b}\\ &< \mathtt{ba}\overline{P_3}\cdots\mathtt{a} < \mathtt{bab}\overline{P_4}\cdots\mathtt{a} < \cdots < \mathtt{bab}^{k-3}\overline{Q_k}\cdots\mathtt{a}.\end{aligned}$$

For integer i, we can see that $\mathtt{ba}^i\mathtt{bab}^{i-2}$ is lexicographically smaller than $\mathtt{ba}^i\mathtt{bb}$. Thus, the first conjugate in lexicographic order starting with $\mathtt{ba}$ is the one followed by the longest run of $\mathtt{a}$, and it can be found by $\mathtt{ba}^k\mathtt{b}$ of $\overline{Q_k}$, followed by conjugates starting with $\mathtt{ba}^i\mathtt{bab}^{i-2}$ of $\overline{E_i}$ and $\mathtt{ba}^i\mathtt{bb}$ of $\overline{P_i}$ for all $i \in [2..k-1]$ taken in decreasing order. By construction of $\overline{E_i}$, for $i \in [2..k-1]$, these conjugates must end with a $\mathtt{b}$. Otherwise, for $\overline{P_i}$, conjugates also end with $\mathtt{b}$, with the exception of a conjugate starting with $\overline{P_3}$, since it is preceded by an $\mathtt{a}$ from $\overline{P_2}$. The remaining conjugates starting with $\mathtt{ba}$ are exactly those conjugates that have the prefix of the suffix $\mathtt{bab}^{i-2}\overline{P_{i+1}}$ if $i \in [2..k-2]$ or $\mathtt{bab}^{k-3}\overline{Q_k}$. All of these conjugates end with $\mathtt{a}$, since they are preceded by $\mathtt{a}$. □

Lemma 20. $\beta(\mathtt{bba}, \overline{W_k}) = \mathtt{b}^{2k-8}\mathtt{abba}$.

Proof. The conjugates starting with $\mathtt{bba}$ in $\mathcal{M}(\overline{W_k})$ are

$$\begin{aligned}\mathtt{b}\overline{Q_k}\cdots\mathtt{b} &< \mathtt{b}\overline{E_{k-1}}\cdots\mathtt{b} < \mathtt{b}\overline{P_{k-1}}\cdots\mathtt{b} < \cdots < \mathtt{b}\overline{E_5}\cdots\mathtt{b} < \mathtt{b}\overline{P_5}\cdots\mathtt{b} < \mathtt{b}\overline{E_4}\cdots\mathtt{b}\\ &< \mathtt{b}\overline{P_4}\cdots\mathtt{a} < \mathtt{b}\overline{E_3}\cdots\mathtt{b} < \mathtt{b}\overline{E_2}\cdots\mathtt{b} < \mathtt{b}\overline{P_2}\cdots\mathtt{a}.\end{aligned}$$

These conjugates are obtained by following four cases.

Case 1: concatenating suffix $\mathtt{b}$ of $\overline{P_j}$ with $\overline{E_j}$ for all $j \in [2..k-1]$,
Case 2: concatenating suffix $\mathtt{b}$ of $\overline{E_j}$ with $\overline{P_{j+1}}$ for all $j \in [3..k-2]$,
Case 3: concatenating suffix $\mathtt{b}$ of $\overline{E_{k-1}}$ with $\overline{Q_k}$,
Case 4: concatenating suffix $\mathtt{b}$ of $\overline{Q_k}$ with $\overline{P_2}$.

The first conjugate in lexicographic order starting with $\mathtt{bba}$ is the one followed by the longest run of $\mathtt{a}$. The smallest of these conjugates can be found by Case 3, concatenation of the suffix $\mathtt{b}$ of $\overline{E_{k-1}}$ with $\overline{Q_k}$. We can directly observe that $\mathtt{bba}^j\mathtt{bab}^{j-2} < \mathtt{bba}^j\mathtt{bb}$ holds for every integer $j \geq 0$. Thus, the next conjugate will have the prefix $\mathtt{b}\overline{E_j}$ from Case 1 and $\mathtt{b}\overline{P_j}$ from Case 2 repeating in decreasing order. Since $\mathtt{b}\overline{E_j}$ of Case 1 and $\mathtt{b}\overline{Q_k}$ of Case 3 is preceded by a $\mathtt{b}$, those end with a $\mathtt{b}$. On the other hand, $\mathtt{b}\overline{P_{j+1}}$ precedes $\mathtt{b}$ for all $j \in [4..k-2]$ until $\mathtt{b}\overline{P_4}$ appears since it precedes an $\mathtt{a}$. Lastly, conjugates with the prefix $\mathtt{bbaaa}$ and $\mathtt{bbaa}$ by Case 1 end with a $\mathtt{b}$. The greatest lexicographic conjugate is from Case 4 as it has the smallest runs of $\mathtt{a}$ which is *two* and ends with $\mathtt{a}$.

We can sort all of these conjugates according to the order of the words in

$$\{\mathtt{b}\overline{Q_k}\}\bigcup_{j=4}^{k-1}\{\mathtt{b}\overline{E_j}\mathtt{b}\overline{P_j}\}\cup\bigcup_{j'=2}^{3}\{\overline{E_{j'}}\}\cup\{\mathtt{b}\overline{P_2}\}.$$

□

Lemma 21. $\beta(\mathtt{bbba}, \overline{W_k}) = \mathtt{b}(\mathtt{ab})^{k-6}\mathtt{a}^5$.

Proof. The conjugates in $\mathcal{M}(\overline{W_k})$ starting with bbba are

$$\mathtt{bb}\overline{Q_k}\cdots\mathtt{b} < \mathtt{bb}\overline{E_{k-1}}\cdots\mathtt{a} < \mathtt{bb}\overline{P_{k-1}}\cdots\mathtt{b} < \cdots < \mathtt{bb}\overline{E_6}\cdots\mathtt{a} < \mathtt{bb}\overline{P_6}\cdots\mathtt{b}$$
$$< \mathtt{bb}\overline{E_5}\cdots\mathtt{a} < \mathtt{bb}\overline{P_5}\cdots\mathtt{a} < \mathtt{bb}\overline{E_4}\cdots\mathtt{a} < \mathtt{bb}\overline{E_3}\cdots\mathtt{a} < \mathtt{bb}\overline{E_2}\cdots\mathtt{a}.$$

Some of the conjugates starting with bbba can be obtained by two cases.

Case 1: from the concatenation of the suffix bb of $\overline{E_{j-1}}$ with a prefix of ba of $\overline{P_j}$ for all $j \in [5..k-1]$ or $\overline{Q_k}$ if $j = k$;

Case 2: from the concatenation of the suffix bb of $\overline{P_j}$ with prefix ba of $\overline{E_j}$ for all $j \in [2..k-1]$.

Thus, all conjugates starting with bbba are sorted according to the lexicographic order of the words in $\{\mathtt{bb}\overline{Q_k}\} \cup \bigcup_{j=5}^{k-1}\{\mathtt{bb}\overline{E_j}\mathtt{bb}\overline{P_j}\} \cup \bigcup_{j=2}^{4}\{\overline{E_j}\}$. All conjugates starting with $\mathtt{bb}\overline{P_j}$ for all $j \in [6..k-1]$ or $\mathtt{bb}\overline{Q_k}$ in Case 1 end with b. Otherwise, conjugates starting with $\mathtt{bb}\overline{P_5}$ of Case 1 or $\mathtt{bb}\overline{E_j}$ for all $j \in [2..k-1]$ of Case 2 end with a. □

Lemma 22. *$\beta(\mathtt{b}^j\mathtt{a}, \overline{W_k}) = \mathtt{b}^{k-j-2}\mathtt{a}$ for all $j \in [4..k-2]$.*

Proof. All runs of b of length of a range $j \in [4..k-3]$ appear only by concatenating suffix $\mathtt{b}^{j-1}$ of E_{j+1} with prefix ba of $\overline{P_{j'}}$ for all $j' \in [j+2..k-1]$ in decreasing order. All of these conjugates end with a b, with the exception of a conjugate $\mathtt{b}^{j-1}\overline{P_{j+2}}$ which ends with an a since suffix $\mathtt{b}^{j-1}$ precedes an a. Hence, the last conjugate in lexicographic order starting with $\mathtt{b}^{k-2}\mathtt{a}$ is within $\mathtt{b}^{k-3}\overline{Q_k}$ and since the run of b is maximal it ends with a, and the claim follows. □

The following theorem presents the shape of the BWT of $\overline{W_k}$.

Theorem 3. *For every $k \geq 6, r(\overline{W_k}) = 6k - 12$. cf. Table 8.*

Table 8. Classification of the number of runs obtained in Theorem 3. The total number of runs is $6k - 12$.

BWT of $\overline{W_k}$	Runs
$\beta(\mathtt{a}^k\mathtt{b}) = \mathtt{b}$	1
$\beta(\mathtt{a}^i\mathtt{b}) = \mathtt{ba}^{2(k-1-i)+1}\mathtt{b}$ for all $i \in [2..k-1]$	$2k-3$ but, when merged, $2k-4$
$\beta(\mathtt{ab}) = \mathtt{ba}^{k-2}\mathtt{baa}^{k-5}\mathtt{baaab}^{k-5}$	7 but, when merged, 6
$\beta(\mathtt{ba}) = \mathtt{b}^{2k-6}\mathtt{abba}^{k-2}$	4 but, when merged, 3
$\beta(\mathtt{bba}) = \mathtt{b}^{2k-8}\mathtt{abba}$	4
$\beta(\mathtt{bbba}) = \mathtt{b}(\mathtt{ab})^{k-6}\mathtt{a}^5$	$2k-10$
$\beta(\mathtt{b}^i\mathtt{a}) = \mathtt{b}^{k-i-2}\mathtt{a}$, for all $i \in [4..k-2]$	$2k-12$

Proof. Let us put the result from Lemma 16 to Lemma 22 together. Every conjugate of contributing a character to $\beta(\mathtt{a}^i\mathtt{b})$ is smaller than a conjugate contributing a character to $\beta(\mathtt{a}^{i'}\mathtt{b})$, for every $1 \leq i' \leq i \leq k$. Symmetrically, every conjugate in $\beta(\mathtt{b}^j\mathtt{a})$ is greater than every conjugate in $\beta(\mathtt{b}^{j'}\mathtt{a})$, when $1 \leq j' \leq j \leq k-2$. Since we considered all the disjoint ranges of conjugates of $\overline{W_k}$ based on their common prefix, the word $\prod_{i=0}^{k-1} \beta(\mathtt{a}^{k-i}\mathtt{b}) \cdot \prod_{i=1}^{k-2} \beta(\mathtt{b}^i\mathtt{a})$ is the BWT of $\overline{W_k}$.

With the structure of BWT($\overline{W_k}$), we can derive its number of runs. The word $\prod_{i=0}^{k-1} \beta(\mathtt{a}^{k-i}\mathtt{b})$ has exactly $2k+3$ runs: we start with 1 run from $\beta(\mathtt{a}^k\mathtt{b})$ but it is merged by a b from $\beta(\mathtt{a}^{k-1}\mathtt{b})$. Then, concatenating each $\beta(\mathtt{a}^{k-1}\mathtt{b})$ up to $\beta(\mathtt{aab})$ adds 3 runs each. However, the boundaries between these words merge because b appears continuously. Thus, each $\beta(\mathtt{a}^i\mathtt{b})$ for $i \in [2..k-1]$ makes 2 runs each. By counting, we observe that $\beta(\mathtt{ab})$ runs 7 times. The remaining part of the BWT, that is, $\prod_{i=1}^{k-2}\beta(\mathtt{b}^i\mathtt{a})$ has $4k-12$ runs: the word $\beta(\mathtt{ba})$, has 4 runs, but the first b merges with a b from $\beta(\mathtt{ab})$, so we only charge 3 runs for this word. Then, $\beta(\mathtt{bba})$ and $\beta(\mathtt{bbba})$ add 4 and $1+2(k-6)+1$ runs, respectively. Finally, $\prod_{i=4}^{k-2}\beta(\mathtt{b}^i\mathtt{a})$ runs for 2 until $i = k-3$. The word $\beta(\mathtt{b}^{k-2}\mathtt{a})$ does not add new runs, as it consists only of an a that merges with the previous one. Altogether, we have $2(k-2)+7+3+4+1+2(k-6)+1+2(k-6) = 6k-12$, and the claim holds. □

The following lemmas describe the BWT of $\overline{W_k}$ after applying one specific edit operation. $\overline{W_k^\flat}\mathtt{c}$ is a word obtained by replacing the last character b of $\overline{W_k}$ with c, where c is lexicographically larger than b. The number of runs in the BWT of $\overline{W_k^\flat}\mathtt{c}$ can be derived by comparing the BWT of $\overline{W_k^\flat}\mathtt{c}$ to the BWT of $\overline{W_k}$, for which we explicitly counted the number of runs, so we omit these parts of the proof using $\mathcal{M}(\overline{W_k^\flat}\mathtt{c})$, which is a list of lexicographically sorted conjugates of word $\overline{W_k^\flat}\mathtt{c}$. Substituting the last character with c in $\overline{W_k}$ also increases the number of runs by $\Theta(k)$.

Lemma 23. $\beta(\mathtt{a}^k\mathtt{c}, \overline{W_k^\flat}\mathtt{c}) = \mathtt{b}$.

Proof. The first conjugate in $\mathcal{M}(\overline{W_k^\flat}\mathtt{c})$ starts with $\mathtt{a}^k\mathtt{c}\cdots\mathtt{b}$. The first conjugate in lexicographic order must start with the longest run of a. By the definition of $\overline{W_k^\flat}\mathtt{c}$, the longest run of a is obtained by suffix $\mathtt{a}^k\mathtt{c}$ of $\overline{Q_k^\flat}\mathtt{c}$, preceded by a b. □

Lemma 24. $\beta(\mathtt{a}^i\mathtt{b}, \overline{W_k^\flat}\mathtt{c}) = \mathtt{ba}^{2k-2i-2}\mathtt{b}$ *for all* $i \in [2..k-1]$.

Proof. The conjugates in $\mathcal{M}(\overline{W_k^\flat}\mathtt{c})$ starting with the prefix $\mathtt{a}^i\mathtt{b}$ for $i \in [2..k-1]$ are

$$\begin{aligned}\mathtt{a}^i\mathtt{bab}^{i-2}\overline{P_{i+1}}\cdots\mathtt{b} < \mathtt{a}^i\mathtt{bab}^{i-1}\overline{P_{i+2}}\cdots\mathtt{a} < \cdots < \mathtt{a}^i\mathtt{bab}^{k-4}\overline{P_{k-1}}\cdots\mathtt{a} < \mathtt{a}^i\mathtt{bab}^{k-3}\overline{Q_k^\flat}\mathtt{c}\cdots\mathtt{a}\\ < \mathtt{a}^i\mathtt{bb}\overline{E_{k-1}}\cdots\mathtt{a} < \cdots < \mathtt{a}^i\mathtt{bb}\overline{E_{k-2}}\cdots\mathtt{a} < \cdots < \mathtt{a}^i\mathtt{bb}\overline{E_{i+1}}\cdots\mathtt{a}\\ < \mathtt{a}^i\mathtt{bb}\overline{E_i}\cdots\mathtt{b}.\end{aligned}$$

For every integer $i \in [2..k-1]$, the conjugates in $\mathcal{M}(\overline{W_k^\flat}\mathtt{c})$ starting with $\mathtt{b}^i\mathtt{a}$ can only be obtained from two cases:

Case 1: $\mathtt{a}^i\mathtt{bab}^{i-2}$ of $\overline{E_j}$ for all $j \in [i..k-1]$,
Case 2: $\mathtt{a}^i\mathtt{bb}$ of $\overline{P_j}$ for all $j \in [i..k-1]$.

We can sort these conjugates according to the lexicographic order of $\bigcup_{j=i}^{k-2}\{\mathtt{a}^i\mathtt{bab}^{j-2}\overline{P_{j+1}}\} \cup \mathtt{a}^i\mathtt{bab}^{k-3}\overline{Q_k^\flat} \cup \bigcup_{j=i}^{k-1}\{\mathtt{a}^i\mathtt{bb}\overline{E_j}\}$. Note that all these conjugates end with an a, with the exception of the conjugate starting with $\mathtt{a}^i\mathtt{bab}^{i-2}\overline{P_{i+1}}$ and $\mathtt{a}^i\mathtt{bb}\overline{E_i}$, since these are the only places where the occurrence of $\mathtt{a}^i\mathtt{b}$ can be found. □

Lemma 25. $\beta(\mathtt{a}^i\mathtt{c}, \overline{W_k^\flat}\mathtt{c}) = \mathtt{a}$ *for all* $i \in [1..k-1]$.

Proof. The only conjugate in $\mathcal{M}(\overline{W_k^\flat}\mathtt{c})$ starting with $\mathtt{a}^i\mathtt{bc}$ for all $i \in [1..k-1]$ has a prefix of $\mathtt{a}^i\mathtt{bc}\overline{P_2}\cdots\mathtt{a}$. For all two distinct integers i, i' with $i > i' \geq 0$, we have $\mathtt{a}^i\mathtt{bc} < \mathtt{a}^{i'}\mathtt{bc}$. Also, since the lexicographic order of a word in $\overline{W_k^\flat}\mathtt{c}$ is $\mathtt{a} < \mathtt{b} < \mathtt{c}$, it is also clear that $\mathtt{a}^i\mathtt{b} < \mathtt{a}^i\mathtt{c}$. The conjugates starting with $\mathtt{a}^i\mathtt{c}$ are obtained from $\mathtt{a}^i\mathtt{c}$ from $\overline{Q_k^\flat}\mathtt{c}$ and since the length of a is k, all conjugates with $\mathtt{a}^i$ with $i \in [1..k-1]$ end with a. □

Lemma 26. $\beta(\mathtt{ab}, \overline{W_k^\flat}\mathtt{c}) = \mathtt{ba}^{k-2}\mathtt{ba}^{k-5}\mathtt{baaab}^{k-5}$.

Proof. In $\mathcal{M}(\overline{W_k^\flat}\mathtt{c})$, the conjugates starting with ab are

$$\begin{aligned}\mathtt{a}\overline{P_3}\cdots\mathtt{b} &< \mathtt{aba}\overline{P_3}\cdots\mathtt{a} < \mathtt{abab}\overline{P_4}\cdots\mathtt{a} < \cdots < \mathtt{abab}^{k-3}\overline{Q_k^\flat}\mathtt{c}\cdots\mathtt{a}\\ &< \mathtt{ab}\overline{P_4}\cdots\mathtt{b}\\ &< \mathtt{abb}\overline{E_{k-1}}\cdots\mathtt{a} < \cdots < \mathtt{abb}\overline{E_5}\cdots\mathtt{a}\\ &< \mathtt{abb}\overline{P_5}\cdots\mathtt{b} < \mathtt{abb}\overline{E_4}\cdots\mathtt{a} < \mathtt{abb}\overline{E_3}\cdots\mathtt{a} < \mathtt{abb}\overline{E_2}\cdots\mathtt{a}\\ &< \mathtt{abbb}\overline{P_6}\cdots\mathtt{b} < \cdots < \mathtt{ab}^{k-3}\overline{Q_k^\flat}\mathtt{c}\cdots\mathtt{b}.\end{aligned}$$

We have as many circular occurrences of ab as the number of maximal runs of a in $\overline{W_k^\flat}\mathtt{c}$. Then, for all $i \in [2..k-1]$, we have two cases.

Case 1: one run in $\overline{P_i}$ obtained by concatenating suffix abb of $\overline{P_i}$ with $\overline{E_i}$, for each $i \in [2..k-1]$, and

Case 2: two runs in $\overline{E_i}$.

For Case 1, since each run of a within each word of $\bigcup_{i=2}^{k-1}\mathtt{abb}\overline{E_i}$ is of length of at least 2, all conjugates in Case 1 end with an a.

For Case 2, for all $i \in [2..k-1]$, we can distinguish between two sub-cases, based on where ab starts, if either

Case 2 (a): from the first a in $\overline{E_i}$ or

Case 2 (b): from the second a in $\overline{E_i}$.

For Case 2 (a), we can see that these conjugates are of the type $\mathtt{abab}^{i-2}\overline{P_{i+1}}$ if $i \in [2..k-2]$ or $\mathtt{abab}^{k-3}\overline{Q_k^\flat}\mathtt{c}$. Similarly to Case 1, each conjugate for Case 2 (a) ends with a. Each conjugate in Case 2 (b) is obtained by shifting two characters on the right each conjugate in Case 2 (a). Therefore, all of these conjugates end with a b and have prefix of type $\mathtt{ab}^{i-2}\overline{P_{i+1}}$, if $i \in [2..k-2]$ or $\mathtt{ab}^{k-3}\overline{Q_k^\flat}\mathtt{c}$. All these conjugates end with a b since a is preceded by b. Observe that only for Item Case 2 (b), we have conjugates starting with abaaab which is $\mathtt{a}\overline{P_3}$. Hence, it is the first conjugate in lexicographic order, followed by those starting with $\mathtt{aba}\overline{P_3} < \mathtt{abab}\overline{P_4} < \cdots < \mathtt{abab}^{k-3}\overline{Q_k^\flat}\mathtt{c}$ from Item Case 2 (a) and these conjugates start with abab.

