Insertions Yielding Equivalent Double Occurrence Words

Daniel A. Cruz[*]**, Margherita Maria Ferrari, Nataša Jonoska**

University of South Florida

Tampa FL, 33620, USA

{dcruz, mmferrari, jonoska}@mail.usf.edu

Lukas Nabergall

University of Waterloo

Waterloo ON, N2L 3G1, Canada

lnaberga@uwaterloo.ca

Masahico Saito

University of South Florida

Tampa FL, 33620, USA

saito@usf.edu

Abstract. A double occurrence word (DOW) is a word in which every symbol appears exactly twice; two DOWs are equivalent if one is a symbol-to-symbol image of the other. We consider the so called repeat pattern ($\alpha\alpha$) and the return pattern ($\alpha\alpha^R$), with gaps allowed between the α's. These patterns generalize square and palindromic factors of DOWs, respectively. We introduce a notion of inserting repeat/return words into DOWs and study how two distinct insertions into the same word can produce equivalent DOWs. Given a DOW w, we characterize the structure of w which allows two distinct insertions to yield equivalent DOWs. This characterization depends on the locations of the insertions and on the length of the inserted repeat/return words and implies that when one inserted word is a repeat word and the other is a return word, then both words must be trivial (i.e., have only one symbol). The characterization also introduces a method to generate families of words recursively.

[*]Address for correspondence: University of South Florida, Tampa FL, 33620, USA

1.   Introduction

A word w over an alphabet Σ is a double occurrence word (DOW) if each element of Σ appears either zero or two times. DOWs have been studied in relation to knot theory [4, 10, 19], mathematical logic [8], and algebraic combinatorics [17]. DOWs are also known as Gauss words and are closely related to linear diagrams, chord diagrams, and circle graphs. In the context of genomics, DOWs and operations on DOWs have been used in studies of DNA rearrangement [2, 3, 9, 11]. By modeling the DNA rearrangement process using DOWs, it was observed that over 95% of the scrambled genome of the ciliate *Oxytricha trifallax* could be described by iterative insertions of the "repeat pattern" ($\alpha\alpha$) and the "return pattern" ($\alpha\alpha^R$) [5]. Roughly speaking, a *pattern* is a sequence of variables, and we say that an instance of a pattern *appears in a word* w if each variable of the pattern can be mapped to a non-empty factor of w [11]. The repeat pattern $\alpha\alpha$ generalizes square factors while the return pattern $\alpha\alpha^R$ generalizes palindromic factors. We refer to instances of the repeat and return pattern as *repeat words* and *return words*, respectively [1, 5, 11].

   Patterns in DNA rearrangement are discussed in [5], while transformations on DOWs where instances of patterns are deleted or inserted are considered in [11]. In studies of DNA rearrangement, it has been observed that the insertion of a repeat or return pattern may have evolutionary significance [6], and the process of obtaining one word from another by the insertion of a repeat or return word may be of interest. Relatedly, similar operations on words have been studied with applications to computational linguistics and natural language processing. In the literature, four so-called edit operations are primarily considered: insertion of a symbol, deletion of a symbol, substitution of one symbol with another, and transposition of two adjacent symbols [12, 15, 16]. The pattern-based word transformations considered here may be regarded as a generalization of these edit operations.

   Here we define a notion of inserting repeat and return words in DOWs at prespecified indices. We consider equivalence classes of DOWs where two DOWs are equivalent if one is obtained from the other by a symbol-to-symbol morphism. Equivalent words correspond to the same chord diagrams, as well as isomorphic assembly graphs. The main question considered here is under which conditions two distinct insertions into the same word can produce equivalent DOWs. A pair of insertions in a given DOW falls in one of the three types: interleaving, nested, and sequential (see Section 3 for definitions). The paper characterizes the words that yield the equivalent results in each of these situations.