Next, conjugates with a prefix of abba which is $\mathtt{ab}\overline{P_4}$ from Case 2 (b) follow, then those having prefix abbba either start with $\mathtt{abb}\overline{E_i}$ for all $i \in [5..k-1]$ from Case 1 follow in decreasing order. Then, $\mathtt{abb}\overline{P_5}$ from Case 2 (b) and $\mathtt{abb}\overline{E_4}$, $\mathtt{abb}\overline{E_3}$, $\mathtt{abb}\overline{E_2}$ from Case 1 follow.

The remaining conjugates are those which start with a prefix of $\mathtt{ab}^i\mathtt{a}$ for $i \in [4..k-2]$, which are obtained by $\mathtt{ab}^{i-1}\overline{P_{i+2}}$ if $i \in [4..k-3]$ or $\mathtt{ab}^{k-3}\overline{Q_k^\flat}\mathtt{c}$, from Case 2 (b). These conjugates are sorted according to the length of the run of a following the common prefix. Then, the result is

$$\begin{aligned}&\{\mathtt{a}\overline{P_3}\} \cup \bigcup_{j=2}^{k-2}\{\mathtt{abab}^{j-2}\overline{P_{j+1}}\} \cup \{\mathtt{aba}^{k-3}\overline{Q_k^\flat}\mathtt{c}\} \cup \{\mathtt{ab}\overline{P_4}\} \cup \bigcup_{i=0}^{k-6}\{\mathtt{abb}\overline{E_{k-i-1}}\}\\ &\cup \{\mathtt{abb}\overline{P_5}\} \cup \bigcup_{i=0}^{2}\{\mathtt{abb}\overline{E_{4-i}}\} \cup \bigcup_{j=4}^{k-3}\{\mathtt{a}^{j-1}\overline{P_{j+2}}\} \cup \{\mathtt{ab}^{k-3}\overline{Q_k^\flat}\mathtt{c}\}.\end{aligned}$$

□

Lemma 27. $\beta(\mathtt{ba}, \overline{W_k^\flat}\mathtt{c}) = \mathtt{b}^{2k-6}\mathtt{abca}^{k-2}$.

Proof. The conjugates in $\mathcal{M}(\overline{W_k^\flat}\mathtt{c})$ starting with the prefix $\mathtt{ba}$ are

$$\begin{aligned}\mathtt{ba}^k\mathtt{c}\overline{P_2}\cdots\mathtt{b} &< \mathtt{ba}^{k-1}\mathtt{ab}^{k-3}\overline{Q_k^\flat}\mathtt{c}\cdots\mathtt{b} < \mathtt{ba}^{k-1}\mathtt{bb}\overline{E_{k-1}}\cdots\mathtt{b} < \cdots < \mathtt{ba}^4\mathtt{babb}\overline{P_5}\cdots\mathtt{b}\\ &< \mathtt{ba}^4\mathtt{bb}\overline{E_4}\cdots\mathtt{b} < \mathtt{ba}^3\mathtt{bab}\overline{P_4}\cdots\mathtt{b} < \mathtt{ba}^3\mathtt{bb}\overline{E_3}\cdots\mathtt{a} < \mathtt{baaba}\overline{P_3}\cdots\mathtt{b}\\ &< \mathtt{baabb}\overline{E_2}\cdots\mathtt{c} < \mathtt{ba}\overline{P_3}\cdots\mathtt{a} < \mathtt{bab}\overline{P_4}\cdots\mathtt{a} < \cdots < \mathtt{bab}^{k-3}\overline{Q_k^\flat}\mathtt{c}\cdots\mathtt{a}.\end{aligned}$$

There are many occurrences of a conjugate starting with the prefix $\mathtt{ba}$, and it occurs in three parts.

Case 1: one run of $\mathtt{ba}$ in $\overline{P_j}$, for all $j \in [2..k-1]$,
Case 2: two runs from $\overline{E_j}$, for all $j \in [2..k-1]$,
Case 3: one run from $\overline{Q_k^\flat}\mathtt{c}$.

The conjugates in Case 1 start with $\mathtt{ba}^j$ for all $j \in [2..k-1]$. Since $\mathtt{ba}^i < \mathtt{ba}^{i'}$ if only $i > i'$, the conjugates are sorted in decreasing order. All conjugates for $j \in [3..k-1]$ end with a $\mathtt{b}$, except for conjugates with prefix $\overline{P_2}$ since it is preceded by $\mathtt{c}$.

In Case 2, we can distinguish between two sub-cases based on where $\mathtt{ba}$ starts:

Case 2 (a): first run of $\mathtt{ba}$ from the prefix of $\overline{E_j}$,
Case 2 (b): from the second run of $\mathtt{ba}$ in $\overline{E_j}$.

The conjugates in Case 2 (a) are the type of $\mathtt{ba}^j\mathtt{bab}^{j-2}\overline{P_{j+1}}$, if $j \in [2..k-2]$ or $\mathtt{ba}^{k-1}\mathtt{bab}^{k-3}\overline{Q_k^\flat}\mathtt{c}$. All of these conjugates are preceded by $\overline{P_j}$, thus ending with $\mathtt{b}$. The conjugates in Case 2 (b) start from $\mathtt{bab}^{j-2}\overline{P_{j+1}}$ if $j \in [2..k-2]$ or $\mathtt{bab}^{k-3}\overline{Q_k^\flat}\mathtt{c}$ and end with an $\mathtt{a}$.

In Case 3, only one conjugate can be found by a prefix of $\mathtt{ba}^k\mathtt{c}$, which ends with $\mathtt{b}$.

Observe that only for Case 3 we have a conjugate with the longest run of $\mathtt{a}$ after $\mathtt{b}$. Hence, the first conjugate in lexicographic order is $\mathtt{ba}^k\mathtt{c}\overline{P_2}$ from Case 3. It is followed by $\mathtt{ba}^{k-1}\mathtt{bab}^{k-3}\overline{Q_k^\flat}\mathtt{c} < \mathtt{ba}^{k-1}\mathtt{bb}\overline{E_{k-1}} < \mathtt{ba}^{k-2}\mathtt{bab}^{k-4}\overline{P_{k-1}} < \mathtt{ba}^{k-2}\mathtt{bb}\overline{E_{k-2}} < \cdots < \mathtt{ba}^4\mathtt{babb}\overline{P_5} < \mathtt{ba}^4\mathtt{bb}\overline{E_4}$. All of these conjugates end with a $\mathtt{b}$.

Among the remaining conjugates, those having prefix $\mathtt{baaab}$ either start with $\mathtt{baaabab}\overline{P_4}$ from Case 2 (a) or $\mathtt{baaabb}\overline{E_3}$ from Case 1. Then, the remaining conjugates with prefix $\mathtt{baab}$ are those starting with $\mathtt{baaba}\overline{P_3}$ from Case 2 (a) or $\mathtt{baabb}\overline{E_2}$ from Case 1. Lastly, $k-2$ conjugates from Case 2 (b) follow, which are $\mathtt{bab}^{j-2}\overline{P_{j+1}}$ for all $j \in [2..k-2]$ or $\mathtt{bab}^{k-3}\overline{Q_k^\flat}\mathtt{c}$. All of these conjugates end with an $\mathtt{a}$.

We prove our claim by sorting lexicographically the conjugates in

$$\{\mathtt{ba}^k\mathtt{c}\overline{P_2}\} \cup \bigcup_{j=0}^{k-3}\{\mathtt{ba}^{k-j-1}\mathtt{bab}^{k-j-3}\overline{P_{k-j}} \cdot \mathtt{ba}^{k-j-1}\mathtt{bb}\overline{E_{k-j-1}}\} \cup \bigcup_{j=2}^{k-2}\{\mathtt{bab}^{j-2}\overline{P_{j+1}}\} \cup \{\mathtt{bab}^{k-3}\overline{Q_k^\flat}\mathtt{c}\}.$$

□

Lemma 28. $\beta(\mathtt{bba}, \overline{W_k^\flat}\mathtt{c}) = \mathtt{b}^{2k-8}\mathtt{abb}$.

Proof. The conjugates in $\mathcal{M}(\overline{W_k^\flat}\mathtt{c})$ starting with prefix $\mathtt{bba}$ are

$$\begin{aligned}\mathtt{bba}^k\mathtt{c}\overline{P_2}\cdots\mathtt{b} &< \mathtt{bba}^{k-1}\mathtt{bab}^{k-3}\overline{Q_k^\flat}\mathtt{c}\cdots\mathtt{b} < \mathtt{bba}^{k-1}\mathtt{bb}\overline{E_{k-1}}\cdots\mathtt{b} < \cdots\\ &< \mathtt{bba}^5\mathtt{bab}^3\overline{P_6}\cdots\mathtt{b} < \mathtt{bba}^5\mathtt{bb}\overline{E_5}\cdots\mathtt{b} < \mathtt{bba}^4\mathtt{babb}\overline{P_5}\cdots\mathtt{b}\\ &< \mathtt{bba}^4\mathtt{bb}\overline{E_4}\cdots\mathtt{a} < \mathtt{bba}^3\mathtt{bab}\overline{P_4}\cdots\mathtt{b} < \mathtt{bbaaba}\overline{P_3}\cdots\mathtt{b}.\end{aligned}$$

The smallest conjugate with prefix $\mathtt{bba}$ can be obtained by three cases.

Case 1: concatenating suffix $\mathtt{b}$ of $\overline{E_{k-1}}$ with $\overline{Q_k^\flat}\mathtt{c}$,
Case 2: concatenation of suffix $\mathtt{b}$ of $\overline{E_j}$ with $\overline{P_{j+1}}$ if $j \in [3..k-2]$ or $\overline{Q_k^\flat}\mathtt{c}$,

Case 3: concatenating suffix $\mathtt{b}$ of $\overline{P_j}$ with $\overline{E_j}$, for all $j \in [2..k-1]$.

The conjugates in Case 1 and Case 3 end with $\mathtt{b}$. Also, conjugates from Case 2 end with $\mathtt{b}$ with an exception of a conjugate starting with $\mathtt{b}\overline{P_4}$ since it is preceded by an $\mathtt{a}$. We conclude this proof by sorting lexicographically the conjugates in

$$\{\mathtt{bba}^k\mathtt{c}\overline{P_2}\} \cup \bigcup_{j=0}^{k-5}\{\mathtt{bba}^{k-i-1}\mathtt{bab}^{k-i-3}\overline{P_{k-i}}\mathtt{bba}^{k-i-1}\mathtt{bb}\overline{E_{k-i-1}}\} \cup \bigcup_{j=0}^{1}\{\mathtt{bba}^{3-j}\mathtt{bab}^{1-j}\overline{P_{4-j}}\}.$$

□

Lemma 29. $\beta(\mathtt{bbba}, \overline{W_k^\flat}\mathtt{c}) = \mathtt{b}(\mathtt{ab})^{k-6}\mathtt{aaaaa}$.

Proof. The conjugates in $\mathcal{M}(\overline{W_k^\flat}\mathtt{c})$ starting with the prefix $\mathtt{bbba}$ are

$$\begin{aligned}\mathtt{bb}\overline{Q_k^\flat}\mathtt{c}\cdots\mathtt{b} < \mathtt{bb}\overline{E_{k-1}}\cdots\mathtt{a} < \mathtt{bb}\overline{P_{k-1}}\cdots\mathtt{b} < \cdots < \mathtt{bb}\overline{E_6}\cdots\mathtt{a} < \mathtt{bb}\overline{P_6}\cdots\mathtt{b}\\ < \mathtt{bb}\overline{E_5}\cdots\mathtt{a} < \mathtt{bb}\overline{P_5}\cdots\mathtt{a} < \mathtt{bb}\overline{E_4}\cdots\mathtt{a} < \mathtt{bb}\overline{E_3}\cdots\mathtt{a} < \mathtt{bb}\overline{E_2}\cdots\mathtt{a}.\end{aligned}$$

Analogously to Lemma 28, the conjugates starting with $\mathtt{bbba}$ can be obtained from three cases.

Case 1: concatenating suffix $\mathtt{bb}$ of $\overline{E_{k-1}}$ with $\overline{Q_k^\flat}\mathtt{c}$,
Case 2: concatenation of suffix $\mathtt{bb}$ of $\overline{E_j}$ with $\overline{P_{j+1}}$ if $j \in [4..k-2]$ or $\overline{Q_k^\flat}\mathtt{c}$,
Case 3: concatenating a suffix $\mathtt{bb}$ of $\overline{P_j}$ with $\overline{E_j}$, for all $j \in [2..k-1]$.

The conjugate in Case 1 is the smallest conjugate starting with $\mathtt{bbba}$ since it has a longest run of $\mathtt{a}$ and ends with a $\mathtt{b}$. In addition, the conjugates of Case 3 end with a $\mathtt{a}$ since $\mathtt{bb}$ are preceded by an $\mathtt{a}$. In Case 2, all the conjugates end with $\mathtt{b}$ with an exception of a conjugate starting with $\mathtt{bb}\overline{P_5}$ since it is preceded by an $\mathtt{a}$. We can sort these conjugates by

$$\{\mathtt{bb}\overline{Q_k^\flat}\mathtt{c}\} \cup \bigcup_{j=0}^{k-6}\{\mathtt{bba}^{k-i-1}\mathtt{bab}^{k-i-3}\overline{P_{k-i}}\mathtt{bba}^{k-i-1}\mathtt{bb}\overline{E_{k-i-1}}\} \cup \bigcup_{j=0}^{2}\{\mathtt{bba}^{4-j}\mathtt{bab}^{2-j}\overline{P_{5-j}}\}.$$

□

Lemma 30. $\beta(\mathtt{b}^j\mathtt{a}, \overline{W_k^\flat}\mathtt{c}) = \mathtt{b}^{k-j-2}\mathtt{a}$ *for all* $j \in [4..k-2]$.

Proof. In $\mathcal{M}(\overline{W_k^\flat}\mathtt{c})$, the conjugates starting with prefix $\mathtt{b}^j\mathtt{a}$ for all $j \in [4..k-2]$ are

$$\mathtt{b}^{j-1}\overline{Q_k^\flat}\mathtt{c}\cdots\mathtt{b} < \mathtt{b}^{j-1}\overline{P_{k-1}}\cdots\mathtt{b} < \mathtt{b}^{j-1}\overline{P_{k-2}}\cdots\mathtt{b} < \cdots < \mathtt{b}^{j-1}\overline{P_{j+3}}\cdots\mathtt{b} < \mathtt{b}^{j-1}\overline{P_{j+2}}\cdots\mathtt{a}.$$

Observe that the only conjugates with the prefix $\mathtt{b}^j\mathtt{a}$ for $j \in [4..k-2]$ start with concatenating $\mathtt{b}^{j-1}$ either to $\overline{Q_k^\flat}\mathtt{c}$ or $\overline{P_{j'}}$ if $j' \in [j+2..k-1]$. One can see that these conjugates taken in this order are already sorted, and all conjugates end with a $\mathtt{b}$, with the exception of a conjugate starting with $\mathtt{b}^{j-1}\overline{P_{j+2}}$, since it is preceded by an $\mathtt{a}$, therefore ending with an $\mathtt{a}$. We have all conjugates ordered according to the lexicographic order of the words in $\mathtt{b}^{j-1}\overline{Q_k^\flat}\mathtt{c} \cup \bigcup_{j'=0}^{k-j-3}\{\mathtt{b}^{j-1}\overline{P_{k-j'-1}}\}$. This concludes our proof. □

Lemma 31. $\beta(\mathtt{c}, \overline{W_k^\flat}\mathtt{c}) = \mathtt{a}$.

Proof. The only conjugate in $\mathcal{M}(\overline{W_k^\flat \mathtt{c}})$ that starts with prefix $\mathtt{c}$ is $\mathtt{c}\overline{P_2}\cdots\mathtt{a}$. Since $\mathtt{c}$ is lexicographically larger than other characters such as $\mathtt{a}$, $\mathtt{b}$, it is the biggest conjugate in $\mathcal{M}(\overline{W_k^\flat \mathtt{c}})$, and it ends with an $\mathtt{a}$. □

The following theorem puts the lemmas above together.

Theorem 4. *Substituting the last character* $\mathtt{b}$ *of* $\overline{W_k}$ *by* $\mathtt{c}$ *increases* r *by* $2k-5$, *cf. Table 9.*

Table 9. Classification of the number of runs obtain in Theorem 4. The total number of runs is $8k-17$.

BWT of $\overline{W_k^\flat \mathtt{c}}$	Runs
$\beta(\mathtt{a}^k\mathtt{c}) = \mathtt{b}$	1
$\beta'(\mathtt{a}^{k-1}\mathtt{b}) = \mathtt{bb}$	1 but, when merged, 0
$\beta'(\mathtt{a}^i\mathtt{b}) = \mathtt{ba}^{2k-2i-2}\mathtt{b}$ for all $i \in [2..k-2]$	$3k-9$
$\beta'(\mathtt{a}^i\mathtt{c}) = \mathtt{a}$ for all $i \in [1..k-1]$	$k-1$
$\beta'(\mathtt{ab}) = \mathtt{ba}^{k-2}\mathtt{ba}^{k-5}\mathtt{baaab}^{k-5}$	7
$\beta'(\mathtt{ba}) = \mathtt{b}^{2k-6}\mathtt{abca}^{k-2}$	5
$\beta'(\mathtt{bba}) = \mathtt{b}^{2k-8}\mathtt{abb}$	3
$\beta'(\mathtt{bbba}) = \mathtt{b}(\mathtt{ab})^{k-6}\mathtt{a}^5$	$2k-10$ but, when merged, $2k-11$
$\beta'(\mathtt{b}^i\mathtt{a}) = \mathtt{b}^{k-i-2}\mathtt{a}$ for all $i \in [4..k-2]$	$2k-12$
$\beta'(\mathtt{c}) = \mathtt{a}$	1 but, when merged, 0

Proof. Every conjugate contributing a character to $\beta(\mathtt{a}^i\mathtt{b})$ is smaller than a conjugate contributing a character to $\beta(\mathtt{a}^{i'}\mathtt{b})$ for every $1 \le i' < i \le k-1$. By symmetry, every conjugate contributing a character to $\beta(\mathtt{b}^j\mathtt{a})$ is greater than each conjugate contributing a character to $\beta(\mathtt{b}^{j'}\mathtt{a})$ for every $1 \le j' \le j \le k-2$. With the structure of the BWT of $(\overline{W_k^\flat \mathtt{c}})$, we can easily derive its number of runs. $\beta(\mathtt{a}^k\mathtt{c}) \cdot \beta(\mathtt{a}^{k-1}\mathtt{c}) \cdot \prod_{i=1}^{k-2}\beta(\mathtt{a}^i\mathtt{b}) \cdot \beta(\mathtt{a}^i\mathtt{c})$ has exactly $4k-2$ runs: we start from 1 run from $\beta(\mathtt{a}^k\mathtt{c})$ but it is merged with $\beta(\mathtt{a}^{k-1}\mathtt{b})$. $\beta(\mathtt{a}^{k-1}\mathtt{b})$ and $\beta(\mathtt{a}^{k-1}\mathtt{c})$ add 2 runs. Then, concatenating each $\beta(\mathtt{a}^i\mathtt{b})$ and $\beta(\mathtt{a}^i\mathtt{c})$ for all $i \in [2..k-2]$ in a decreasing order, we add 3 and 1 runs each, which results in $4(k-3)$ runs. By counting, we observe that $\beta(\mathtt{ab}), \beta(\mathtt{a\#})$ adds 7 and 1 runs, respectively.

The word $\beta(\mathtt{ba}), \beta(\mathtt{bba}), \beta(\mathtt{bbba})$ has exactly 5, 3, $2k-10$ runs each, but since the boundaries between $\beta'(\mathtt{bba})$ and $\beta(\mathtt{bbba})$ merge, the first $\mathtt{b}$ of $\beta(\mathtt{bbba})$ does not count, turning into $2k-11$. The remaining part of BWT, that is,$\prod_{j=4}^{k-3}\beta(\mathtt{b}^j\mathtt{a}) \cdot \beta(\mathtt{b}^{k-2}\mathtt{a}) \cdot \beta(\mathtt{c})$ has $2k-12$ runs: we start by concatenating each $\beta'(\mathtt{b}^4\mathtt{a})$ up to $\beta(\mathtt{b}^{k-3}\mathtt{a})$, which adds 2 runs each. The last $\beta(\mathtt{b}^{k-2}\mathtt{a}), \beta(\mathtt{c})$ does not add new runs, as it consists only of an $\mathtt{a}$ that merges with the previous one. Altogether, we have $2+4(k-3)+7+1+5+3+2k-11+2(k-6) = 8k-17$, and the claim holds.

The main difference between $\overline{W_k}$ and $\overline{W_k^\flat \mathtt{c}}$ comes from $\mathtt{a}^i\mathtt{b}$ that is concatenated with $\mathtt{a}^i\mathtt{c}$ for $i \in [2..k-1]$, which repeats $\mathtt{baba}$, while $\overline{W_k}$ repeats $\mathtt{ba}$ only, making $2k-5 = \Theta(k)$ more runs. Tables 10–13 describe the scheme of the BWT of word $\overline{W_k^\flat \mathtt{c}}$. We have $r(\overline{W_K^\flat \mathtt{c}}) = r(\overline{W_k}) + 2k - 5$. From Definition 1, we have $k = \Theta(\sqrt{n})$. Thus, $r(\overline{W_K^\flat \mathtt{c}}) - r(\overline{W_k}) = 2k-5 = \Theta(\sqrt{n})$. □

Table 10. Lexicographically sorted conjugates of $\overline{W_k^\flat}\texttt{c}$ studied in Theorem 4, Part 1.

Prefix	Remaining Part	BWT
$\texttt{a}^k\texttt{c}$	$\overline{P_2}$	b
$\texttt{a}^{k-1}\texttt{b}$	$\texttt{ab}^{k-2}$ $\texttt{bba}^{k-1}$	b b
$\texttt{a}^{k-1}\texttt{c}$	$\overline{P_2}$	a
$\texttt{a}^{k-2}\texttt{b}$	$\texttt{ab}^{k-3}$ $\texttt{ab}^{k-2}$ $\texttt{bba}^{k-1}$ $\texttt{bba}^{k-2}$	b a a b
$\texttt{a}^{k-2}\texttt{c}$	$\overline{P_2}$	a
$\texttt{a}^{k-3}\texttt{b}$	$\texttt{ab}^{k-4}$ $\texttt{ab}^{k-3}$ $\texttt{ab}^{k-2}$ $\texttt{bba}^{k-1}$ $\texttt{bba}^{k-2}$ $\texttt{bba}^{k-3}$	b a a a a b
$\texttt{a}^{k-3}\texttt{c}$	$\overline{P_2}$	a
...	...	...
aab	ab $\texttt{ab}^2$ ... $\texttt{ab}^{k-2}$ $\texttt{bba}^{k-1}$ ... $\texttt{bba}^3$ $\texttt{bba}^2$	b a a a a a a b
aac	$\overline{P_2}$	a

Table 11. Lexicographically sorted conjugates of $\overline{W_k^\flat}\texttt{c}$ studied in Theorem 4, Part 2.

Prefix	Remaining Part	BWT
ab	aaabb $\texttt{a}\overline{P_3}$ $\texttt{ab}\overline{P_4}$ $\texttt{abb}\overline{P_5}$ ... $\texttt{ab}^{k-4}\overline{P_{k-1}}$ $\texttt{ab}^{k-3}\overline{Q_k^\flat}\texttt{c}$ $\overline{P_4}$ $\texttt{b}\overline{E_{k-1}}$ $\texttt{b}\overline{E_{k-2}}$ ... $\texttt{b}\overline{E_5}$ $\texttt{b}\overline{P_5}$ $\texttt{b}\overline{E_4}$ $\texttt{b}\overline{E_3}$ $\texttt{b}\overline{E_2}$ $\texttt{bb}\overline{P_6}$ $\texttt{bbb}\overline{P_7}$ $\texttt{bbbb}\overline{P_8}$ ... $\texttt{b}^{k-5}\overline{P_{k-1}}$ $\texttt{b}^{k-4}\overline{Q_k^\flat}\texttt{c}$	b a a a a a a b a a a a b a a a b b b b b b
ac	$\overline{P_2}$	a

Table 12. Lexicographically sorted conjugates of $\overline{W_k^\flat}$c studied in Theorem 4, Part 3.

Prefix	Remaining Part	BWT
ba	$a^{k-1}c$	b
	$a^{k-2}bab^{k-3}$	b
	$a^{k-2}bb$	b
	$a^{k-3}bab^{k-4}$	b
	$a^{k-3}bb$	b
	...	...
	aaababb	b
	aaabb	b
	aaabab	b
	aabb	a
	aba	b
	abb	#
	$\overline{P_3}$	a
	b$\overline{P_4}$	a
	...	a
	$b^{k-4}\overline{P_{k-1}}$	a
	$b^{k-3}\overline{Q_k^\flat}c$	a
bba	$a^{k-1}c$	b
	$a^{k-2}bab^{k-3}$	b
	$a^{k-2}bb$	b
	...	...
	a^4bab^3	b
	a^4bb	b
	a^3babb	b
	aaabb	a
	aabab	b
	aba	b

Table 13. Lexicographically sorted conjugates of $\overline{W_k^\flat}$c studied in Theorem 4, Part 4.

Prefix	Remaining Part	BWT
bbba	$a^{k-1}c$	b
	$a^{k-2}bab^{k-3}$	a
	$a^{k-2}bb$	b
	$a^{k-3}bab^{k-4}$	a
	$a^{k-3}bb$	b
	...	...
	a^5bab^4	a
	a^5bb	b
	a^4bab^3	a
	a^4bb	a
	a^3bab^2	a
	aabab	a
	aba	a
bbbba	$a^{k-1}c$	b
	$a^{k-2}bb$	b
	...	b
	a^6bb	b
	a^5bb	a
...	...	...
$b^{k-3}a$	$a^{k-1}c$	b
	$a^{k-2}bb$	a
$b^{k-2}a$	$a^{k-1}c$	a
c	$\overline{P_2}$	a

6. Multiplicative Sensitivity of ρ by $\Omega(\log n)$

Recall that $\rho(W) = \mathtt{runs}(\text{BBWT}(W))$. In this section, we return our attention to Fibonacci words. Similar to Section 4, we use them to construct a family of words with a multiplicative sensitivity of $\Theta(\log n)$ for the number of runs ρ in the BBWT. Before that, we start with some helpful lemmas known in the literature.

Lemma 32 ([41], Lemma 3)**.** *The* $2k$*th Fibonacci word* F_{2k} *is* $X_{2k}\mathtt{ab}$*. The Lyndon conjugate of the Fibonacci word* F_{2k} *is* $L_{2k} = \mathtt{a}X_{2k}\mathtt{b}$*.*

Lemma 33 ([37], Lemma 6)**.** *We let* L_{2k} *be the Lyndon conjugate of the Fibonacci word* F_{2k}*. Then, r(BBWT(*L_{2k}*))= 2.*

Lemma 34 ([41], Lemma 8)**.** *If* $k < n$*, then the Lyndon conjugate of* F_k *is a prefix or a suffix of* $\mathtt{a}X_n\mathtt{b}$*. If* $F_k = P_k\mathtt{ba}$*, then its Lyndon conjugate* $\mathtt{a}P_k\mathtt{b}$ *is a prefix of* $\mathtt{a}P_n\mathtt{b}$*; and if* $F_k = P_k\mathtt{ab}$*, then its Lyndon conjugate* $\mathtt{a}P_k\mathtt{b}$ *is a suffix of* $\mathtt{a}P_n\mathtt{b}$*.*

The next lemma addresses the extended Burrows–Wheeler transform [16], which takes a subset of steps from the BBWT by expecting the input to be a set of primitive words (i.e., the Lyndon factors in case of the BBWT). We translate the following known result to the BBWT:

Lemma 35 (Corollary 4 of [47])**.** *We let* $\{T_1, \ldots, T_m\}$ *be a conjugate-free set of primitive words and let* r' *be the number of runs of its extended Burrows–Wheeler transform. Then,* $m \leq r'$*.*

Corollary 1. *We let* $T_1, \ldots, T_m$ *be the Lyndon factors of word* T*, then* $m \leq \rho(T)$*.*

In what follows, we establish a lower bound on the multiplicative sensitivity of ρ with the Lyndon conjugates of Fibonacci words by leveraging Corollary 1.

6.1. Editing the Last Position of L_{2k}

We start with deleting the last character of L_{2k}, which directly leads to the following insight.

Theorem 5. $\rho(L^{\flat}_{2k}) \geq k$*.*

Proof. $L^{\flat}_{2k} = \mathtt{a}X_{2k}$ is not a Lyndon word; therefore, its Lyndon is factorized and has more than one factor. According to Lemma 34, the Lyndon word of the Fibonacci word $F_{2k} = X_{2k}\mathtt{ab}$ is $\mathtt{a}X_{2k}\mathtt{b}$. The central word X_{2k} is $X_{2k-1}baX_{2k-2}$, so the Lyndon word of F_{2k} is $\mathtt{a}X_{2k}\mathtt{b} = \mathtt{a}X_{2k-1}\mathtt{ba}X_{2k-2}\mathtt{b}$. $\mathtt{a}X_{2k-1}\mathtt{b}$ refers to L_{2k-1}, which is $\mathtt{a}X_{2k-1}\mathtt{b}$ and the suffix $\mathtt{a}X_{2k-2}\mathtt{b}$ is L_{2k-2}.

However, by deleting the last character $\mathtt{b}$, $L^{\flat}_{2k}$ becomes $\mathtt{a}X_{2k-1}\mathtt{ba}X_{2k-2}$, meaning that L_{2k-2} does not exist. Thus, we can say that L_{2k-1} is one of the Lyndon factors, since it is not followed by L_{2k-2}. The remaining part of $L^{\flat}_{2k}$ is $\mathtt{a}X_{2k-2}$. The same as X_{2k}, central word X_{2k-2} can be divided as $X_{2k-3}\mathtt{ba}X_{2k-4}$; thus, $\mathtt{a}X_{2k-2} = \mathtt{a}X_{2k-3}\mathtt{ba}X_{2k-4}$. We can find Lyndon factor $L_{2k-3} = \mathtt{a}X_{2k-3}\mathtt{b}$ in the prefix. The remaining part is $\mathtt{a}X_{2k-4}$, which is not a Lyndon word, same as $\mathtt{a}X_{2k-2}$ above, so $\mathtt{a}X_{2k-4}$ is Lyndon factorized and makes L_{2k-5} as a prefix, and the remaining $\mathtt{a}X_{2k-6}$ makes L_{2k-7} as a prefix. And finally, $\mathtt{a}X_4$ is divided as $\mathtt{a}X_3\mathtt{ba}X_2$, where X_2 is ε. Therefore, $L^{\flat}_{2k}$'s Lyndon factor is L_{2i-1} for $i \in [2..k]$ and the last remaining part $\mathtt{a}$ is the Lyndon word itself. Thus, $L^{\flat}_{2k}$ has Lyndon factors L_{2i-1} for every $i \in [2..k]$ and $\mathtt{a}$ as a Lyndon factor. The number of the Lyndon factor is k, which we depict in Figure 6. □

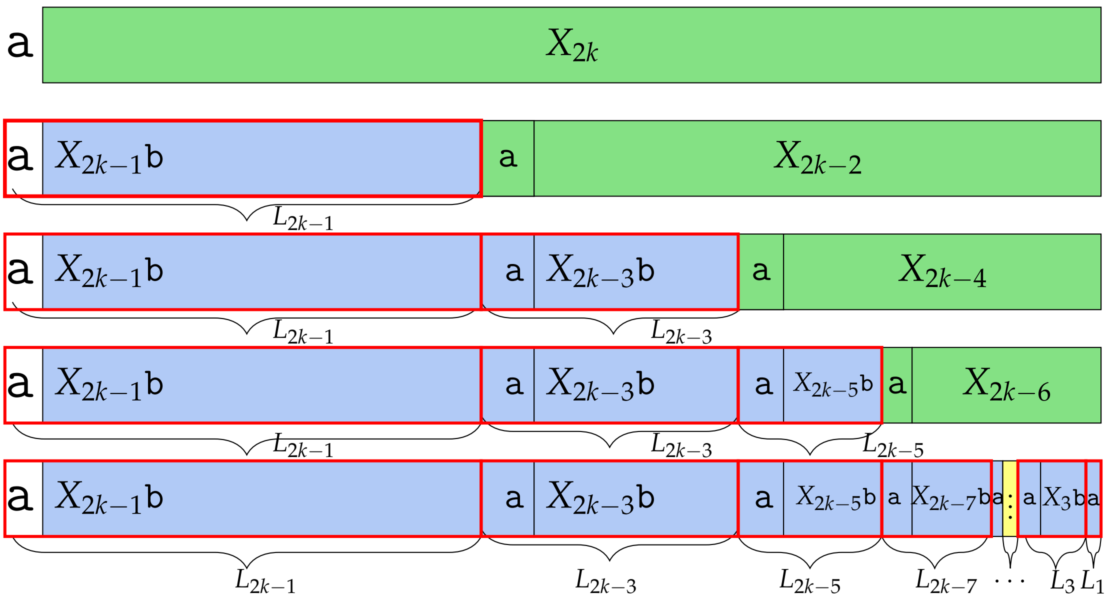

Figure 6. Factorization of $L_{2k}^{\flat}$ into Lyndon factors studied in the proof of Theorem 5. $L_{2k}^{\flat}$ has k Lyndon factors.

By Lemma 33 and Theorem 5, we conclude that the multiplicative sensitivity for deleting the last character of L_{2k} is $\Omega(k)$.

Theorem 6. *We let $L_{2k}^{\flat}$# be the word obtained by substituting the last character* b *of L_{2k} by* #*. Then, $\rho(L_{2k}^{\flat}\#) \geq k+1$.*

Proof. Since # is lexicographically smaller than a, $L_{2k}^{\flat}$# is not a Lyndon word; it makes Lyndon factors. Since # is smaller than both a and b, # is a Lyndon word. In addition, $L_{2k}^{\flat}$ is Lyndon factorized as Theorem 5, which produces Lyndon factors L_{2i-1} for $i \in [2..k]$ and the last Lyndon factor a. $L_{2k}^{\flat}$# makes one more Lyndon factor, which is #, which therefore makes $k+1$ a number of Lyndon factors. We depict the Lyndon factorization in Figure 7. □

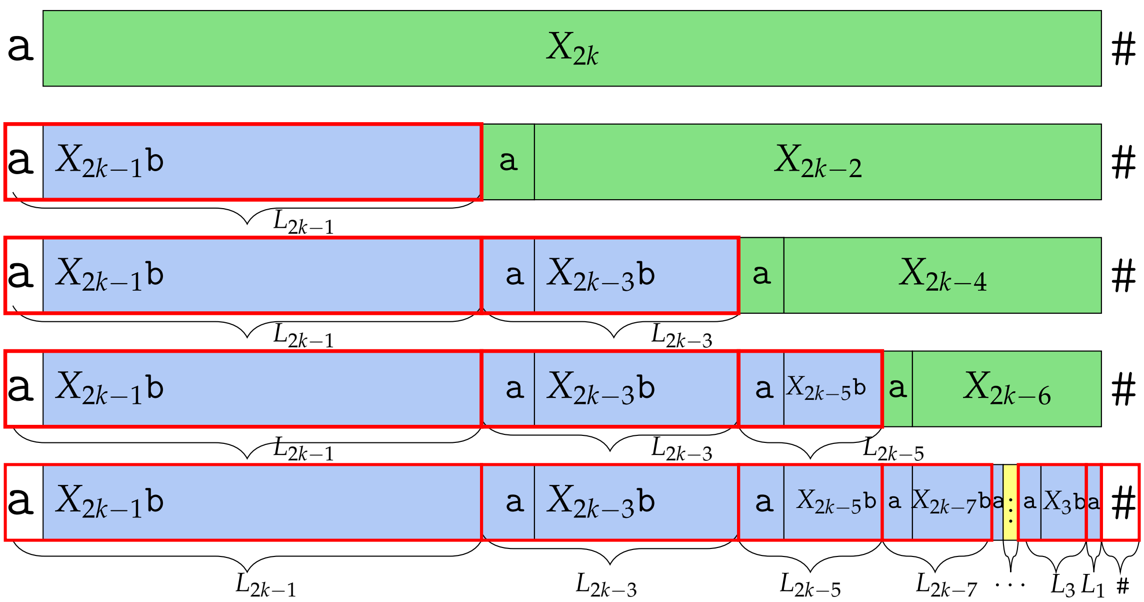

Figure 7. Factorization of $L_{2k}^{\flat}$# into Lyndon factors studied in the proof of Theorem 6. $L_{2k}^{\flat}$# has $k+1$ Lyndon factors.

By Lemma 33 and Theorem 6, we conclude that the multiplicative sensitivity for substituting the last character of L_{2k} is $\Omega(k)$. We observe a similar result when substituting the last character with a larger character instead of a smaller one (#).

Theorem 7. $\rho(L^{\flat}_{2k}\mathtt{c}) \geq k$.

Proof. The lexicographic order between a, b, and c is $\mathtt{a} < \mathtt{b} < \mathtt{c}$. Recall that $L^{\flat}_{2k}$ makes L_{2i-1} for $i \in [2..k]$, and a as a Lyndon factor. In $L^{\flat}_{2k}\mathtt{c}$, c is in position f_{2k}; therefore, it does not affect anything until the last Lyndon factor a. ac is the Lyndon word itself because $\mathtt{a} < \mathtt{c}$. Therefore, $L^{\flat}_{2k}\mathtt{c}$ makes a k number of Lyndon factors, shown in Figure 8. □

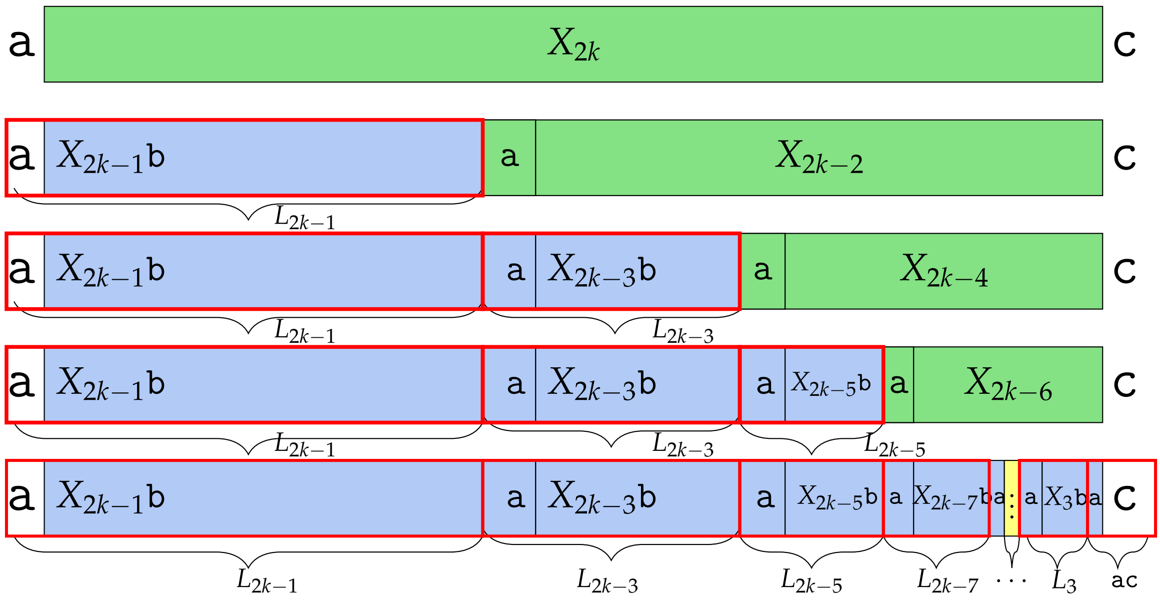

Figure 8. Factorization of $L^{\flat}_{2k}\mathtt{c}$ into Lyndon factors studied in the proof of Theorem 7. $L^{\flat}_{2k}\mathtt{c}$ has k Lyndon factors.

6.2. Insertions at Specific Locations

According to Corollary 1, ρ is lower bounded by the number of distinct Lyndon factors. After editing L_{2k} at any position, we can still find consecutive Lyndon conjugates of lower order which can merge to a higher order. For instance, $L_{2k-1} \cdot L_{2k-2}$ merge into L_{2k}, which can decrease the number of the Lyndon factor. Also, $L_{2k-3} \cdot L_{2k-2}$ merge into L_{2k-1}. Our idea is to avoid consecutive Fibonacci Lyndon conjugates so that they do not merge because doing so avoids a decrease in a number of *distinct* Lyndon factors. Now, we consider editing the specific location of Fibonacci Lyndon conjugates, also resulting in an increase in runs. The following theorems describe the bijective BWT of L_{2k} after some specific edit operations are applied.