2.   Background

An *ordered alphabet* Σ is a countable set with a linear order that is bounded below which can naturally be identified with the set of natural numbers $\mathbb{N} = \{1, 2, \ldots\}$, so we set $\Sigma = \mathbb{N}$ throughout the rest of the paper. For $n \in \mathbb{N}$, we denote $\{1, 2, \ldots, n\}$ by $[n]$. For the remainder of the paper, we reserve the letters a, b as *symbols* in Σ, and reserve the letters $s, t, u, \ldots, z$ as "words" in Σ^* (defined below).

   We use standard definitions and conventions (e.g., [7, 8, 13, 11]). A *word* w over Σ is a finite sequence of symbols $a_1 \cdots a_n$ in Σ; the *length of w*, denoted $|w|$, is n. The set of all words over Σ is denoted by Σ^* and includes the empty word ϵ whose length is 0; and $\Sigma^+ = \Sigma^* \setminus \{\epsilon\}$. The set of all symbols $\{a_1, \ldots, a_n\}$ comprising w is denoted by $\Sigma[w]$. The *reverse of* $w = a_1 a_2 \cdots a_n$ ($a_i \in \Sigma$) is the word $w^R = a_n \cdots a_2 a_1$. The word v is a *factor* of the word w, denoted $v \sqsubseteq w$, if $\exists w_1, w_2 \in \Sigma^*$

such that $w = w_1 v w_2$; if $w_1 = \epsilon$ then v is a *prefix* of w, while if $w_2 = \epsilon$ then v is a *suffix* of w; the word $w_1 w_2$ is denoted by $w - v$. If $|v| < |w|$, then v is a *proper* factor, prefix, or suffix as appropriate. The set of common factors of $u, v \in \Sigma^*$ is denoted by $u \cap v$. Let $1 \leq d \leq n = |w|$, and write $w = a_1 a_2 \cdots a_n$. The prefix (suffix resp.) of w with length d is denoted $\mathrm{pre}(w, d) = a_1 a_2 \cdots a_d$ ($\mathrm{suf}(w, d) = a_{n-d+1} \cdots a_{n-1} a_n$ resp.).

A word w in Σ^* is a *double occurrence word* (DOW) if every symbol in w appears either zero or twice. We use Σ_{DOW} to denote the set of all double occurrence words over Σ. *Single occurrence words* (SOWs) denoted Σ_{SOW} are nonempty words with distinct symbols. Given $w \in \Sigma_{DOW}$, $|w|/2$ is the *size of w*. Given $u \in \Sigma_{SOW}$, we say that uu is a *repeat word in w* if $w = z_1 u z_2 u z_3$ for some $z_1, z_2, z_3 \in \Sigma^*$. Similarly, uu^R is a *return word in w* if $w = z_1 u z_2 u^R z_3$ for some $z_1, z_2, z_3 \in \Sigma^*$.

A morphism f on Σ^* induced by a bijection (symbol-to-symbol map) on Σ is called an *equivalence map*. We write $w_1 \sim w_2$ if there is an equivalence map f such that $f(w_1) = w_2$. The relation $\sim$ is an equivalence on Σ^*. A word $w = a_1 a_2 \cdots a_n$ is in *ascending order* if a_1 is the least element in the alphabet and the first appearance of a symbol is one greater than the largest of all preceding symbols in the word [3]. Since a word in ascending order is unique in the considered alphabet, we take words in ascending order as *class representatives* of the equivalence classes determined by the relation $\sim$.

Example 2.1. Consider the words $w = 121323$ and $w' = 131232$. Note that $w \sim w'$, but w is in ascending order while the word w' is not because the symbol 3 appears before 2.