Theorem 8. *We let L_{2k} be a Fibonacci Lyndon conjugate. By inserting* a *at position α in L_{2k}, ρ is at least k.*

Proof. We let α be the number of additions of odd Fibonacci numbers $f_{2k-3} + f_{2k-5} + \cdots + f_3 + f_1$. Recall that the Fibonacci word $F_i = X_i c$ with $c \in \{\mathtt{ab}, \mathtt{ba}\}$ has the Lyndon conjugate $L_i = \mathtt{a}X_i\mathtt{b}$. Further, $L_{2k} = L_{2k-1} \cdot L_{2k-2} = \mathtt{a}X_{2k-1}\mathtt{b} \cdot \mathtt{a}X_{2k-2}\mathtt{b}$. Thus, we start with $\mathtt{a}X_{2k-1}\mathtt{b} \cdot \mathtt{a}X_{2k-2}\mathtt{b}$. To obtain many distinct Lyndon factors, we aim to produce Lyndon factors that are not consecutive. Knowing $X_{2k-1} = X_{2k-3}\mathtt{ba}X_{2k-2}$, $\mathtt{a}X_{2k-2}$ merges with $\mathtt{a}X_{2k-3}\mathtt{b}$ into L_{2k-1}, so it is best to divide X_{2k-2}. $\mathtt{a}X_{2k-2}$ divides into $\mathtt{a}X_{2k-3}\mathtt{ba}X_{2k-4}$. In this case, it is best to add $\mathtt{a}X_{2k-3}\mathtt{b}$ as a new Lyndon factor since it is smallest among those Lyndon factors that are not consecutive with X_{2k-1}, the same as $\mathtt{a}X_{2k-2}$; $\mathtt{a}X_{2k-4}$ divides into $\mathtt{a}X_{2k-5}\mathtt{ba}X_{2k-6}$, and we add $\mathtt{a}X_{2k-5}\mathtt{b}$ as a Lyndon factor. $\mathtt{a}X_{2k-6}$ divides into $\mathtt{a}X_{2k-7}\mathtt{ba}X_{2k-8}$ as we add $\mathtt{a}X_{2k-7}\mathtt{b}$ as a Lyndon factor. The addition of Lyndon factors of $2i-1$ for $i \in [1..k-1]$ continues until $\mathtt{a}X_5 = \mathtt{a}X_3\mathtt{ba}X_4$ appears since $\mathtt{a}X_3\mathtt{b}$ is the second smallest Lyndon factor in Fibonacci. Thus, we need $X_1 = \mathtt{a}$ as the last Lyndon factor and it is obtained by inserting a # in $\mathtt{a}X_4$, dividing $\mathtt{a}X_4$ into $\mathtt{a\#}X_4$. Since # is lexicographically smaller than any words from right to #, the right words become the Lyndon factor. Thus, we can obtain k Lyndon factors by inserting # in L_{2k} : $k-1$ factors from $L_{2k-3} \cdot L_{2k-5} \cdots L_1$ and one from # concatenated with the remaining words. And this is shown in Figure 9. □

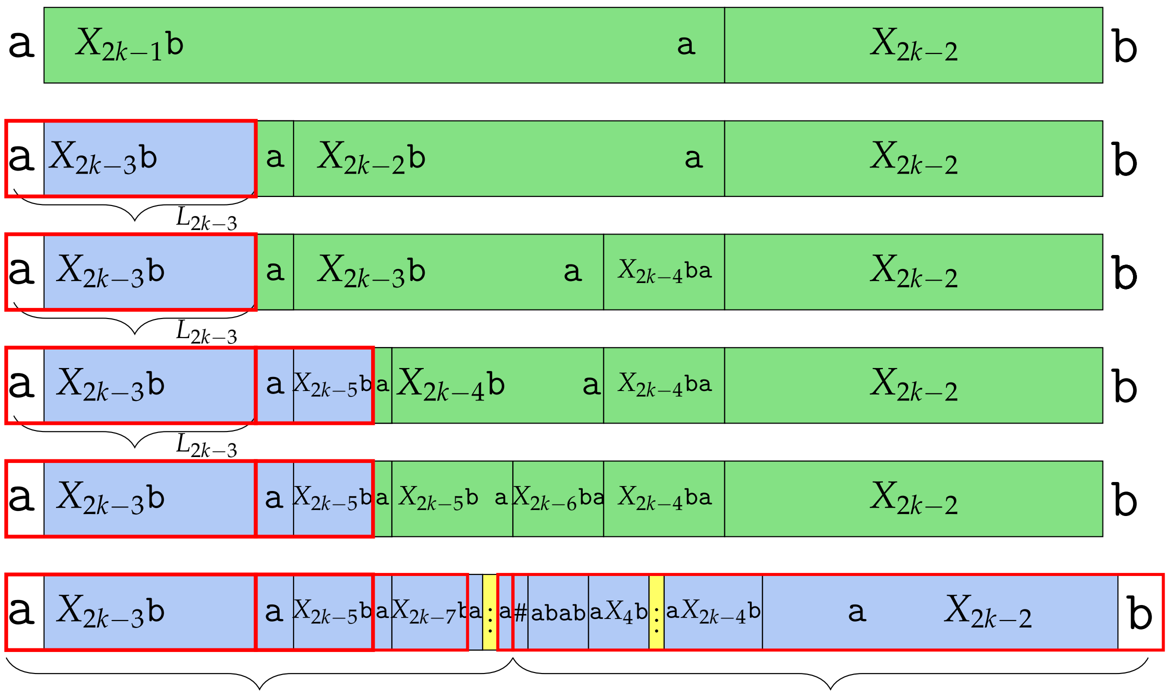

Figure 9. Inserting # at position α in L_{2k} considered in the proof in Theorem 8.

By Lemma 33 and Theorem 8, we conclude that the multiplicative sensitivity for inserting a character into L_{2k} is $\Omega(k)$. In the same way, we can also insert the special character # to observe a similar behavior:

Theorem 9. *We let L_{2k} be a Fibonacci Lyndon conjugate. By inserting # at position $f_{2k}-2$ in L_{2k}, ρ is at least $k+1$.*

Proof. Unlike Theorem 8, we can obtain some Lyndon factors on the right side of $\mathtt{a}X_{2k}\mathtt{b}$, adding $\mathtt{a}X_{2k-1}\mathtt{b} = L_{2k-1}$ as a Lyndon factor. We divide $\mathtt{a}X_{2k-2}\mathtt{b}$ into $\mathtt{a}X_{2k-3}\mathtt{ba}X_{2k-4}\mathtt{b}$ and obtain $\mathtt{a}X_{2k-3}\mathtt{b} = L_{2k-3}$. Further, we divide $\mathtt{a}X_{2k-4}\mathtt{b}$ into $\mathtt{a}X_{2k-5}\mathtt{ba}X_{2k-6}\mathtt{b}$, making L_{2k-5}. We divide $\mathtt{a}X_{2k-6}\mathtt{b}$ and can obtain Lyndon factors such as $L_{2k-7} \cdots L_5$. Lastly, $\mathtt{a}X_4\mathtt{b}$ divides into $\mathtt{a}X_3\mathtt{ba}X_2\mathtt{b}$, but since X_2 is ε, the last Lyndon factor obtained here is L_3. To make more Lyndon factors, we can add # between a and b, turning into a#b, adding 2 Lyndon factors which are $\mathtt{a} = L_1$ and #b. Thus, we can obtain $k+1$ Lyndon factors here: k factors by $L_{2k-1}, L_{2k-3} \cdots L_1$ and one from #b. We visualize the Lyndon factorization in Figure 10. □

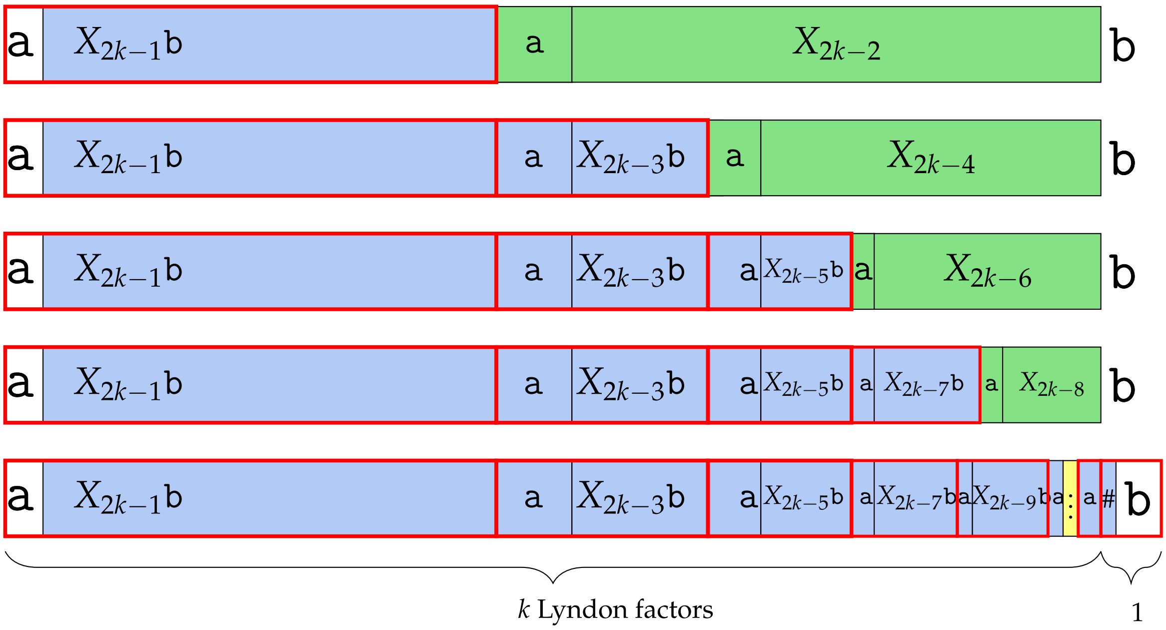

Figure 10. Insertion of # at position $f_{2k} - 2$ in L_{2k} increases ρ by at least the number of distinct Lyndon factors $k+1$ studied in Theorem 9.

7. Additive Sensitivity of ρ by $\Omega(\sqrt{n})$

Here, we study the additive sensitivity of ρ with an approach similar to Section 5. In what follows, we establish that the additive sensitivity of ρ is at least $\Theta(\sqrt{n})$. To that end, we again make use of the word W_k. Recall that $W_k = (\prod_{i=2}^{k-1} P_i E_i) Q_k = \left(\prod_{i=2}^{k-1} \mathtt{ab}^i\mathtt{aaab}^i\mathtt{aba}^{i-2}\right)\mathtt{ab}^k\mathtt{a}$.

Lemma 36. *The Lyndon conjugate C_k of W_k is* $\mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a} \cdot \left(\prod_{i=2}^{k-2} P_i E_i\right) \cdot P_{k-1}\mathtt{ab}^{k-1}\mathtt{ab}$.

Proof. The Lyndon conjugate of W_k starts with the longest runs of $\mathtt{a}$, which can be obtained by concatenating suffix $\mathtt{a}^{k-3}$ of E_{k-1} with prefix $\mathtt{a}$ of Q_k. Therefore, $C_k = \mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a} \cdot \left(\prod_{i=2}^{k-2} P_i E_i\right) \cdot P_{k-1}\mathtt{ab}^{k-1}\mathtt{ab} = \mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a} \cdot \left(\prod_{i=2}^{k-2} \mathtt{ab}^i\mathtt{aaab}^i\mathtt{aba}^{i-2}\right) \cdot \mathtt{ab}^{k-1}\mathtt{aaab}^{k-1}\mathtt{ab}$. □

Lemma 37. $\rho(C_k) = 6k - 12$.

Proof. According to Lemma 1, all conjugates have the same BWT, thus $r(W_k) = r(C_k) = 6k - 12$. Also, since C_k is a Lyndon word, $r(C_k) = \rho(C_k) = 6k - 12$. □

Recall that the runs in the BBWT and BWT are the same if the input word is Lyndon, cf. Lemma 1. Thus, we can leverage BWT computation if the input word is Lyndon since we can obtain the number of runs in the same way as in Section 5 by using $\beta(W)$ for word W. In this section, we focus on three variations of the word C_k: deleting its last character and substituting its last character $\mathtt{b}$ with $\mathtt{c}$ or #.

7.1. Deletions and Edits of C_k with a Character Smaller than $\mathtt{a}$

Recall that $C_k = \mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a} \cdot \left(\prod_{i=2}^{k-2} \mathtt{ab}^i\mathtt{aaab}^i\mathtt{aba}^{i-2}\right) \cdot \mathtt{ab}^{k-1}\mathtt{aaab}^{k-1}\mathtt{ab}$. Thus, $C_k^\flat$, which is obtained by deleting the last character $\mathtt{b}$, is $C_k^\flat = \mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a} \cdot \left(\prod_{i=2}^{k-2} \mathtt{ab}^i\mathtt{aaab}^i\mathtt{aba}^{i-2}\right) \cdot \mathtt{ab}^{k-1}\mathtt{aaab}^{k-1}\mathtt{a}$. Recall that the Lyndon conjugate of $C_k^\flat$ is the strictly smallest conjugate of all conjugates of $C_k^\flat$. Since we obtain the longest runs of $\mathtt{a}$s from a conjugate of $C_k^\flat$ by concatenating the last $\mathtt{a}$ with $\mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a}$, $C_k^\flat$ cannot be a Lyndon word. In fact, it has *two*

Lyndon factors, which are $\mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a} \cdot \left(\prod_{i=2}^{k-2} \mathtt{ab}^i\mathtt{aaab}^i\mathtt{aba}^{i-2}\right) \cdot \mathtt{ab}^{k-1}\mathtt{aaab}^{k-1}$, and we refer to both Lyndon factors as D_k (first factor) and $\mathtt{a}$ from now on. Figure 11 shows the Lyndon factorization. Since $r(\mathtt{a})$ is 1, the only thing left to check is $r(D_k)$. In D_k, we made a slight modification to the subword E_{k-1}. In fact, $E_{k-1} = \mathtt{ab}^{k-1}\mathtt{aba}^{k-3}$ was changed to $\mathtt{ab}^{k-1}\mathtt{a}^{k-3}$, which we call H_{k-1} in this section. Since D_k is a Lyndon word, we determine $\rho(D_k)$ using $\mathcal{M}(D_k)$ with the BWT as we did before.

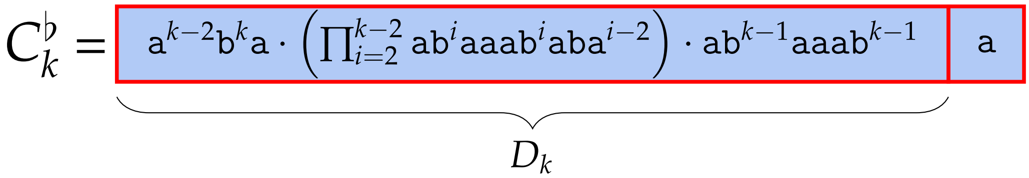

Figure 11. Introducing D_k from $C_k^\flat$ studied in Section 7.1. D_k is the first Lyndon factor of $C_k^\flat$.

Lemma 38. $\beta(\mathtt{a}^{k-2}\mathtt{b}, D_k) = \mathtt{b}$.

Proof. The *only* conjugate in $\mathcal{M}(D_k)$ starting with prefix $\mathtt{a}^{k-2}\mathtt{b}$ is $\mathtt{a}^{k-2}\mathtt{b}P_2 \cdots \mathtt{b}$. The first conjugate in lexicographic order must start with the longest run of $\mathtt{a}$. By the definition of D_k, the longest run of $\mathtt{a}$ has length $k-2$, and it is obtained by concatenating suffix $\mathtt{a}^{k-3}$ with prefix $\mathtt{a}$ of Q_k that is preceded by $\mathtt{b}$ (otherwise, we could extend the sequence of $\mathtt{a}$ characters). □

Lemma 39. $\beta(\mathtt{a}^i\mathtt{b}, D_k) = \mathtt{ba}^{k-i-2}$, *for all* $i \in [4..k-3]$.

Proof. With integer $i \in [3..k-3]$, the conjugates in $\mathcal{M}(D_k)$ starting with $\mathtt{a}^i\mathtt{b}$ are

$$\mathtt{a}^{i-1}P_{i+2} \cdots \mathtt{b} < \mathtt{a}^{i-1}P_{i+3} \cdots \mathtt{a} < \cdots < \mathtt{a}^{i-1}P_{k-1} \cdots \mathtt{a} < \mathtt{a}^i\mathtt{b}^k\mathtt{a}P_2 \cdots \mathtt{a}.$$

For all $i \in [4..k-3]$, the factor $\mathtt{a}^i\mathtt{b}$ can only be obtained from the concatenation of suffix $\mathtt{a}^{i-1}$ from E_{j-1},

- with the prefix $\mathtt{ab}$ of P_j for a $j \in [i+2..k-1]$ or
- with the prefix $\mathtt{ab}$ of Q_k, if $j = k$.

We can sort these conjugates according to the lexicographic order of $\bigcup_{j=i+2}^{k-1} P_j \cup Q_k$. All these conjugates end with an $\mathtt{a}$, with the exception of the conjugate starting with $\mathtt{a}^i P_{i+2}$, since D_k has a unique occurrence of $\mathtt{ba}^i\mathtt{b}$. □

Lemma 40. $\beta(\mathtt{aaab}, D_k) = \mathtt{bbbbb}(\mathtt{ab})^{k-7}\mathtt{baa}$.

Proof. The conjugates in $\mathcal{M}(D_k)$ starting with $\mathtt{aaab}$ are

$$\begin{aligned}\mathtt{aa}E_2 \cdots \mathtt{b} < \mathtt{aa}E_3 \cdots \mathtt{b} < \mathtt{aa}E_4 \cdots \mathtt{b} < \mathtt{aa}P_5 \cdots \mathtt{b} < \mathtt{aa}E_5 \cdots \mathtt{b} \\ < \mathtt{aa}P_6 \cdots \mathtt{a} < \mathtt{aa}E_6 \cdots \mathtt{b} < \cdots < \mathtt{aa}P_{k-2} \cdots \mathtt{a} < \mathtt{aa}E_{k-2} \cdots \mathtt{b} \\ < \mathtt{aa}H_{k-1} \cdots \mathtt{b} < \mathtt{aa}P_{k-1} \cdots \mathtt{a} < \mathtt{aa}Q_k \cdots \mathtt{a}.\end{aligned}$$

The above conjugates are obtained in the following cases.

Case 1: by concatenating the suffix $\mathtt{aa}$ of E_{i-1} with the prefix $\mathtt{ab}$ of P_i, if only $i \in [5..k-1]$,

Case 2: by concatenating the suffix $\mathtt{aa}$ of P_i, with the prefix $\mathtt{ab}$ of E_i, for all $i \in [2..k-2]$ or with H_{k-1} if $i = k-1$,

Case 3: by concatenating the suffix $\mathtt{aa}$ of H_{k-1} with the prefix $\mathtt{ab}$ of Q_k.

All these conjugates starting with aaab are sorted according to the lexicographic order of the words in $\bigcup_{i=2}^{4}\{\mathtt{aa}E_i\} \cup \bigcup_{j=5}^{k-2}\{\mathtt{aa}P_j \cdot \mathtt{aa}E_j\} \cup \{\mathtt{aa}H_{k-1}\} \cup \{\mathtt{aa}P_{k-1}\} \cup \{\mathtt{aa}Q_k\}$. The conjugates starting either with $\mathtt{aa}P_i$, for all $i \in [6..k-1]$ in Case 1 or Case 3, end with an a. On the other hand, conjugates of Case 2 or $\mathtt{aa}P_5$ in Case 1 end with a b. □

Lemma 41. $\beta(\mathtt{aab}, D_k) = \mathtt{baaba}^{2k-8}$.

Proof. The conjugates in $\mathcal{M}(D_k)$ starting with aab are

$$\mathtt{a}P_2\cdots\mathtt{b} < \mathtt{a}E_2\cdots\mathtt{a} < \mathtt{a}E_3\cdots\mathtt{a} < \mathtt{a}P_4\cdots\mathtt{b} < \mathtt{a}E_4\cdots\mathtt{a}$$
$$< \mathtt{a}P_5\cdots\mathtt{a} < \mathtt{a}E_5\cdots\mathtt{a} < \cdots < \mathtt{a}H_{k-1}\cdots\mathtt{a} < \mathtt{a}P_{k-1}\cdots\mathtt{a} < \mathtt{a}Q_k\cdots\mathtt{a}.$$

Each of the cases from Case 1 to Case 3 in Lemma 40 induces a conjugate starting with aab, obtained by shifting on the left character a. It follows that all of these conjugates end with a. The other *two* conjugates that start with an aab are obtained by

- concatenating the suffix a of Q_k with the prefix ab of P_2 or
- concatenating suffix a of E_3 to the prefix ab of P_4.

In both cases, the obtained conjugates end with b. We conclude this proof by sorting lexicographically the conjugates in $\mathtt{a}P_2 \cup \bigcup_{i=2}^{3}\{\mathtt{a}E_i\} \cup \bigcup_{i'=4}^{k-2}\{\mathtt{a}P_{i'} \cdot \mathtt{a}E_{i'}\} \cup \{\mathtt{a}H_{k-1}\} \cup \{\mathtt{a}P_{k-1}\} \cup \{\mathtt{a}Q_k\}$. □

Lemma 42. $\beta(\mathtt{ab}, D_k) = \mathtt{b}^{k-3}\mathtt{aaba}^{2k-6}$.

Proof. The conjugates in $\mathcal{M}(D_k)$ starting with ab are

$$\mathtt{aba}^{k-4}P_{k-1}\cdots\mathtt{b} < \cdots < \mathtt{ab}P_3\cdots\mathtt{b}$$
$$< P_2\cdots\mathtt{a} < E_2\cdots\mathtt{a} < P_3\cdots\mathtt{b}$$
$$< E_3\cdots\mathtt{a} < P_4\cdots\mathtt{a} < E_4\cdots\mathtt{a} < \cdots < P_{k-2}\cdots\mathtt{a} < E_{k-2}\cdots\mathtt{a}$$
$$< H_{k-1}\cdots\mathtt{a} < P_{k-1}\cdots\mathtt{a} < Q_k\cdots\mathtt{a}.$$

The above conjugates are obtained in the following cases.

Case 1: P_i for all $i \in [2..k-1]$,
Case 2: prefix ab of E_i, for all $i \in [2..k-1]$,
Case 3: $\mathtt{aba}^{i-2}$ from E_i, for all $i \in [2..k-2]$ or ab from H_{k-1},
Case 4: ab from Q_k.

For two distinct integers i, i' with $i > i' \geq 0$, we have $\mathtt{aba}^{i} > \mathtt{aba}^{i'}$. Thus, the first conjugate in lexicographic order starting with ab is the one followed by the longest run of as. The smallest of these conjugates can be found by concatenating the suffix $\mathtt{aba}^{k-4}$ with the prefix ab of P_{k-1} from Case 3. Then, the remaining conjugates in Case 3 which are $\mathtt{aba}^{i-2}$ of E_i for all $i \in [2..k-3]$ follow in decreasing order. By construction of E_i, for all $i \in [2..k-2]$, these conjugates must end with a b. Note that the remaining cases are obtained by shifting the character a from the conjugates starting with aab from Lemma 41 with the exception of the character starting with P_3. It follows that the latter ends with a b, while all the other conjugates end with a. □

Lemma 43. $\beta(\mathtt{ba}, D_k) = \mathtt{ba}^{k-6}\mathtt{bbbab}^{k-4}\mathtt{ab}^{k-3}\mathtt{a}$.

Proof. The conjugates in $\mathcal{M}(D_k)$ starting with $\mathtt{ab}$ are

$$\begin{aligned}\mathtt{ba}^{k-2}\mathtt{b}^{k}\mathtt{a}P_2\cdots\mathtt{b} < \mathtt{ba}^{k-4}P_{k-1}\cdots\mathtt{a} < \mathtt{ba}^{k-5}P_{k-2}\cdots\mathtt{a} < \cdots < \mathtt{baaa}P_6\cdots\mathtt{a}\\ < \mathtt{baa}E_2\cdots\mathtt{b} < \mathtt{baa}E_3\cdots\mathtt{b} < \mathtt{baa}E_4\cdots\mathtt{b} < \mathtt{baa}P_5\cdots\mathtt{a}\\ < \mathtt{baa}E_5\cdots\mathtt{b} < \cdots < \mathtt{baa}E_{k-2}\cdots\mathtt{b} < \mathtt{baa}H_{k-1}\cdots\mathtt{b} < \mathtt{ba}P_2\cdots\mathtt{b}\\ < \mathtt{ba}P_4\cdots\mathtt{a}\\ < \mathtt{baba}^{k-4}P_{k-1}\cdots\mathtt{b} < \cdots < \mathtt{bab}P_3\cdots\mathtt{b}\\ < \mathtt{b}P_3\cdots\mathtt{a}.\end{aligned}$$

The conjugates above are obtained by following cases.

Case 1: suffix $\mathtt{baa}$ of P_i concatenating with E_i for all $i \in [2..k-2]$ or H_{k-1} if $i = k-1$,
Case 2: runs in E_i for all $i \in [2..k-2]$,
Case 3: suffix $\mathtt{ba}$ from Q_k concatenating with P_2,
Case 4: $\mathtt{ba}^{k-3}$ of H_{k-1} concatenating with Q_k.

We have as many circular occurrences of $\mathtt{ba}$ as the number of maximal runs of $\mathtt{b}$s in D_k. For Case 1, we have one conjugate starting with $\mathtt{baa}E_i$ for all $i \in [2..k-2]$ or $\mathtt{baa}H_{k-1}$. Since each run of $\mathtt{b}$s within each word from $\bigcup_2^{k-1} P_i$ is of length of at least 2, all conjugates of Case 1 end with $\mathtt{b}$.

For Case 2, for all $i \in [2..k-2]$, we can distinguish two subcases based on where $\mathtt{ba}$ starts:

Case 2 (a): the first run of $\mathtt{ba}$ in E_i, which has a type of $\mathtt{baba}^{i-2}$ for all $i \in [2..k-2]$,
Case 2 (b): the second run of $\mathtt{ba}$ in E_i, which has a type of $\mathtt{ba}^{i-2}$ for all $i \in [2..k-2]$.

- For Case 2 (a), we can see that these conjugates start with $\mathtt{baba}^{i-2}P_{i+1}$, if $i \in [2..k-2]$. Similarly to Case 1, each conjugate for Case 2 (a) ends with a $\mathtt{b}$. Each conjugate in Case 2 (b) is obtained by shifting two characters on the right each conjugate in Case 2 (a). Therefore, all of these conjugates end with an $\mathtt{a}$ and have prefixes of the type $\mathtt{ba}^{i-2}P_{i+1}$, if $i \in [2..k-2]$.
- For Case 3, the conjugate starting with $\mathtt{ba}$ in Q_k has $\mathtt{ba}P_2$ as a prefix, and it is preceded by a $\mathtt{b}$.
- Lastly, for Case 4, the conjugates start with $\mathtt{ba}^{k-3}$ concatenating with Q_k which ends with a $\mathtt{b}$.
- Observe that only for Case 4 and Case 2 (b) we have conjugates starting with $\mathtt{baaaa}$. Hence, the first conjugate in lexicographic order is the one from Case 4 starting with $\mathtt{ba}^{k-3}Q_k$, followed by those from Case 2 (b) which are $\mathtt{ba}^{k-4}P_{k-1} < \mathtt{ba}^{k-5}P_{k-2} < \cdots < \mathtt{baaa}P_6$.

Among the remaining conjugates, those having prefix $\mathtt{baaa}$ either start with $\mathtt{baa}P_5$ from Case 2 (b) and from Case 1 starting with $\mathtt{baa}E_i$ for all $i \in [2..k-2]$ or $\mathtt{baa}H_{k-1}$ if $i = k-1$. We can sort them according to the order of the words in

$$\bigcup_{i=2}^{4}\{\mathtt{baa}E_i\} \cup \{\mathtt{baa}P_5\} \cup \bigcup_{i=5}^{k-2}\{\mathtt{baa}E_i\} \cup \{\mathtt{baa}H_{k-1}\}.$$

Then, the remaining conjugates with prefix $\mathtt{baa}$ are those starting with $\mathtt{ba}P_2$ from Case 3 and $\mathtt{ba}P_4$ from Case 2 (b). Finally, let us focus on the conjugates from Case 2 (a). These conjugates are sorted according to the length of the run of $\mathtt{a}$s following the common prefix $\mathtt{bab}$. The last conjugate left is the one starting with $\mathtt{b}P_3$ from Case 2 (b). Since this conjugate is greater than each conjugate considered in Case 2 (a), this is the greatest conjugate of D_k starting with $\mathtt{ba}$ and the thesis follows. □

Lemma 44. $\beta(\mathtt{b}^j\mathtt{a}, D_k) = \mathtt{bab}^{2k-2j-2}\mathtt{a}$ *for all* $j \in [2..k-2]$.

Proof. With integer $i \in [2..k-2]$, the conjugates in $\mathcal{M}(D_k)$ starting with the prefix $\mathtt{b}^i\mathtt{a}$ are

$$\mathtt{b}^i\mathtt{a}^{k-3}Q_k\cdots\mathtt{b} < \mathtt{b}^i\mathtt{aa}E_i\cdots\mathtt{a} < \mathtt{b}^i\mathtt{aa}E_{i+1}\cdots\mathtt{b} < \cdots < \mathtt{b}^i\mathtt{aa}E_{k-2}\cdots\mathtt{b} < \mathtt{b}^i\mathtt{aa}H_{k-1}\cdots\mathtt{b}$$
$$< \mathtt{b}^i\mathtt{a}P_2\cdots\mathtt{b} < \mathtt{b}^i\mathtt{aba}^{k-4}P_{k-1}\cdots\mathtt{b} < \cdots < \mathtt{b}^i\mathtt{aba}^{i-1}P_{i+2}\cdots\mathtt{b}$$
$$< \mathtt{b}^i\mathtt{aba}^{i-2}P_{i+1}\cdots\mathtt{a}.$$

Case 1: concatenating $\mathtt{b}^i\mathtt{aa}$ of P_j with E_j for all $j \in [i..k-1]$ or with H_{k-1} if only $j = k-1$,
Case 2: concatenating $\mathtt{b}^i\mathtt{aba}^{j-2}$ of E_j with P_{j+1} if only $j \in [i..k-2]$,
Case 3: concatenating $\mathtt{b}^i\mathtt{a}^{k-3}$ of H_{k-1} with Q_k,
Case 4: concatenating $\mathtt{b}^i\mathtt{a}$ with P_2.

We consider these four cases separately. For all $j \in [i..k-2]$, the conjugate starting within P_j has a prefix of $\mathtt{b}^i\mathtt{aa}E_j$ or $\mathtt{b}^i\mathtt{aa}H_{k-1}$ (Case 1). For all $j \in [i..k-2]$, the conjugates starting within E_j have a prefix of $\mathtt{b}^i\mathtt{aba}^{j-2}P_{j+1}$ (Case 2). In addition, conjugate starting within a word in Case 3 has a prefix of $\mathtt{b}^i\mathtt{a}^{k-3}Q_k$. Finally, the conjugates starting with Q_k starts with $\mathtt{b}^i\mathtt{a}P_2$ (Case 4). By construction, we can see that first we have all the conjugates first from Case 3 and then from Case 1 sorted according to the lexicographic order into $\bigcup_{j=i}^{k-2}\mathtt{b}^i\mathtt{aa}E_j \cup \mathtt{b}^i\mathtt{aa}H_{k-1}$; then, we have the conjugate from Case 4, then Case 2 sorted according to the decreasing length of the run of as following the common prefix $\mathtt{b}^i\mathtt{ab}$. Moreover, we note that only when the run of bs is exactly of length i, the conjugate ends with a. Thus, only the conjugates ending with an a are those starting within $\mathtt{b}^i\mathtt{aa}E_i$ and $\mathtt{b}^i\mathtt{aba}^{i-2}P_{i+1}$. □

Lemma 45. $\beta(\mathtt{b}^{k-1}\mathtt{a}, D_k)$= aab.

Proof. There are *three* conjugates in $\mathcal{M}(D_k)$ starting with prefix $\mathtt{b}^{k-1}\mathtt{a}$. These conjugates are

$$\mathtt{b}^{k-1}\mathtt{a}^{k-3}Q_k\cdots\mathtt{a} < \mathtt{b}^{k-1}\mathtt{aaab}^{k-1}\cdots\mathtt{a} < \mathtt{b}^{k-1}\mathtt{a}P_2\cdots\mathtt{b}.$$

Observe that the only conjugates with prefix $\mathtt{b}^{k-1}\mathtt{a}$ have the prefixes, respectively, of $\mathtt{b}^{k-1}\mathtt{a}^{k-3}Q_k$, $\mathtt{b}^{k-1}\mathtt{aa}H_{k-1}$, and $\mathtt{b}^{k-1}\mathtt{a}P_2$. One can see that these conjugates taken in this order are already sorted, and only the conjugate starting within Q_k ends with b, while the other two have a. □

Lemma 46. $\beta(\mathtt{b}^k\mathtt{a}, D_k) = \mathtt{a}$.

Proof. The last conjugate in $\mathcal{M}(D_k)$ with prefix $\mathtt{b}^k\mathtt{a}$ is $\mathtt{b}^k\mathtt{a}P_2\cdots\mathtt{a}$. Finally, the only occurrence of $\mathtt{b}^k$ is within Q_k. Hence, the last conjugate in lexicographic order starts with $\mathtt{b}^k\mathtt{a}P_2$, and since the run of b′ is maximal, it ends with an a, and the thesis follows. □

We summarize the above lemmas as follows.

Lemma 47. *For integer* $k \geq 10$, $\rho(D_k) = 8k - 18$, *cf. Table 14. The BWT of the word* D_k *is given by* $BBWT(D_k) = \prod_{i=2}^{k-1}\beta(\mathtt{a}^{k-i}\mathtt{b}) \cdot \prod_{i=1}^{k}\beta(\mathtt{b}^i\mathtt{a})$.

Proof. Every conjugate of $\beta(\mathtt{a}^i\mathtt{b})$ is smaller than each conjugate of $\beta(\mathtt{a}^{i'}\mathtt{b})$ for every $1 \leq i' < i \leq k-2$. Symmetrically, every conjugate of $\beta(\mathtt{b}^j\mathtt{a})$ is greater than any conjugate of $\beta(\mathtt{b}^{j'}\mathtt{a})$, for every $1 \leq j' < j \leq k$. Since we considered all the disjoint ranges of conjugates of D_k based on their common prefix, the word $\prod_{i=2}^{k-1}\beta(\mathtt{a}^{k-i}\mathtt{b}) \cdot \prod_{i=1}^{k}\beta(\mathtt{b}^i\mathtt{a})$ is the BWT of D_k.

With the structure of $\mathrm{BBWT}(D_k)$, we can easily derive its number of runs. The word $\prod_{i=2}^{k-4}\beta(\mathtt{a}^{k-i}\mathtt{b})$ has exactly $2(k-6)$ runs. We start with 2 runs from $\beta(\mathtt{a}^{k-2}\mathtt{b})\beta(\mathtt{a}^{k-3}\mathtt{b}) = \mathtt{bba}$, and then, concatenating each $\beta(\mathtt{a}^i\mathtt{b})$ up to $\beta(\mathtt{a}^4\mathtt{b})$ adds 2 new runs each. By counting, we observe that $\beta(\mathtt{aaab}), \beta(\mathtt{aab}), \beta(\mathtt{ab})$ have $2(k-6)$, 4, 4. The boundaries between these words do not yet merge. The word $\beta(\mathtt{ba})$ has exactly 8 runs. The remaining part of the BWT, that is, $\prod_{i=2}^{k} \beta(\mathtt{b}^i\mathtt{a})$, has $4(k-3)+2$ runs. Concatenating each $\beta(\mathtt{b}^2\mathtt{a})$ to $\beta(\mathtt{b}^{k-2}\mathtt{a})$ adds 4 new runs each. The word $\beta(\mathtt{b}^k\mathtt{a})$ adds only one run by $\mathtt{b}$, as it contains an $\mathtt{a}$ that merges with the previous one. Finally, $\beta(\mathtt{b}^k\mathtt{a})$ adds one run. Altogether, we have $2(k-6)+2(k-6)+4+4+8+4(k-3)+2 = 8k-18$, and the claim holds. □

Table 14. Classification of the number of runs obtain in Lemma 47. The total number of runs is $8k-18$.

BBWT of $\overline{W_k^\flat}\mathtt{c}$	Runs
$\beta(\mathtt{a}^{k-2}\mathtt{b}) = \mathtt{b}$	1
$\beta(\mathtt{a}^i\mathtt{b}) = \mathtt{ba}^{k-i-2}$, for all $i \in [4..k-3]$	$2k-12$ but, when merged, $2k-13$
$\beta(\mathtt{aaab}) = \mathtt{bbbbb}(\mathtt{ab})^{k-7}\mathtt{baa}$	$2k-12$
$\beta(\mathtt{aab})=\mathtt{baaba}^{2k-8}$	4
$\beta(\mathtt{ab})= \mathtt{b}^{k-3}\mathtt{aaba}^{2k-6}$	4
$\beta(\mathtt{ba}) = \mathtt{ba}^{k-6}\mathtt{bbbab}^{k-4}\mathtt{ab}^{k-3}\mathtt{a}$	8
$\beta(\mathtt{b}^j\mathtt{a}) = \mathtt{bab}^{2k-2j-2}\mathtt{a}$ for all $j \in [2..k-2]$	$4k-12$
$\beta(\mathtt{b}^{k-1}\mathtt{a})= \mathtt{aab}$	2 but, when merged, 1
$\beta(\mathtt{b}^k\mathtt{a})= \mathtt{a}$	1

Using Lemma 47 above, we can finally obtain the runs of $C_k^\flat = D_k\mathtt{a}$.

Theorem 10. *$\rho(C_k^\flat) = 8k-17$.*

Proof. $C_k^\flat = \mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a} \cdot \left(\prod_{i=2}^{k-2} \mathtt{ab}^i\mathtt{aaab}^i\mathtt{aba}^{i-2}\right) \cdot \mathtt{ab}^{k-1}\mathtt{aaab}^{k-1}\mathtt{a}$. The Lyndon conjugate of $C_k^\flat$ is the smallest conjugate starting with the longest runs of $\mathtt{a}$, thus it is the one starting with $\mathtt{a}^{k-1}$. Therefore, it is obvious that $C_k^\flat$ is not a Lyndon word, then it is Lyndon factorized by an $\mathtt{a}$ and the residual which is D_k. Figure 12 depicts the Lyndon factorization of $C_k^\flat$. Since the lexicographic order between $\mathtt{a}$ and D_k is $\mathtt{a} < D_k$, the runs of $C_k^\flat$ add one run because the first conjugate of D_k from Lemma 38 ends with a $\mathtt{b}$. Therefore, $\rho(C_k^\flat) = \rho(\mathtt{a}) + \rho(D_k) = 8k-17$. □

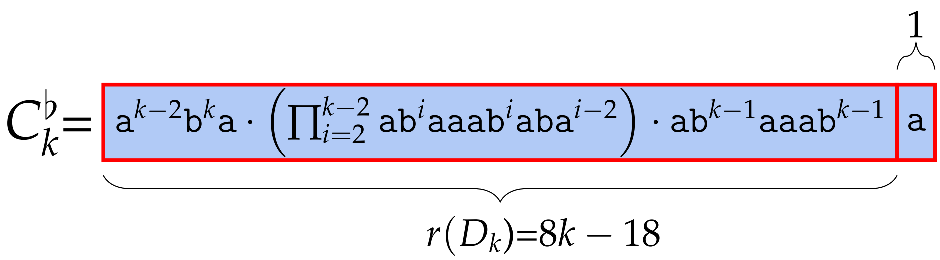

Figure 12. Lyndon factorization of $C_k^\flat$. We obtain $\rho(C_k^\flat)$ by knowing the number of runs of both its Lyndon factors and where these conjugates are sorted in the BBWT. The analysis is in the proof of Theorem 10.

With Lemma 37 and Theorem 10, we determine that the additive sensitivity of ρ for C_k is $\Theta(\log n)$ when deleting the last character.

Theorem 11. *$\rho(C_k^\flat\#)= 8k-16$.*

Proof. $C_k^\flat\# = \mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a} \cdot \left(\prod_{i=2}^{k-2} \mathtt{ab}^i\mathtt{aaab}^i\mathtt{aba}^{i-2}\right) \cdot \mathtt{ab}^{k-1}\mathtt{aaab}^{k-1}\mathtt{a}\#$. $C_k^\flat\#$ is Lyndon factorized into three parts, which are D_k, $\mathtt{a}$ and #, because the lexicographic order of $\mathtt{a}$ is lower than D_k, and moreover # is smaller than both D_k and $\mathtt{a}$. Therefore, the $\rho(C_k^\flat\#) = \rho(\#) + \rho(\mathtt{a}) + \rho(D_k) = 8k - 16$. We show a sketch in Figure 13. □

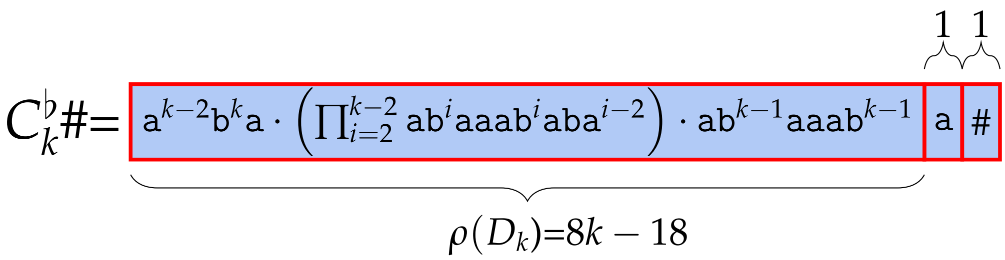

Figure 13. Lyndon factorization of $C_k^\flat\#$. Compared to Figure 12, we have one additional Lyndon factor. The analysis is in proof of Theorem 11.

With Lemma 37 and Theorem 11, we obtain that the additive sensitivity of ρ for C_k is $\Theta(\log n)$ when substituting the last character.

7.2. Editing C_k with a Character Larger than b

Now, we consider the editing operation C_k with a character c that is lexicographically larger than any character in C_k. In this part, we consider two edit operations that add c in the last part of C_k, and substitute the last character of C_k into c.

7.2.1. Appending c to C_k

Now, we prove that adding c to C_k, i.e., C_k becomes $C_k\mathtt{c}$, also adds $\Theta(\sqrt{n})$ in runs in BBWT. $C_k\mathtt{c} = \mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a} \cdot \left(\prod_{i=2}^{k-2} \mathtt{ab}^i\mathtt{aaab}^i\mathtt{aba}^{i-2}\right) \cdot \mathtt{ab}^{k-1}\mathtt{aaab}^{k-1}\mathtt{abc}$. We illustrate C_k in Figure 14. Similar to Section 7.1, we slightly modify E_{k-1} to $\mathtt{ab}^{k-1}\mathtt{abca}^{k-3}$. In this section, we call this modified subword S_{k-1}. The lexicographic order of c is larger than any words in C_k. Thus, $C_k\mathtt{c}$ is a Lyndon word itself. Recall that the runs of a Lyndon word are the same in both the BWT or the BBWT, so we obtain $\rho(C_k\mathtt{c})$ by using BWT with $\mathcal{M}(C_k\mathtt{c})$ the same way we did in previous lemmas.

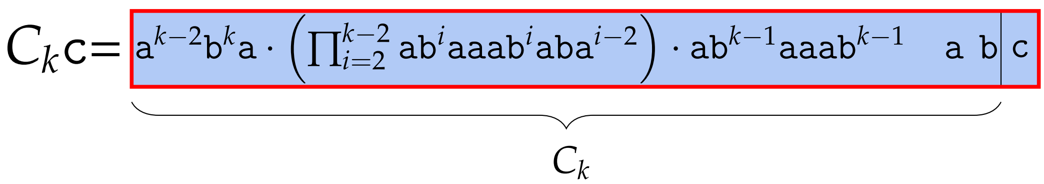

Figure 14. Introducing the Lyndon word $C_k\mathtt{c}$ studied in Section 7.2

Lemma 48. $\beta(\mathtt{a}^{k-2}\mathtt{b}, C_k\mathtt{c}) = \mathtt{c}$.