Definition 2.2. Let $w = z_1 z_2 z_3 \in \Sigma_{DOW}$ for some $z_1, z_2, z_3 \in \Sigma^*$ be in ascending order. Let u be a SOW over $\Sigma \setminus \Sigma[w]$ which is in ascending order and $|u| = \nu$. Suppose k and ℓ are such that $k - 1 = |z_1|$ and $\ell - 1 = |z_1 z_2|$. Then

- $w' = z_1 u z_2 u z_3$ is obtained from w by a *repeat insertion* denoted $w' = w \star \rho(\nu, k, \ell)$, and
- $w' = z_1 u z_2 u^R z_3$ is obtained from w by a *return insertion* denoted $w' = w \star \tau(\nu, k, \ell)$.

We do not specify the word u in the notation of repeat and return insertions because the inserted word has distinct symbols from w and consists of symbols immediately following the largest symbol of w; hence it is uniquely determined by its length.

Example 2.3. Let $w = 1232314554$, then

$$w \star \rho(2, 4, 6) = 123\underline{67}23\underline{67}14554 = w_1 \qquad w \star \tau(2, 7, 11) = 12323\underline{167}4554\underline{76} = w_3$$

$$w \star \rho(2, 2, 4) = 1\underline{67}23\underline{67}2314554 = w_2 \qquad w \star \tau(2, 9, 9) = 1232314\underline{567}7654 = w_4$$

Observe that $w_1 \sim w_2$ and $w_3 \sim w_4$ but $w_1 \nsim w_3$.

Let v be a repeat or return word in $w \in \Sigma_{DOW}$; we write $v = uu'$ where $u' = u$ if v is a repeat word and $u' = u^R$ if v is a return word. For the rest of the paper, we use u' for a SOW u to denote u or u^R as is appropriate in the context of repeat and return words. We use the notation $w \star \mathfrak{I}(\nu, k, \ell)$ to indicate that the insertion is either a repeat insertion or a return insertion. We denote by *Rep* (*Ret* resp.) the set of all repeat insertions (return insertions resp.), and we write $\mathfrak{I} \in \mathrm{Rep}$ ($\mathfrak{I} \in \mathrm{Ret}$ resp.) to indicate that the insertion is a repeat insertion (return insertion resp.). Observe that

$(w \star \Im(\nu, k, \ell))^R \sim w^R \star \Im(\nu, |w| - \ell + 2, |w| - k + 2)$. We say that uu or uu^R is a *trivial* repeat or return word, respectively, if $|u| = 1$, and an insertion $\Im(1, k, \ell)$ is called *trivial*. A trivial repeat insertion into w is also a trivial return insertion, so we focus on nontrivial insertions. As a convention, if $k = \ell$ we set $w \star \Im(\nu, k, k) = z_1 uu' z_2 z_3$ as shown in w_4 in Example 2.3.

Remark 2.4. Suppose $w \star \Im_1(\nu_1, k_1, \ell_1) \sim w \star \Im_2(\nu_2, k_2, \ell_2)$, then $\nu_1 = \nu_2 = \nu$, and because the inserted word has distinct symbols from w, by definition we have that $u_1 u_1'$ inserted by $\Im_1$ and $u_2 u_2'$ by $\Im_2$ are such that $u_1 = u_2 = u$ (and $u_1', u_2' \in \{u, u^R\}$). For all $w' \sim w$ we also have that $w' \star \Im_1(\nu, k_1, \ell_1) \sim w' \star \Im_2(\nu, k_2, \ell_2)$. This follows from the fact that if $f(w \star \Im_1(\nu, k_1, \ell_1)) = w \star \Im_2(\nu, k_2, \ell_2)$ and $g(w') = w$ for equivalence maps f and g, then $f(g(w') \star \Im_1(\nu, k_1, \ell_1)) = g(w') \star \Im_2(\nu, k_2, \ell_2)$.

Considering the above remark, we have the following definition for equality of two insertions.

Definition 2.5. We say that two nontrivial insertions $\Im_1(\nu, k_1, \ell_1)$ and $\Im_2(\nu, k_2, \ell_2)$ are *equal for a DOW* w if $w \star \Im_1(\nu, k_1, \ell_1) \sim w \star \Im_2(\nu, k_2, \ell_2)$ and $(k_1, \ell_1) = (k_2, \ell_2)$.

It becomes a natural question to consider the situations when two distinct (unequal) nontrivial insertions into w yield equivalent words.

The following three results are used repeatedly throughout Section 3. The first generalizes a well known lemma by Lyndon and Schützenberger [14]; it describes the structures of two equivalent words where a suffix of one word is a prefix of the other.

Lemma 2.6. Let $s \in \Sigma^+$ and $t, z \in \Sigma^*$ such that $f(sz) = zt$ for an equivalence map f. Then, $s = s_1 s_2$ with $z = f(s) \cdots f^{h-1}(s) f^h(s_1)$, and $t = f^h(s_2) f^{h+1}(s_1)$ for $h = \lceil |z|/|s| \rceil$.
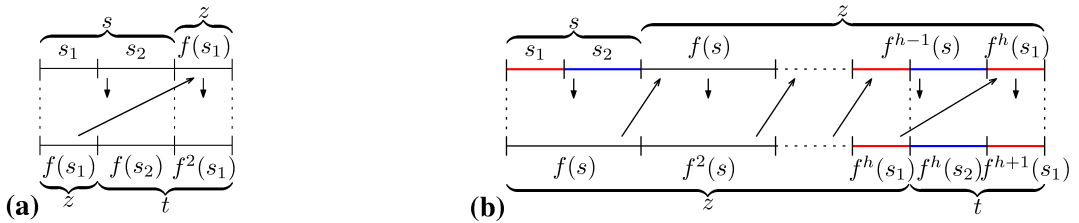

Figure 1: Representation of $sz \sim zt$ when **(a)** $0 < |z| < |s|$, or **(b)** $|z| \geq |s|$.

Proof:
If $z = \epsilon$, then $f(t) = s$ and the statement holds with $h = 0$ and $s = s_1$. If $0 < |z| < |s|$, then write $s = s_1 s_2$ where $|s_1| = |z|$. It follows that $f(s_1) = z$ because $f(s_1)$ is a prefix of $f(s)$, and $f(s_2)$ is a prefix of t because $f(s_2)$ is a suffix of $f(s)$ and $f(s) \cap t \neq \emptyset$. Then $sz \sim zt$ implies $f(z) = f^2(s_1)$ is a suffix of t, so $t = f(s_2) f^2(s_1)$ (Figure 1a). If $|z| \geq |s|$, let $h = \lceil |z|/|s| \rceil$ and write $s = s_1 s_2$ where $|s_1| = |z| - (h-1)|s|$. Note that $s_1 \neq \epsilon$. Then $z = f(s) \cdots f^{h-1}(s) f^h(s_1)$ and $t = f^h(s_2) f^{h+1}(s_1)$ (Figure 1b). $\square$

In Lemma 2.6, note that if $sz = zt$, then f is the identity mapping, and so we arrive at the lemma from [14]. In Section 3, we use the lemma symmetrically when $s = t$, and $sz \sim zs$. An illustration of the situation described with Lemma 2.6 can be seen within the prefixes of w_1 and w_2, as well as in the suffixes of w_3 and w_4.

Lemma 2.7. Suppose $x, y_1, y_2 \in \Sigma^+$ with $|y_1| = |y_2|$ and f is an equivalence map. Let $f(xy_1) = y_1 x$ and either $f(x^R y_2) = y_2 x^R$ or $f(y_2 x) = xy_2$. Then either $|x| \leq |y_1| = |y_2|$ or $\Sigma[x] \cap \Sigma[y_2] \neq \emptyset$.

Proof:
Let $x = x'a$ for $a \in \Sigma$ and let $|x| > |y_1| = |y_2|$. Then $f(x'ay_1) = y_1 x$ implies $f(a) \in \Sigma[x]$. In both cases when $f(ax'^R y_2) = y_2 x^R$ or $f(y_2 x'a) = xy_2$ we have $f(a) \in \Sigma[y_2]$. □

The last lemma from [11] details useful properties of repeat and return words.