Proof. The *first* conjugate in $\mathcal{M}(C_k\mathtt{c})$ is $\mathtt{a}^{k-2}\mathtt{ba}P_2 \cdots \mathtt{c}$. The first conjugate must start with the longest run of as. In $C_k\mathtt{c}$, the longest run of a has a length of $k-2$ which is a prefix of itself, and it is obtained by concatenating the suffix $\mathtt{a}^{k-3}$ of S_{k-1} with Q_k, and it is preceded by a c. □

Lemma 49. $\beta(\mathtt{a}^i\mathtt{b}, C_k\mathtt{c}) = \mathtt{ba}^{k-i-2}$ *for all* $i \in [4..k-3]$.

Proof. In $\mathcal{M}(C_k\mathtt{c})$, the conjugates starting with $\mathtt{a}^i\mathtt{b}$ for $i \in [4..k-3]$ are

$$\mathtt{a}^{i-1}P_{i+2} \cdots \mathtt{b} < \mathtt{a}^{i-1}P_{i+3} \cdots \mathtt{a} < \cdots < \mathtt{a}^{i-1}P_{k-1} \cdots \mathtt{a} < \mathtt{a}^{i-1}Q_k \cdots \mathtt{a}.$$

For all $i \in [4..k-3]$, the factor $\mathtt{a}^i\mathtt{b}$ can only be obtained, for all $j \in [i+2..k]$, from the concatenation of the suffix $\mathtt{a}^{i-1}$ of E_{j-1} with prefix $\mathtt{ab}$ of P_j, if $j \in [i+2..k-1]$ or from the concatenation with $\mathtt{a}^{i-1}$ of S_{k-1} with prefix $\mathtt{ab}$ of Q_k. We can sort these conjugates according to the lexicographic order of $\bigcup_{j=i}^{k-1} \mathtt{a}^{i-1}P_j \cup \mathtt{a}^{i-1}Q_k$. Note that all these conjugates end with an $\mathtt{a}$, with the exception of the conjugate starting with $\mathtt{a}^{i-1}P_{i+2}$, since it is here the only occurrence of $\mathtt{ba}^i\mathtt{b}$ can be found. □

Lemma 50. $\beta(\mathtt{aaab}, C_k\mathtt{c}) = \mathtt{bbbbb}(\mathtt{ab})^{k-6}\mathtt{a}$.

Proof. In $\mathcal{M}(C_k\mathtt{c})$, the conjugates starting with $\mathtt{aaab}$ are

$$\begin{aligned} \mathtt{aa}E_2\cdots\mathtt{b} &< \mathtt{aa}E_3\cdots\mathtt{b} < \mathtt{aa}E_4\cdots\mathtt{b} < \mathtt{aa}P_5\cdots\mathtt{b} < \mathtt{aa}E_5\cdots\mathtt{b} \\ &< \mathtt{aa}P_6\cdots\mathtt{a} < \mathtt{aa}E_6\cdots\mathtt{b} < \cdots < \mathtt{aa}P_{k-2}\cdots\mathtt{a} \\ &< \mathtt{aa}E_{k-2}\mathtt{b} < \mathtt{aa}P_{k-1}\cdots\mathtt{a} < \mathtt{aa}S_{k-1}\cdots\mathtt{b} \\ &< \mathtt{aa}Q_k\cdots\mathtt{a}. \end{aligned}$$

Similarly to Lemma 49, $\mathtt{aaab}$ can be obtained from concatenation of the suffix $\mathtt{aa}$ of E_{j-1}, with the prefix $\mathtt{ab}$ of P_j, if $j \in [5..k-1]$, or concatenating $\mathtt{aa}$ of S_{k-1} with prefix $\mathtt{ab}$ of Q_k. On the other hand, there are more conjugates from concatenating suffix $\mathtt{aa}$ of $P_{j'}$ to the prefix $\mathtt{ab}$ of $E_{j'}$, for all $j' \in [2..k-2]$, or with S_{k-1} if $j' = k-1$. All the conjugates starting with $\mathtt{aaab}$ are sorted according to the lexicographic order of the words in $\bigcup_{j=2}^{4}\{\mathtt{aa}E_j\} \cup \{\mathtt{aa}P_5 \cdot \mathtt{aa}E_5\} \cup \bigcup_{j'=6}^{k-2}\{\mathtt{aa}P_{j'} \cdot \mathtt{aa}E_{j'}\} \cup \{\mathtt{aa}P_{k-1} \cdot \mathtt{aa}S_{k-1}\} \cup \{\mathtt{aa}Q_k\}$. Note that all the conjugates starting either with $\mathtt{aa}P_j$, for all $j \in [6..k-1]$, or with $\mathtt{aa}Q_k$, end with $\mathtt{a}$. On the other hand, the conjugates starting either with $\mathtt{aa}P_5$ or with $\mathtt{aa}E_j$, for all $j \in [2..k-2]$ or $\mathtt{aa}S_{k-1}$, end with $\mathtt{b}$. □

Lemma 51. $\beta(\mathtt{aab}, C_k\mathtt{c}) = \mathtt{baaba}^{2k-8}$.

Proof. The conjugates starting with $\mathtt{aab}$ in $\mathcal{M}(C_k\mathtt{c})$ are

$$\begin{aligned} &\mathtt{a}P_2\cdots\mathtt{b} < \mathtt{a}E_2\cdots\mathtt{a} < \mathtt{a}E_3\cdots\mathtt{a} < \mathtt{a}P_4\cdots\mathtt{b} \\ &\quad < \mathtt{a}E_4\cdots\mathtt{a} < \mathtt{a}P_5\cdots\mathtt{a} < \mathtt{a}E_4\cdots\mathtt{a} < \cdots < \mathtt{a}P_{k-1}\cdots\mathtt{a} < \mathtt{a}S_{k-1}\cdots\mathtt{a} < \mathtt{a}Q_k\cdots\mathtt{a}. \end{aligned}$$

Each of the conjugates starting with $\mathtt{aaab}$ from Lemma 50 induces a conjugate starting with $\mathtt{aab}$, obtained by shifting one character on the left $\mathtt{a}$. It follows that all of these conjugates end with $\mathtt{a}$. The other conjugates starting with $\mathtt{aab}$ are the ones obtained by concatenating the suffix $\mathtt{a}$ of E_3 and the prefix $\mathtt{ab}$ of P_4, and the one obtained by concatenating the suffix $\mathtt{a}$ of Q_k and the prefix $\mathtt{ab}$ of P_2. Moreover, both conjugates end with a $\mathtt{b}$. We conclude this proof by sorting lexicographically the conjugates in $\{\mathtt{a}P_2\} \cup \bigcup_{i=2}^{3}\{\mathtt{a}E_i\} \cup \bigcup_{i=4}^{k-2}\{\mathtt{a}P_i \cdot \mathtt{a}E_i\} \cup \{\mathtt{a}P_{k-1} \cdot \mathtt{a}S_{k-1}\} \cup \{\mathtt{a}Q_k\}$. □

Lemma 52. $\beta(\mathtt{ab}, C_k\mathtt{c}) = \mathtt{b}^{k-3}\mathtt{aaba}^{2k-6}\mathtt{b}$.

Proof. The conjugates in $\mathcal{M}(C_k\mathtt{c})$ starting with $\mathtt{ab}$ are

$$\begin{aligned} &\mathtt{ab}^{k-4}P_{k-1}\cdots\mathtt{b} < \mathtt{ab}^{k-5}P_{k-2}\cdots\mathtt{b} < \cdots < \mathtt{ab}P_3\cdots\mathtt{b} \\ &\quad < P_2\cdots\mathtt{a} < E_2\cdots\mathtt{a} < P_3\cdots\mathtt{b} \\ &\quad < E_3\cdots\mathtt{a} < P_4\cdots\mathtt{a} < E_4\cdots\mathtt{a} < \cdots < P_{k-1}\cdots\mathtt{a} < S_{k-1}\cdots\mathtt{a} < Q_k\cdots\mathtt{a} \\ &\quad < \mathtt{abc}\cdots\mathtt{b}. \end{aligned}$$

For all two distinct integers i, i' with $i > i' \geq 0$, we have $\mathtt{ab}^i\mathtt{ab} < \mathtt{ab}^{i'}\mathtt{ab}$. Thus, the first conjugate in lexicographic order starting with ab is the one followed by the longest run of as. The smallest of these conjugates can be found by concatenating the suffix $\mathtt{aba}^{k-4}$ of E_{k-2} with P_{k-1}, followed by the suffix $\mathtt{aba}^{i-3}$ of E_{i-1} concatenated with P_i, for all $i \in [3..k-2]$, taken in decreasing order. By construction of E_i, for all $i \in [2..k-2]$, these conjugates all end with a b. The remaining conjugates starting with ab are exactly those conjugates that have as prefix either P_i or E_i, for all $i \in [2..k-2]$, P_{k-1}, S_{k-1} or Q_k. Note that all of these conjugates are obtained by shifting one character on the left a from the conjugates starting with aab from Lemma 51, with the exception of one starting with P_3. It follows that the latter ends with a b, while all the other conjugates end with a. Finally, the conjugate starting with the prefix abc follows, which ends with b. □

Lemma 53. $\beta(\mathtt{ba}, C_k\mathtt{c}) = \mathtt{a}^{k-6}\mathtt{bbbab}^{k-4}\mathtt{ab}^{k-3}\mathtt{ab}$.

Proof. In $\mathcal{M}(C_k\mathtt{c})$, the conjugates starting with ba are

$$\begin{aligned}\mathtt{ba}^{k-4}P_{k-1}\cdots\mathtt{a} < \mathtt{ba}^{k-5}P_{k-2}\cdots\mathtt{a} < \cdots < \mathtt{baaa}P_6\cdots\mathtt{a} < \mathtt{baa}E_2\cdots\mathtt{b} \\ < \mathtt{baa}E_3\cdots\mathtt{b} < \mathtt{baa}E_4\cdots\mathtt{b} < \mathtt{baa}P_5\cdots\mathtt{a} \\ < \mathtt{baa}E_5\cdots\mathtt{b} < \mathtt{baa}E_6\cdots\mathtt{b} < \cdots < \mathtt{baa}E_{k-2}\cdots\mathtt{b} < \mathtt{baa}S_{k-1}\cdots\mathtt{b} \\ < \mathtt{ba}P_2\cdots\mathtt{b} < \mathtt{ba}P_4\cdots\mathtt{a} < \mathtt{baba}^{k-4}P_{k-1}\cdots\mathtt{b} < \cdots < \mathtt{baba}P_4\cdots\mathtt{b} \\ < \mathtt{bab}P_3\cdots\mathtt{b} < \mathtt{b}P_3\cdots\mathtt{a} < \mathtt{babca}^{k-3}Q_k\cdots\mathtt{b}.\end{aligned}$$

We have as many circular occurrences of ba as the number of maximal runs of b in $C_k\mathtt{c}$. We have *four* cases.

Case 1: one run of bs in P_i, for all $i \in [2..k-1]$,
Case 2: two runs in E_i for all $i \in [2..k-2]$,
Case 3: one run of ba in Q_k,
Case 4: one run of ba in S_{k-1}.

For Case 1, we have one conjugate starting with $\mathtt{baa}E_i$, for each $i \in [2..k-2]$, or $\mathtt{baa}S_{k-1}$. Since each run of bs within each word from $\bigcup_{i=2}^{k-1}\{P_i\}$ is of length of at least 2, all conjugates in Case 1 end with a b.

For Case 2 and all $i \in [2..k-2]$, we can distinguish between two subcases based on where ba starts:

Case 2 (a): a first run of ba in E_i, which has a type of $\mathtt{baba}^{i-2}$ for all $i \in [2..k-2]$,
Case 2 (b): a second run of ba in E_i, which has a type of $\mathtt{ba}^{i-2}$ for all $i \in [2..k-2]$.

- For Case 2 (a), we can see that these conjugates are of the type $\mathtt{baba}^{i-2}P_{i+1}$, for $i \in [2..k-2]$. Analogously to Case 1, each conjugates for Case 2 (a) end with a b. Each conjugate in Case 2 (b) is obtained by shifting two characters on the right each conjugate in Case 2 (a). Therefore, all of these conjugates end with an a and have prefixes of the type $\mathtt{ba}^{i-2}P_{i+1}$, for all $i \in [2..k-2]$.
- For Case 3, the conjugate starting with ba in Q_k has $\mathtt{ba}P_2$ as prefix, and it is preceded by a b.
- Finally, in Case 4, there is one run of ba, having a prefix of $\mathtt{babca}^{k-3}Q_k$, ending with b.
- Only for Case 2 (b) we have conjugates starting with baaaa. Hence, the first conjugate in lexicographic order is the one starting with $\mathtt{ba}^{k-4}P_{k-1}$, followed by those $\mathtt{ba}^{k-5}P_{k-2} < \cdots < \mathtt{baaa}P_6$.

Among the remaining conjugates, those having prefix baaa either start with $\mathtt{baa}P_5$ from Case 2 (b) or $\mathtt{baa}E_i$ from Case 1, for all $i \in [2..k-2]$ or $\mathtt{baa}S_{k-1}$ if $i = k-1$. We can sort these conjugates by following the order of $\bigcup_{i=2}^{4}\{\mathtt{baa}E_i\} \cup \{\mathtt{baa}P_5\} \cup \bigcup_{i=5}^{k-2}\{\mathtt{baa}E_i\} \cup \{\mathtt{baa}S_{k-1}\}$. Then, the remaining conjugates with prefix baa are those starting with $\mathtt{ba}P_2$

from Case 3 and $\mathtt{ba}P_4$ from Case 2 (b). Finally, we focus on the conjugates from Case 2 (a). These conjugates are sorted according to the length of the run of as following the common prefix bab. The last two conjugates left are one starting with $\mathtt{b}P_3$ from Case 2 (b), and the one from Case 4, which is $\mathtt{babca}^{k-3}Q_k$. These two conjugates are already sorted. Since these conjugates are greater than other conjugates, these are the greatest conjugates of $\mathcal{M}(C_k\mathtt{c})$ starting with ba. □

Lemma 54. $\beta(\mathtt{b}^i\mathtt{a}, C_k\mathtt{c}) = \mathtt{ab}^{2k-2i-2}\mathtt{ab}$ *for all* $i \in [2..k-2]$.

Proof. With integer $i \in [2..k-2]$, conjugates in $\mathcal{M}(C_k\mathtt{c})$ with prefix $\mathtt{b}^i\mathtt{a}$ are

$$\begin{aligned} \mathtt{b}^i\mathtt{aa}E_i\cdots\mathtt{a} < \mathtt{b}^i\mathtt{aa}E_{i+1}\cdots\mathtt{b} < \cdots < \mathtt{b}^i\mathtt{aa}E_{k-2}\cdots\mathtt{b} < \mathtt{b}^i\mathtt{aa}S_{k-1}\cdots\mathtt{b} \\ < \mathtt{b}^i\mathtt{a}P_2\cdots\mathtt{b} < \mathtt{b}^i\mathtt{aba}^{k-4}P_{k-1}\cdots\mathtt{b} \\ < \mathtt{b}^i\mathtt{aba}^{k-5}P_{k-2}\cdots\mathtt{b} < \cdots < \mathtt{b}^i\mathtt{aba}^{i-1}P_{i+2}\cdots\mathtt{b} \\ < \mathtt{b}^i\mathtt{aba}^{i-2}P_{i+1}\cdots\mathtt{a} < \mathtt{b}^i\mathtt{abca}^{k-3}Q_k\cdots\mathtt{b}. \end{aligned}$$

With integer $i \in [2..k-2]$, these conjugates are obtained in the following cases.

Case 1: concatenating $\mathtt{b}^i\mathtt{aa}$ of P_j with E_j, for all $j \in [i..k-2]$ or with S_{k-1} if $j = k-1$,
Case 2: concatenating $\mathtt{b}^i\mathtt{aba}^{j-2}$ of E_j with P_{j+1} for all $i \in [2..k-2]$,
Case 3: concatenating $\mathtt{b}^i\mathtt{a}$ of Q_k with P_2,
Case 4: concatenating $\mathtt{b}^i\mathtt{abca}^{k-3}$ of S_{k-1} with Q_k.

We consider the four cases separately. For all $j \in [i..k-1]$, the conjugate starting within P_j (Case 1) has as prefix $\mathtt{b}^i\mathtt{aa}E_j$ if $j \in [i..k-2]$ or $\mathtt{b}^i\mathtt{aa}S_{k-1}$ if $j = k-1$. Also, when $j \in [i..k-2]$, the conjugate starting within E_j (Case 2) has the prefix of $\mathtt{b}^i\mathtt{aba}^{j-2}P_{j+1}$. In addition, the conjugate starting within Q_k(Case 3) has as prefix $\mathtt{b}^i\mathtt{a}P_2$. Finally, the conjugate that begins within S_{k-1} (Case 4) has a prefix of $\mathtt{b}^i\mathtt{abca}^{k-3}$. By construction, we can see that all the conjugates from Case 1 are sorted according to the lexicographic order of the words in $\bigcup_{j=i}^{k-2}\{\mathtt{b}^i\mathtt{aa}E_j\} \cup \{\mathtt{b}^i\mathtt{aa}S_{k-1}\}$; then, we have the conjugate from Case 3. Following, we have the conjugate from Case 2, sorted according to the decreasing length of the run of as following the common prefix $\mathtt{b}^i\mathtt{ab}$. Finally, the conjugate of Case 4 follows. Moreover, we note that only when the run of bs is exactly of length i ends the conjugate with an a. Thus, only conjugates ending with an a are those starting within P_i and E_i, i.e., those with prefixes $\mathtt{b}^i\mathtt{aa}E_i$ and $\mathtt{b}^i\mathtt{aba}^{i-2}P_{i+1}$. □

Lemma 55. $\beta(\mathtt{b}^{k-1}\mathtt{a}, C_k\mathtt{c})$=aba.

Proof. In $\mathcal{M}(C_k\mathtt{c})$, the conjugates with prefix $\mathtt{b}^{k-1}\mathtt{a}$ are

$$\mathtt{b}^{k-1}\mathtt{aa}S_{k-1}\cdots\mathtt{a} < \mathtt{b}^{k-1}\mathtt{a}P_2\cdots\mathtt{b} < \mathtt{b}^{k-1}\mathtt{abca}^{k-3}Q_k\cdots\mathtt{a}.$$

Observe that the only conjugates with prefix $\mathtt{b}^{k-1}\mathtt{a}$ start within P_{k-1}, Q_k and S_{k-1}. These conjugates have prefixes of, respectively, $\mathtt{b}^{k-1}\mathtt{aa}S_{k-1}$, $\mathtt{b}^{k-1}\mathtt{a}P_2$, $\mathtt{b}^{k-1}\mathtt{abca}^{k-3}Q_k$. One can see that these conjugates taken in this order are already sorted, and only the conjugate starting within Q_k ends with b, while the other two have a. □

Lemma 56. $\beta(\mathtt{b}^k\mathtt{a}, C_k\mathtt{c})$=a.

Proof. In $\mathcal{M}(C_k\mathtt{c})$, the conjugate with prefix $\mathtt{b}^k\mathtt{a}$ is $\mathtt{b}^k\mathtt{a}P_2\cdots\mathtt{a}$. The only occurrence of $\mathtt{b}^k\mathtt{a}$ is within Q_k. Since the run of bs is maximal, it ends with a. □

Lemma 57. $\beta(\mathtt{bc}, C_k\mathtt{c}) = \mathtt{a}$.

Proof. In $\mathcal{M}(C_k\texttt{c})$, the conjugate starting with $\texttt{bc}$ is $\texttt{bca}^{k-3}Q_k \cdots \texttt{a}$. The only occurrence of $\texttt{bc}$ is in S_{k-1}, preceded by an $\texttt{a}$. □

Lemma 58. $\beta(\texttt{c}, C_k\texttt{c}) = \texttt{b}$.

Proof. In $\mathcal{M}(C_k\texttt{c})$, the last conjugate is $\texttt{ca}^{k-3}Q_k \cdots \texttt{b}$ since $\texttt{c}$ is biggest character in $C_k\texttt{c}$. The only occurrence of $\texttt{c}$ is in the last character of $C_k\texttt{c}$. Hence, the last conjugate in lexicographic order starts with $\texttt{ca}^{k-3}Q_k$. Since $\texttt{c}$ is preceded by $\texttt{b}$, the conjugate $C_k\texttt{c}$ contributes a $\texttt{b}$ to the BWT. □

The following theorem puts the above lemmas together.

Theorem 12. *$\rho(C_k\texttt{c}) = 8k - 12$, cf. Table 15. It holds that BBWT$(C_k\texttt{c})$ = BWT$(C_k\texttt{c})$ = $\prod_{i=2}^{k-1} \beta(\texttt{a}^{k-i}\texttt{b}) \cdot \prod_{i=1}^{k} \beta(\texttt{b}^i\texttt{a}) \cdot \beta(\texttt{bc}) \cdot \beta(\texttt{c})$.*

Table 15. Classification of the number of runs obtain in Theorem 12. The total number of runs is $8k - 12$.