Lemma 2.8. Let $w \in \Sigma_{DOW}$, let xx' and yy' be two repeat (return resp.) words in w, and let $u \in x \cap y$. Then both $(x - u)(x' - u)$ $((x - u)(x' - u^R)$ resp.) and $(y - u)(y' - u)$ $((y - u)(y' - u^R)$ resp.) are repeat (return resp.) words in w. Furthermore, if xx' and yy' are repeat and return words, respectively, in w, then $x \cap y = \{\epsilon\}$, $|x| = 1$, or $|y| = 1$.

3. Insertions yielding equivalent words

In this section, we fix $w \in \Sigma_{DOW}$ in ascending order of length n. Let $\mathcal{I}_i(\nu_i, k_i, \ell_i)$ for $i = 1, 2$ be distinct insertions into w such that $w_i = w \star \mathcal{I}_i(\nu_i, k_i, \ell_i)$ are equivalent with $f(w_1) = w_2$ for an equivalence map f. By Remark 2.4 and Definition 2.5, $\nu_1 = \nu_2 = \nu$ and $(k_1, \ell_1) \neq (k_2, \ell_2)$. Without loss of generality, we assume that $k_1 \leq k_2$. Because the inserted words have no symbols in common with w, we consider that uu'_1 is inserted with $\mathcal{I}_1$ and uu'_2 is inserted with $\mathcal{I}_2$ ($u'_1, u'_2 \in \{u, u^R\}$). If $\mathcal{I}_1, \mathcal{I}_2 \in \mathrm{Rep}$ ($\mathcal{I}_1, \mathcal{I}_2 \in \mathrm{Ret}$ resp.), then we consider that they both insert uu'.

Observe that $k_1 \neq k_2$, because if $k_1 = k_2$, then $f(u) = u$, implying $\ell_1 = \ell_2$, and hence the insertions are equal. Up to symmetry, the indices k_1, k_2, ℓ_1, ℓ_2 can have the following possibilities.

- *Interleaving* insertions, $k_1 < k_2 \leq \ell_1 < \ell_2$: $\mathcal{I}_2$ inserts u at a location before $\mathcal{I}_1$ inserts u'_1.
- *Nested* insertions, $k_1 < k_2 \leq \ell_2 < \ell_1$: $\mathcal{I}_2$ inserts u and u'_2 at locations before $\mathcal{I}_1$ inserts u'_1.
- *Sequential* insertions, $k_1 \leq \ell_1 < k_2 \leq \ell_2$: $\mathcal{I}_2$ inserts u and u'_2 at locations after $\mathcal{I}_1$ inserts u'_1.

Further, without loss of generality, we can assume $k_1 = 1$ and $\ell = \max\{\ell_1, \ell_2\} = n + 1$. For the remainder of the section, we set $w = z_1 z_2 z_3$ where $|z_1| = k_2 - 1$ in the case of interleaving and nested insertions and $|z_1| = \ell_1 - 1$ in the case of sequential insertions. Also $|z_1 z_2|$ equals $\ell_1 - 1$, $\ell_2 - 1$, or $k_2 - 1$ for interleaving, nested, or sequential insertions, respectively. Hence the four positions for the two insertions into w are 1, $|z_1| + 1$, $|z_1 z_2| + 1$, and $n + 1$. We consider that w is in ascending order.

3.1. Interleaving and nested insertions

Let $t \in \Sigma_{DOW}$ be in ascending order. We say that t is an *interleaving sequence of return words* if there exist integers $h, \nu \geq 1$ such that $t = x_1 x_2 \cdots x_h x_1^R x_2^R \cdots x_h^R$, where $|x_i| = \nu$ for all i; in this case we write $t = \text{Int}(h, \nu)$. Clearly, $x_i \in \Sigma_{SOW}$ for all i. Note that if $\nu = 1$, then $\text{Int}(h, 1)$ is a repeat word of size $h \geq 1$.