BWT of $C_k\texttt{c}$	Runs
$\beta(\texttt{a}^{k-2}\texttt{b}) = \texttt{c}$	1
$\beta(\texttt{a}^i\texttt{b}) = \texttt{ba}^{k-i-2}$ for all $i \in [4..k-3]$	$2k - 12$
$\beta(\texttt{aaab}) = \texttt{bbbbb}(\texttt{ab})^{k-6}\texttt{a}$	$2k - 10$
$\beta(\texttt{aab}) = \texttt{baaba}^{2k-8}$	4
$\beta(\texttt{ab}) = \texttt{b}^{k-3}\texttt{aaba}^{2k-6}\texttt{b}$	5
$\beta(\texttt{ba}) = \texttt{a}^{k-6}\texttt{bbbab}^{k-4}\texttt{ab}^{k-3}\texttt{ab}$	8
$\beta(\texttt{b}^i\texttt{a}) = \texttt{ab}^{2k-2i-2}\texttt{ab}$ for all $i \in [2..k-2]$	$4k - 12$
$\beta(\texttt{b}^{k-1}\texttt{a}) = \texttt{aba}$	3
$\beta(\texttt{b}^k\texttt{a}) = \texttt{a}$	1 but when merged 0
$\beta(\texttt{bc}) = \texttt{a}$	1 but when merged 0
$\beta(\texttt{c}) = \texttt{b}$	1

Proof. Every conjugate of $\beta(\texttt{a}^i\texttt{b})$ is smaller than any conjugate of $\beta(\texttt{a}^{i'}\texttt{b})$, for all $1 \le i' \le i \le k-2$. Symmetrically, every conjugate of $\beta(\texttt{b}^j\texttt{a})$ is greater than any conjugate of $\beta(\texttt{b}^{j'}\texttt{a})$, for every $1 \le j' \le j \le k$. Since we considered all the disjoint ranges of conjugates of $C_k\texttt{c}$ based on their common prefix, $\prod_{i=2}^{k-1} \beta(\texttt{a}^{k-i}\texttt{b}) \cdot \prod_{i=1}^{k} \beta(\texttt{b}^i\texttt{a}) \cdot \beta(\texttt{bc}) \cdot \beta(\texttt{c})$ is the BBWT and BWT of $C_k\texttt{c}$.

With the structure of BWT$(C_k\texttt{c})$, we can easily derive its number of runs. The word $\prod_{i=2}^{k-4} \beta(\texttt{a}^{k-i}\texttt{b})$ has exactly $2k - 11$ runs: we start with 1 run from $\beta(\texttt{a}^{k-2}\texttt{b}) = \texttt{c}$, and then concatenating each from $\beta(\texttt{a}^{k-3}\texttt{b})$ to $\beta(\texttt{aaaab})$ adds 2 runs each. By counting, we observe that $\beta(\texttt{aaab}), \beta(\texttt{aab}), \beta(\texttt{ab})$, have $2k - 10$, 4, 5 runs, respectively. The boundaries between these words do not merge. The word $\beta(\texttt{ba})$ has exactly 8 runs. The remaining parts of the BWT $\prod_{i=2}^{k} \beta(\texttt{b}^i\texttt{a})$ have $4(k-3) + 4$ runs: we start adding 4 runs each by concatenating each $\beta(\texttt{bba})$ to $\beta(\texttt{b}^{k-2}\texttt{a})$. And $\beta(\texttt{b}^{k-1}\texttt{a})$ adds 3 runs. On the other hand, the words $\beta(\texttt{b}^k\texttt{a})$ and $\beta(\texttt{bc})$ do not add new runs, as they consist only of an $\texttt{a}$ that merges with the previous one. For the last element, $\beta(\texttt{c})$ adds one run. Altogether, we have $2k - 11 + 2k - 10 + 8 + 4 + 5 + 4k - 12 + 3 + 1 = 8k - 12$, and the claim holds. □

With Lemma 37 and Theorem 12, we obtain that the additive sensitivity of ρ for C_k is $\Theta(\log n)$ when appending a character.

7.2.2. Substituting the Last Position of C_k with c

Here, we focus on the word $C_k^\flat\mathtt{c}$ that we obtain by substituting the last character of C_k with c, which is lexicographically larger than any character in C_k. See Figure 15 for a visualization. The same as Section 7.2.1, E_{k-1} changes to $\mathtt{ab}^{k-1}\mathtt{aca}^{k-3}$, and we refer to it as R_{k-1} below. According to its definition, $C_k^\flat\mathtt{c} = \mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a} \cdot \left(\prod_{i=2}^{k-2} \mathtt{ab}^i\mathtt{aaab}^i\mathtt{aba}^{i-2}\right) \cdot \mathtt{ab}^{k-1}\mathtt{aaab}^{k-1}\mathtt{ac}$. Recall from the proof of Theorem 10 that $C_k^\flat$ is not a Lyndon factor. The Lyndon factors of $C_k^\flat$ are D_k and a. There, we prove that the run of $C_k^\flat$ is $8k-17$. We start with the first observation that $C_k^\flat\mathtt{c}$ is a Lyndon word.

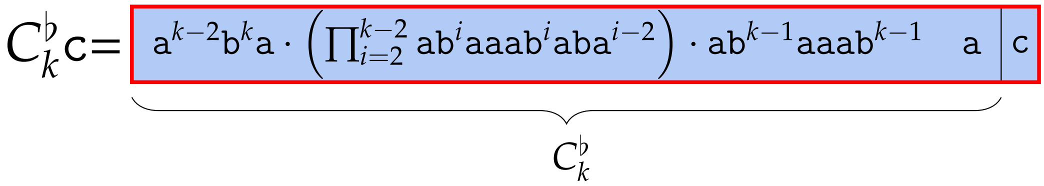

Figure 15. Introducing the Lyndon word $C_k^\flat\mathtt{c}$ studied in Section 7.2.2.

Lemma 59. *$C_k^\flat\mathtt{c}$ is a Lyndon word.*

Proof. The longest run of a has a length of $k-2$, which is a prefix of $C_k^\flat\mathtt{c}$ itself having prefix $\mathtt{a}^{k-2}\mathtt{b}$. Thus, $C_k^\flat\mathtt{c}$ is a Lyndon word. □

Thus, we prove $\rho(C_k^\flat\mathtt{c})$ using the $\mathcal{M}(C_k^\flat\mathtt{c})$ as we did above.

Lemma 60. $\beta(\mathtt{a}^{k-2}\mathtt{b}, C_k^\flat\mathtt{c}) = \mathtt{c}$.

Proof. The first conjugate in $\mathcal{M}(C_k^\flat\mathtt{c})$ is $\mathtt{a}^{k-2}\mathtt{b}^k\mathtt{a}P_2\cdots\mathtt{c}$. The first conjugate in lexicographic order must start with the longest run of as. By the definition of $C_k^\flat$, the longest run of a has length $k-2$, and it is obtained by concatenating the suffix $\mathtt{a}^{k-3}$ of R_{k-1} with Q_k, which is preceded by a c. □

Lemma 61. $\beta(\mathtt{a}^i\mathtt{b}, C_k^\flat\mathtt{c}) = \mathtt{ba}^{k-2-i}$ *for all* $i \in [4..k-3]$.

Proof. All conjugates in $\mathcal{M}(C_k^\flat\mathtt{c})$ starting with the prefix $\mathtt{a}^i\mathtt{b}$ for any $i \in [4..k-3]$ are given below.

$$\mathtt{a}^{i-1}P_{i+2}\cdots\mathtt{b} < \mathtt{a}^{i-1}P_{i+3}\cdots\mathtt{a} < \cdots < \mathtt{a}^{i-1}P_{k-1}\cdots\mathtt{a} < \mathtt{a}^{i-1}Q_k\cdots\mathtt{a}.$$

For all $i \in [4..k-3]$, the factor $\mathtt{a}^i\mathtt{b}$ can only be obtained, for all $j \in [i+2..k-1]$, by concatenating the suffix $\mathtt{a}^{i-1}$ of E_{j-1}, with the prefix ab of P_j, or by concatenating suffix $\mathtt{a}^{k-3}$ of R_{k-1} with the prefix ab of Q_k. We can sort these conjugates according to the lexicographic order of $\bigcup_{j=i}^{k-3}\{\mathtt{a}^{i-1}P_{j+2}\} \cup \{\mathtt{a}^{i-1}Q_k\}$. Note that all these conjugates end with an a, with the exception of the conjugate starting with $\mathtt{a}^{i-1}P_{i+2}$, since it is here the only occurrence of $\mathtt{ba}^i\mathtt{b}$ can be found. □

Lemma 62. $\beta(\mathtt{aaab}, C_k^\flat\mathtt{c}) = \mathtt{bbbbb}(\mathtt{ab})^{k-6}\mathtt{a}$.

Proof. The conjugates in $\mathcal{M}(C_k^\flat\mathtt{c})$ starting with the prefix aaab are

$$\mathtt{aa}E_2\cdots\mathtt{b} < \mathtt{aa}E_3\cdots\mathtt{b} < \mathtt{aa}E_4\cdots\mathtt{b} < \mathtt{aa}P_5\cdots\mathtt{b} < \mathtt{aa}E_5\cdots\mathtt{b}$$
$$< \mathtt{aa}P_6\cdots\mathtt{a} < \mathtt{aa}E_6\cdots\mathtt{b} < \cdots < \mathtt{aa}P_{k-2}\cdots\mathtt{a} < \mathtt{aa}R_{k-1}\cdots\mathtt{b} < \mathtt{aa}Q_k\cdots\mathtt{a}.$$

These conjugates are obtained from the following cases.

Case 1: concatenating suffix aa of P_i with prefix ab of E_i, for all $i \in [2..k-2]$ or with R_{k-1} if $i = k-1$,
Case 2: concatenating suffix aa of E_{i-1} with prefix ab of P_i for all $i \in [5..k-1]$,
Case 3: concatenating suffix aa of R_{k-1} with prefix ab of Q_k.

All these conjugates starting with aaab are sorted according to the lexicographic order of the words in $\bigcup_{i=2}^{4}\{\mathtt{aa}E_i\} \cup \{\mathtt{aa}P_5 \cdot \mathtt{aa}E_5\} \cup \bigcup_{i=6}^{k-2}\{\mathtt{aa}P_i \cdot \mathtt{aa}E_i\} \cup \{\mathtt{aa}P_{k-1} \cdot \mathtt{aa}R_{k-1}\} \cup \{\mathtt{aa}Q_k\}$. Note that all the conjugates starting either with aaP_i , for all $i \in [6..k-1]$ of Case 2, or Case 3 end with a. On the other hand, the conjugates starting either with aaP_5 of Case 2 or Case 1 end with a b. □

Lemma 63. $\beta(\mathtt{aab}, C_k^\flat\mathtt{c}) = \mathtt{baaba}^{2k-8}$.

Proof. The conjugates in $\mathcal{M}(C_k^\flat\mathtt{c})$ that starts with the prefix aab are

$$\begin{aligned}\mathtt{a}P_2\cdots\mathtt{b} &< \mathtt{a}E_2\cdots\mathtt{a} < \mathtt{a}E_3\cdots\mathtt{a} < \mathtt{a}P_4\cdots\mathtt{b} < \mathtt{a}E_4\cdots\mathtt{a} < \mathtt{a}P_5\cdots\mathtt{a}\\ &< \mathtt{a}E_5\cdots\mathtt{a} < \cdots < \mathtt{a}P_{k-2}\cdots\mathtt{a} < \mathtt{a}E_{k-2}\cdots\mathtt{a} < \mathtt{a}P_{k-1}\cdots\mathtt{a} < \mathtt{a}R_{k-1}\cdots\mathtt{a}\\ &< \mathtt{a}Q_k\cdots\mathtt{a}.\end{aligned}$$

Each of the conjugates starting with aaab from Lemma 62 induces a conjugate starting with aab, obtained by shifting one character on the left a. It follows that all of these conjugates end with a. The other conjugates starting with aab are the ones obtained by concatenating suffix a of Q_k with ab of P_2, and another is obtained by concatenating suffix a of E_3 with ab of P_4. Moreover, both conjugates end with a b. We prove our claim by sorting the conjugates according to the lexicographic order of the words in $\{\mathtt{a}P_2 \cdot \mathtt{a}E_2 \cdot \mathtt{a}E_3\} \cup \bigcup_{i=4}^{k-2}\{\mathtt{a}P_i \cdot \mathtt{a}E_i\} \cup \{\mathtt{a}P_{k-1} \cdot \mathtt{a}R_{k-1}\} \cup \{\mathtt{a}Q_k\}$. □

Lemma 64. $\beta(\mathtt{ab}, C_k^\flat\mathtt{c}) = \mathtt{b}^{k-3}\mathtt{aaba}^{2k-6}$.

Proof. In $\mathcal{M}(C_k^\flat\mathtt{c})$, the conjugates which start with prefix ab are

$$\begin{aligned}\mathtt{aba}^{k-4}P_{k-1}\cdots\mathtt{b} &< \mathtt{aba}^{k-5}P_{k-2}\cdots\mathtt{b} < \cdots < \mathtt{ab}P_3\cdots\mathtt{b}\\ &< P_2\cdots\mathtt{a} < E_2\cdots\mathtt{a} < P_3\cdots\mathtt{b} < E_3\cdots\mathtt{a}\\ &< P_4\cdots\mathtt{a} < E_4\cdots\mathtt{a} < \cdots < P_{k-2}\cdots\mathtt{a} < E_{k-2}\cdots\mathtt{a}\\ &< P_{k-1}\cdots\mathtt{a} < R_{k-1}\cdots\mathtt{a} < Q_k\cdots\mathtt{a}.\end{aligned}$$

For all two distinct integers i, i' with $i > i' \geq 0$, we have $\mathtt{aba}^i\mathtt{b} < \mathtt{aba}^{i'}\mathtt{b}$. Thus, the first conjugate in lexicographic order starting with ab is the one which is followed by the longest run of as. The smallest of these conjugates can be found by concatenating the suffix $\mathtt{aba}^{k-4}$ of E_{k-2} with the prefix ab of P_{k-1}, followed by the suffix $\mathtt{aba}^{i-3}$ of E_{i-1} concatenated with the prefix ab of P_i, for all $i \in [3..k-2]$ all taken in decreasing order. By construction of E_i, for all $i \in [2..k-2]$, these conjugates must end with a b. The remaining conjugates starting with ab are exactly those conjugates having as prefix either P_i for all $i \in [2..k-1]$ and $E_{i'}$ for all $i' \in [2..k-2]$ or R_{k-1} and Q_k. Note that all of these conjugates are obtained by shifting one character on the left a from the conjugates starting with aab from Lemma 63, with the exception of one starting with P_3. It follows that the latter ends with a b, while all the other conjugates end with an a. □

Lemma 65. $\beta(\mathtt{ac}, C_k^\flat\mathtt{c}) = \mathtt{b}$.

Proof. In $\mathcal{M}(C_k^\flat \mathtt{c})$, the conjugate that starts with prefix $\mathtt{ac}$ is $\mathtt{aca}^{k-3}Q_k \cdots \mathtt{b}$. The lexicographic order of $\mathtt{c}$ is larger than $\mathtt{b}$ or $\mathtt{a}$, so the prefix $\mathtt{ac}$ is also larger than the prefix $\mathtt{ab}$. $\mathtt{ac}$ is obtained from R_{k-1}, preceded by a $\mathtt{b}$. □

Lemma 66. $\beta(\mathtt{ba}, C_k^\flat \mathtt{c}) = \mathtt{a}^{k-6}\mathtt{bbbab}^{k-4}\mathtt{ab}^{k-3}\mathtt{ab}$.

Proof. In $\mathcal{M}(C_k^\flat \mathtt{c})$, the conjugates starting with the prefix $\mathtt{ba}$ are

$$\begin{aligned}\mathtt{ba}^{k-4}P_{k-1}\cdots\mathtt{a} &< \mathtt{ba}^{k-5}P_{k-2}\cdots\mathtt{a} < \cdots < \mathtt{baaa}P_6\cdots\mathtt{a}\\ &< \mathtt{baa}E_2\cdots\mathtt{b} < \mathtt{baa}E_3\cdots\mathtt{b} < \mathtt{baa}E_4\cdots\mathtt{b} < \mathtt{baa}P_5\cdots\mathtt{a}\\ &< \mathtt{baa}E_5\cdots\mathtt{b} < \mathtt{baa}E_6\cdots\mathtt{b} < \cdots < \mathtt{baa}R_{k-1}\cdots\mathtt{b}\\ &< \mathtt{ba}P_2\cdots\mathtt{b} < \mathtt{ba}P_4\cdots\mathtt{a}\\ &< \mathtt{baba}^{k-4}P_{k-1}\cdots\mathtt{b} < \mathtt{baba}^{k-5}P_{k-2}\cdots\mathtt{b} < \cdots < \mathtt{bab}P_3\cdots\mathtt{b}\\ &< \mathtt{b}P_3\cdots\mathtt{a} < \mathtt{baca}^{k-3}Q_k\cdots\mathtt{b}.\end{aligned}$$

One can notice that we have as many circular occurrences of $\mathtt{ba}$ as the number of maximal runs of $\mathtt{b}$s in $M(C_k^\flat \mathtt{c})$. The conjugates are obtained from the cases below.

Case 1: one run of $\mathtt{b}$s in P_i, for all $i \in [2..k-1]$,
Case 2: two runs in E_i for all $i \in [2..k-2]$,
Case 3: one run in Q_k,
Case 4: one run in R_{k-1}.

For Case 1, we have one conjugate starting with $\mathtt{baa}E_i$ for each $i \in [2..k-1]$. Since each run of $\mathtt{b}$s within each word from $\bigcup_{i=2}^{k-1}\{P_i\}$ is of length of at least 2, all conjugates in Case 1 end with a $\mathtt{b}$.

For Case 2, with integer $i \in [2..k-2]$, we can distinguish between two subcases based on where $\mathtt{ba}$ starts:

Case 2 (a): a first run of $\mathtt{ba}$ in E_i, which has a prefix of $\mathtt{baba}^{i-2}P_{i+1}$ for all $i \in [2..k-2]$,
Case 2 (b): a second run of $\mathtt{ba}$ in E_i, which has a prefix of $\mathtt{ba}^{i-2}P_{i+1}$ for all $i \in [2..k-2]$.

- Similarly to Case 1, all the conjugates in Case 2 (a) end with a $\mathtt{b}$.
- Each conjugate in Case 2 (b) is obtained by shifting two characters on the right each conjugate in Case 2 (a). Therefore, all of these conjugates end with an $\mathtt{a}$.
- For Case 3, the conjugate starting with $\mathtt{ba}$ in Q_k has $\mathtt{ba}P_2$ as a prefix, and it is preceded by a $\mathtt{b}$.
- For Case 4, $\mathtt{ba}$ in R_{k-1} has $\mathtt{baca}^{k-3}$ as a prefix, and it is preceded by a $\mathtt{b}$.
- Observe that only for Case 2 (b) we have conjugates starting with $\mathtt{baaaa}$. Hence, the first conjugate in lexicographic order is the one starting with $\mathtt{ba}^{k-4}P_{k-1}$ followed by $\mathtt{ba}^{k-5}P_{k-2} < \cdots < \mathtt{baaa}P_6$.
- Among the remaining conjugates, those having prefix $\mathtt{baaa}$ either start with $\mathtt{baa}P_5$ from Case 2 (b) or $\mathtt{baa}E_i$ from Case 1 for all $i \in [2..k-1]$. Thus, we can sort them according to the order of the words in $\bigcup_{i=2}^{4}\{\mathtt{baa}E_i\} \cup \{\mathtt{baa}P_5\} \cup \bigcup_{i=5}^{k-2}\{\mathtt{baa}E_i\}$. Then, the remaining conjugates with prefix $\mathtt{baa}$ are those starting with $\mathtt{ba}P_2$ from Case 3 and $\mathtt{ba}P_4$ from Case 2 (b).

Finally, we focus on the conjugates from Case 2 (a). These conjugates are sorted according to the length of the run of $\mathtt{a}$s following the common prefix $\mathtt{bab}$. The last conjugates left are the one starting with $\mathtt{b}P_3$ from Case 2 (b) and the one starting with $\mathtt{baca}^{k-3}$ from Case 4. These conjugates are lexicographically organized and are greater than any other cases, and therefore we analyzed all conjugates. □

Lemma 67. $\beta(\mathtt{b}^i\mathtt{a}, C_k^\flat\mathtt{c}) = \mathtt{ab}^{2k-2i-2}\mathtt{ab}$ *for all* $i \in [2..k-2]$.

Proof. In $\mathcal{M}(C_k^\flat\mathtt{c})$, the conjugates starting with $\mathtt{b}^i\mathtt{a}$ for all $i \in [2..k-2]$ are

$$\mathtt{b}^i\mathtt{aa}E_i\cdots\mathtt{a} < \mathtt{b}^i\mathtt{aa}E_{i+1}\cdots\mathtt{b} < \mathtt{b}^i\mathtt{aa}E_{i+2}\cdots\mathtt{b} < \cdots < \mathtt{b}^i\mathtt{aa}R_{k-1}\cdots\mathtt{b}$$
$$< \mathtt{b}^i\mathtt{a}P_2\cdots\mathtt{b} < \mathtt{b}^i\mathtt{aba}^{k-4}P_{k-1}\cdots\mathtt{b} < \cdots < \mathtt{b}^i\mathtt{aba}^{i-1}P_{i+2}\cdots\mathtt{b}$$
$$< \mathtt{b}^i\mathtt{aba}^{i-2}P_{i+1}\cdots\mathtt{a} < \mathtt{b}^i\mathtt{aca}^{k-3}Q_k\cdots\mathtt{b}.$$

All runs of $\mathtt{b}$s of length of at least $i \in [2..k-2]$ are obtained from the cases below.

Case 1: suffix $\mathtt{b}^i\mathtt{aa}$ in P_j, for all $j \in [i..k-1]$
Case 2: $\mathtt{b}^i\mathtt{aba}^{j-2}$ in E_j for all $j \in [i..k-2]$,
Case 3: $\mathtt{b}^i\mathtt{a}$ in Q_k,
Case 4: $\mathtt{b}^i\mathtt{aca}^{k-3}$ in R_{k-1}.

- Consider the four cases separately. The conjugate starting within P_j (Case 1) has as prefix $\mathtt{b}^i\mathtt{aa}E_j$ if only $j \in [i..k-2]$ or $\mathtt{b}^i\mathtt{aa}R_{k-1}$ if $j = k-1$.
- And for all $j \in [i..k-2]$, the conjugate starting within E_j (Case 2) has as prefix $\mathtt{b}^i\mathtt{aba}^{j-2}P_{j+1}$.
- In addition, the conjugate starting within Q_k (Case 3) has as prefix $\mathtt{b}^i\mathtt{a}P_2$.
- Finally, the conjugate that begins within R_{k-1} (Case 4) has as prefix $\mathtt{b}^i\mathtt{aca}^{k-3}$.

By construction, we have all the conjugates from Case 1 sorted according to the lexicographic order of the words in $\bigcup_{j=i}^{k-2}\{\mathtt{b}^i\mathtt{aa}E_j\} \cup \{\mathtt{b}^i\mathtt{aa}R_{k-1}\}$; then, we have the conjugate from Case 3. Then, the conjugates of Case 2 are sorted according to the decreasing length of the run of $\mathtt{a}$s following the common prefix $\mathtt{b}^i\mathtt{ab}$. Finally, the conjugate of Case 4 follows. Moreover, note that only when the run of $\mathtt{b}$s is exactly of length i the conjugate ends with an $\mathtt{a}$. Thus, only the conjugates ending with an $\mathtt{a}$ are those starting within P_i and E_i, i.e., those with prefix $\mathtt{b}^i\mathtt{aa}E_i$ and $\mathtt{b}^i\mathtt{aba}^{i-2}P_{i+1}$. □

Lemma 68. $\beta(\mathtt{b}^{k-1}\mathtt{a}, C_k^\flat\mathtt{c}) = \mathtt{aba}$.

Proof. In $\mathcal{M}(C_k^\flat\mathtt{c})$, there are exactly *three* conjugates that start with prefix $\mathtt{b}^{k-1}\mathtt{a}$. These are

$$\mathtt{b}^{k-1}\mathtt{aaab}^{k-1}\mathtt{aca}^{k-3}Q_k\cdots\mathtt{a} < \mathtt{b}^{k-1}\mathtt{a}P_2\cdots\mathtt{b} < \mathtt{b}^{k-1}\mathtt{aca}^{k-3}Q_k\cdots\mathtt{a}.$$

Observe that the only conjugates with prefix $\mathtt{b}^{k-1}\mathtt{a}$ start within P_{k-1}, Q_k, and R_{k-1}. These conjugates have prefixes of, respectively, $\mathtt{b}^{k-1}R_{k-1}$, $\mathtt{b}^{k-1}\mathtt{a}P_2$, $\mathtt{b}^{k-1}\mathtt{aca}^{k-3}Q_k$. One can see that these conjugates taken in this order are already sorted, and only the conjugate starting within Q_k ends with $\mathtt{b}$, while the other two end with $\mathtt{a}$. □

Lemma 69. $\beta(\mathtt{b}^k\mathtt{a}, C_k^\flat\mathtt{c}) = \mathtt{a}$.

Proof. In $\mathcal{M}(C_k^\flat\mathtt{c})$, only *one* conjugate starts with a prefix of $\mathtt{b}^k\mathtt{a}$ and it is $\mathtt{b}^k\mathtt{a}P_2\cdots\mathtt{a}$. The only occurrence of $\mathtt{b}^k\mathtt{a}$ is within Q_k, preceded by $\mathtt{a}$. □

Lemma 70. $\beta(\mathtt{c}, C_k^\flat\mathtt{c}) = \mathtt{a}$.

Proof. The *last* conjugate in $\mathcal{M}(C_k^\flat\mathtt{c})$ that starts with prefix $\mathtt{c}$ is $\mathtt{ca}^{k-3}Q_k\cdots\mathtt{a}$. The last conjugate in lexicographic order that starts with $\mathtt{c}$ occurs in R_{k-1}. Since $\mathtt{c}$ is preceded by an $\mathtt{a}$, it ends with $\mathtt{a}$. □

The following theorem puts the above lemmas together.

Theorem 13. *$\rho(C^{\flat}_k\mathtt{c}) = 8k - 13$, cf. Table 16. The BBWT of $C^{\flat}_k\mathtt{c}$ is BBWT($C^{\flat}_k\mathtt{c}$) = $\prod_{i=2}^{k-1} \beta(\mathtt{a}^{k-i}\mathtt{b}) \cdot \beta(\mathtt{ac}) \cdot \prod_{i=1}^{k} \beta(\mathtt{b}^i\mathtt{a}) \cdot \beta(\mathtt{c})$.*

Table 16. Classification of the number of runs obtained in Theorem 13. The total number of runs is $8k - 13$.

BWT of $C^{\flat}_k\mathtt{c}$	**Runs**
$\beta(\mathtt{a}^{k-2}\mathtt{b}) = \mathtt{c}$	1
$\beta(\mathtt{a}^i\mathtt{b}) = \mathtt{ba}^{k-2-i}$ for all $i \in [4..k-3]$	$2k - 12$
$\beta(\mathtt{aaab}) = \mathtt{bbbbb}(\mathtt{ab})^{k-6}\mathtt{a}$	$2k - 10$
$\beta(\mathtt{aab}) = \mathtt{baaba}^{2k-8}$	4
$\beta(\mathtt{ab}) = \mathtt{b}^{k-3}\mathtt{aaba}^{2k-6}$	4
$\beta(\mathtt{ac}) = \mathtt{b}$	1
$\beta(\mathtt{ba}) = \mathtt{a}^{k-6}\mathtt{bbbab}^{k-4}\mathtt{ab}^{k-3}\mathtt{ab}$	8
$\beta(\mathtt{b}^i\mathtt{a}) = \mathtt{ab}^{2k-2i-2}\mathtt{ab}$ for all $i \in [2..k-2]$	$4k - 12$
$\beta(\mathtt{b}^{k-1}\mathtt{a}) = \mathtt{aba}$	3
$\beta(\mathtt{b}^k\mathtt{a}) = \mathtt{a}$	1 but, when merged, 0
$\beta(\mathtt{c}) = \mathtt{a}$	1 but, when merged, 0

Proof. Every conjugate contributing a character to $\beta(\mathtt{a}^i\mathtt{b})$ is smaller than any conjugate of $\beta(\mathtt{a}^{i'}\mathtt{b})$, for all $1 \leq i' \leq i \leq k-2$. Symmetrically, every conjugate contributing a character to $\beta(\mathtt{b}^j\mathtt{a})$ is greater than any conjugate of $\beta(\mathtt{b}^{j'}\mathtt{a})$, for every $1 \leq j' \leq j \leq k$. Since we considered all the disjoint ranges of conjugates of $C_k\mathtt{c}$ based on their common prefix, $\prod_{i=2}^{k-1} \beta(\mathtt{a}^{k-i}\mathtt{b}) \cdot \beta(\mathtt{ac}) \cdot \prod_{i=1}^{k} \beta(\mathtt{b}^i\mathtt{a}) \cdot \beta(\mathtt{c})$ is the BBWT and BWT of $C^{\flat}_k\mathtt{c}$.

With the structure of BWT($C^{\flat}_k\mathtt{c}$), we can easily derive its number of runs. The word $\prod_{i=2}^{k-4} \beta(\mathtt{a}^{k-i}\mathtt{b})$ has exactly $2k - 11$ runs: we start with 1 run from $\beta(\mathtt{a}^{k-2}\mathtt{b}) = \mathtt{c}$, and then concatenating each from $\beta(\mathtt{a}^{k-3}\mathtt{b})$ to $\beta(\mathtt{a}^4\mathtt{b})$ adds 2 runs each. By counting, we observe that $\beta(\mathtt{aaab})$, $\beta(\mathtt{aab})$, and $\beta(\mathtt{ab})$ contribute $2k - 10$, 4, and 4 runs, respectively. The boundaries between these words do not merge. The conjugates in $\beta(\mathtt{ac})$ and $\beta(\mathtt{ba})$ contribute with 1 and 8 runs each. The remaining parts of the BWT $\prod_{i=2}^{k} \beta(\mathtt{b}^i\mathtt{a})$ contribute $4(k-3) + 3$ runs: we start adding 4 runs each by concatenating each $\beta(\mathtt{bba})$ to $\beta(\mathtt{b}^{k-2}\mathtt{a})$. And $\beta(\mathtt{b}^{k-1}\mathtt{a})$ adds 3 runs. $\beta(\mathtt{b}^k\mathtt{a})$ and $\beta(\mathtt{c})$ do not add new runs, as they consist only of an $\mathtt{a}$ that merges with the previous one. The last part $\beta(\mathtt{c})$ contributes one run. In total, we have $2k - 11 + 2k - 10 + 4 + 4 + 1 + 8 + 4k - 12 + 3 = 8k - 13$, and the claim holds. □

8. Conclusions

In this article, we analyzed the sensitivity of the Burrows–Wheeler Transform (BWT) and its bijective variant (BBWT) to single-character edits. We extended previous work on the BWT by a four-character alphabet setting and an alphabet reordering. Our findings reveal that BWT and BBWT exhibit similar sensitivity characteristics, with compression size changes that can follow a multiplicative logarithmic or additive square-root growth. These insights clarify that the BWT and BBWT are not robust repetitiveness measures, which is a crucial property for data compression applications. As future work, we would like to find positions in a word for which we can predict the compression size changes when editing that position. That would allow us to design algorithms to improve the compression power of BWT/BBWT by editing the word in a way that minimizes the compression size changes.

Author Contributions: Conceptualization, D.K.; Writing—original draft, H.J. All authors have read and agreed to the published version of the manuscript.

Funding: JSPS KAKENHI Grant Number 23H04378 and Yamanashi Wakate Grant Number 2291

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Burrows, M.; Wheeler, D.J. *A Block Sorting Lossless Data Compression Algorithm*; Technical Report 124; Digital Equipment Corporation: Palo Alto, CA, USA, 1994.
2. Ferragina, P.; Manzini, G. Opportunistic Data Structures with Applications. In Proceedings of the 41st Annual Symposium on Foundations of Computer Science, Redondo Beach, CA, USA, 12–14 November 2000; pp. 390–398. [CrossRef]
3. Gagie, T.; Navarro, G.; Prezza, N. Optimal-Time Text Indexing in BWT-runs Bounded Space. In Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, LA, USA, 7–10 January 2018; pp. 1459–1477. [CrossRef]
4. Gagie, T.; Navarro, G.; Prezza, N. Fully Functional Suffix Trees and Optimal Text Searching in BWT-Runs Bounded Space. *J. ACM* **2020**, *67*, 2:1–2:54. [CrossRef]
5. Bertram, N.; Fischer, J.; Nalbach, L. Move-r: Optimizing the r-index. In Proceedings of the 22nd International Symposium on Experimental Algorithms (SEA 2024), Vienna, Austria, 23–26 July 2024; Volume 301, pp. 1:1–1:19. [CrossRef]
6. Cobas, D.; Gagie, T.; Navarro, G. A Fast and Small Subsampled R-Index. In Proceedings of the 32nd Annual Symposium on Combinatorial Pattern Matching (CPM 2021), Wrocław, Poland, 5–7 July 2021; Volume 191, pp. 13:1–13:16. [CrossRef]
7. Arakawa, Y.; Navarro, G.; Sadakane, K. Bi-Directional r-Indexes. In Proceedings of the 33rd Annual Symposium on Combinatorial Pattern Matching (CPM 2022), Prague, Czech Republic, 27–29 June 2022; Volume 223, pp. 11:1–11:14. [CrossRef]
8. Shivakumar, V.S.; Ahmed, O.Y.; Kovaka, S.; Zakeri, M.; Langmead, B. Sigmoni: Classification of nanopore signal with a compressed pangenome index. *Bioinformatics* **2024**, *40*, i287–i296. [CrossRef]
9. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef] [PubMed]
10. Li, R.; Yu, C.; Li, Y.; Lam, T.W.; Yiu, S.; Kristiansen, K.; Wang, J. SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics* **2009**, *25*, 1966–1967. [CrossRef] [PubMed]
11. Ferragina, P.; Manzini, G.; Muthukrishnan, S.M. *The Burrows–Wheeler Transform: Ten Years Later*; DIMACS: Piscataway, NJ, USA, 2004.
12. Gagie, T.; Manzini, G.; Navarro, G.; Stoye, J. 25 Years of the Burrows–Wheeler Transform (Dagstuhl Seminar 19241). *Dagstuhl Rep.* **2019**, *9*, 55–68.
13. Adjeroh, D.; Bell, T.; Mukherjee, A. *The Burrows–Wheeler Transform: Data Compression, Suffix Arrays, and Pattern Matching*; Springer: Berlin/Heidelberg, Germany, 2008.
14. Gessel, I.M.; Reutenauer, C. Counting Permutations with Given Cycle Structure and Descent Set. *J. Comb. Theory, Ser. A* **1993**, *64*, 189–215.
15. Schindler, M. A fast block-sorting algorithm for lossless data compression. In Proceedings of the DCC '97. Data Compression Conference, Snowbird, UT, USA, 25–27 March 1997; p. 469. [CrossRef]
16. Mantaci, S.; Restivo, A.; Rosone, G.; Sciortino, M. An extension of the Burrows–Wheeler Transform. *Theor. Comput. Sci.* **2007**, *387*, 298–312. [CrossRef]
17. Kufleitner, M. On Bijective Variants of the Burrows–Wheeler Transform. In Proceedings of the Prague Stringology Conference 2009 (PSC 2009), Prague, Czech Republic, 31 August–2 September 2009; pp. 65–79.
18. Daykin, J.W.; Smyth, W.F. A bijective variant of the Burrows–Wheeler Transform using V-order. *Theor. Comput. Sci.* **2014**, *531*, 77–89. [CrossRef]
19. Daykin, J.W.; Groult, R.; Guesnet, Y.; Lecroq, T.; Lefebvre, A.; Léonard, M.; Prieur-Gaston, É. Binary block order Rouen Transform. *Theor. Comput. Sci.* **2016**, *656*, 118–134. [CrossRef]
20. Gil, J.Y.; Scott, D.A. A Bijective String Sorting Transform. *arXiv* **2012**, arXiv:1201.3077. .. [CrossRef]
21. Likhomanov, K.M.; Shur, A.M. Two Combinatorial Criteria for BWT Images. In *Computer Science–Theory and Applications* ; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6651, pp. 385–396. [CrossRef]
22. Giuliani, S.; Lipták, Z.; Masillo, F.; Rizzi, R. When a dollar makes a BWT. *Theor. Comput. Sci.* **2021**, *857*, 123–146. [CrossRef]
23. Giuliani, S.; Lipták, Z.; Masillo, F. When a Dollar in a Fully Clustered Word Makes a BWT. *Ceur Workshop Proc.* **2022**, *3284*, 122–135.
24. Giuliani, S.; Inenaga, S.; Lipták, Z.; Romana, G.; Sciortino, M.; Urbina, C. Bit Catastrophes for the Burrows–Wheeler Transform. In *Developments in Language Theory* ; Springer: Berlin/Heidelberg, Germany, 2023; Volume 13911, pp. 86–99. [CrossRef]

25. Akagi, T.; Funakoshi, M.; Inenaga, S. Sensitivity of string compressors and repetitiveness measures. *Inf. Comput.* **2023**, *291*, 104999. [CrossRef]
26. Lagarde, G.; Perifel, S. Lempel-Ziv: A "one-bit catastrophe" but not a tragedy. In Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, LA, USA, 7–10 January 2018; pp. 1478–1495. [CrossRef]
27. Ziv, J.; Lempel, A. A universal algorithm for sequential data compression. *IEEE Trans. Inf. Theory* **1977**, *23*, 337–343. [CrossRef]
28. Ziv, J.; Lempel, A. Compression of individual sequences via variable-rate coding. *IEEE Trans. Inf. Theory* **1978**, *24*, 530–536. [CrossRef]
29. Kempa, D.; Prezza, N. At the roots of dictionary compression: String attractors. In Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, Los Angeles, CA, USA, 25–29 June 2018; pp. 827–840. [CrossRef]
30. Navarro, G.; Ochoa, C.; Prezza, N. On the Approximation Ratio of Ordered Parsings. *IEEE Trans. Inf. Theory* **2021**, *67*, 1008–1026. [CrossRef]
31. Nakashima, Y.; Köppl, D.; Funakoshi, M.; Inenaga, S.; Bannai, H. Edit and Alphabet-Ordering Sensitivity of Lex-Parse. In Proceedings of the 49th International Symposium on Mathematical Foundations of Computer Science (MFCS 2024), Bratislava, Slovakia, 26–30 August 2024; Volume 306, pp. 75:1–75:15. [CrossRef]
32. Blumer, A.; Blumer, J.; Haussler, D.; Ehrenfeucht, A.; Chen, M.T.; Seiferas, J.I. The Smallest Automaton Recognizing the Subwords of a Text. *Theor. Comput. Sci.* **1985**, *40*, 31–55. [CrossRef]
33. Fujimaru, H.; Nakashima, Y.; Inenaga, S. On Sensitivity of Compact Directed Acyclic Word Graphs. In *Combinatorics on Words* ; Springer: Berlin/Heidelberg, Germany, 2023; Volume 13899, pp. 168–180. [CrossRef]
34. Olbrich, J.; Ohlebusch, E.; Büchler, T. Generic Non-recursive Suffix Array Construction. *ACM Trans. Algorithms* **2024**, *20*, 18. [CrossRef]
35. Bannai, H.; Kärkkäinen, J.; Köppl, D.; Piątkowski, M. Constructing and Indexing the Bijective and Extended Burrows–Wheeler Transform. *Inf. Comput.* **2024**, *297*, 1–30. [CrossRef]
36. Badkobeh, G.; Bannai, H.; Köppl, D. Bijective BWT based Compression Schemes. In *String Processing and Information Retrieval* ; Springer: Berlin/Heidelberg, Germany, 2024; Volume 14899, pp. 16–25. [CrossRef]
37. Biagi, E.; Cenzato, D.; Lipták, Z.; Romana, G. On the Number of Equal-Letter Runs of the Bijective Burrows–Wheeler Transform. In Proceedings of the International Conference on Information and Communication Technology for Competitive Strategies (ICTCS), Jaipur, India, 8–9 December 2023; Volume 3587, pp. 129–142.
38. Lyndon, R.C. On Burnside's Problem. *Trans. Am. Math. Soc.* **1954**, *77*, 202–215.
39. Chen, K.T.; Fox, R.H.; Lyndon, R.C. Free differential calculus, IV. The quotient groups of the lower central series. *Ann. Math.* **1958**, *68*, 81–95.
40. Lothaire, M. *Combinatorics on Words*, 2 ed.; Cambridge Mathematical Library, Cambridge University Press: Cambridge, UK, 1997.
41. Saari, K. Lyndon words and Fibonacci numbers. *J. Comb. Theory, Ser. A* **2014**, *121*, 34–44. [CrossRef]
42. De Luca, A.; Mignosi, F. Some Combinatorial Properties of Sturmian Words. *Theor. Comput. Sci.* **1994**, *136*, 361–385. [CrossRef]
43. Mantaci, S.; Restivo, A.; Sciortino, M. Burrows–Wheeler transform and Sturmian words. *Inf. Process. Lett.* **2003**, *86*, 241–246. [CrossRef]
44. Christodoulakis, M.; Iliopoulos, C.S.; Ardila, Y.J.P. Simple Algorithm for Sorting the Fibonacci String Rotations. In Proceedings of the SOFSEM, Merin, Czech Republic, 21–27 January 2006; Volume 3831, pp. 218–225. [CrossRef]
45. Séébold, P. Fibonacci Morphisms and Sturmian Words. *Theor. Comput. Sci.* **1991**, *88*, 365–384. [CrossRef]
46. Giuliani, S.; Inenaga, S.; Lipták, Z.; Prezza, N.; Sciortino, M.; Toffanello, A. Novel Results on the Number of Runs of the Burrows–Wheeler-Transform. In Proceedings of the SOFSEM, Bolzano-Bozen, Italy, 25–29 January 2021; Volume 12607, pp. 249–262. [CrossRef]
47. Boucher, C.; Cenzato, D.; Lipták, Z.; Rossi, M.; Sciortino, M. r-Indexing the eBWT. In Proceedings of the SPIRE, Lille, France, 4–6 October 2021; Volume 12944, pp. 3–12. [CrossRef]