In Proposition 3.1, we consider two interleaving insertions into a DOW w which yield equivalent words. To describe the structure of w, we track the image of the inserted word u in the resulting words $w_1, w_2 \in \Sigma_{DOW}$ by using the equivalence map as described by Lemmas 2.6 and 2.7. We also use this "image tracking" method when considering nested and sequential insertions.

Proposition 3.1. If $\mathcal{I}_1$ and $\mathcal{I}_2$ are interleaving insertions, that is $k_1 < k_2 \leq \ell_1 < \ell_2$, then $z_1, z_3 \in \Sigma_{SOW}$ are such that

   (1) If $\mathcal{I}_1, \mathcal{I}_2 \in \text{Rep}$, then $z_1 z_3$ is a repeat word in w.

   (2) If $\mathcal{I}_1, \mathcal{I}_2 \in \text{Ret}$, then $z_1 z_3 \sim \text{Int}(h, \nu)$ where $h = (k_2 - k_1)/\nu$ is a positive integer.

Proof:
We have that $w_1 = u z_1 z_2 u_1' z_3$ and $w_2 = z_1 u z_2 z_3 u_2'$. Because $u z_1 \sim z_1 u$ and u is SOW, we have that z_1 is a SOW by Lemma 2.6. Similarly, we have that z_3 is a SOW, and thus $|z_1| = |z_3|$ since z_2 is a DOW. Moreover, $z_1 = x_1 x_2 \cdots x_h$, where $f^i(u) = x_i$ for $1 \leq i < h$ (if any such exist) and $x_h = f^h(v_1)$ where v_1 is a prefix of u.

(1) Suppose $\mathcal{I}_1, \mathcal{I}_2 \in \text{Rep}$. Then $u = u_1' = u_2'$ and the two equivalences $u z_1 \sim z_1 u$ and $u z_3 \sim z_3 u$ are with the same equivalence map. Since z_1 and z_3 are of the same length, by Lemma 2.6 they have the same structure, i.e., $z_1 = x_1 x_2 \cdots x_h = z_3$.

(2) Suppose $\mathcal{I}_1, \mathcal{I}_2 \in \text{Ret}$. Then $u_1' = u_2' = u^R$. By Lemma 2.7, $|z_1| = |z_3| \geq |u|$ because $f(u z_1) = z_1 u$, $f(u^R z_3) = z_3 u^R$, and $\Sigma[w] \cap \Sigma[u] = \emptyset$. Finally, observe that $|x_h| = |u| = \nu$.

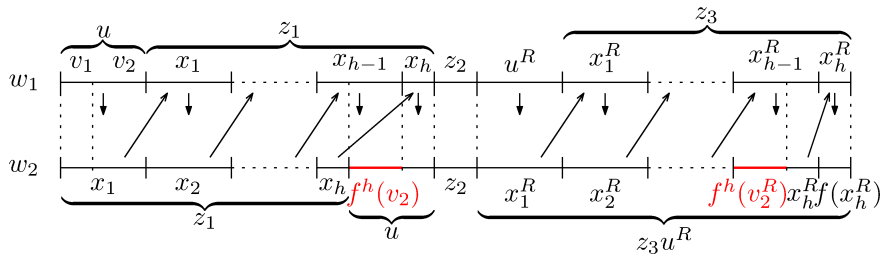
Figure 2: Representation of w_1 and w_2 when $\mathcal{I}_1, \mathcal{I}_2 \in \text{Ret}$ are such that $k_1 < k_2 \leq \ell_1 < \ell_2$.

Since $u^R z_3 \sim z_3 u^R$, by Lemma 2.6, $z_3 = x_1^R x_2^R \cdots x_{h-1}^R x_h^R$ with $f^i(u) = x_i$ $(1 \leq i < h)$ and $u = f^h(v_2) f^{h+1}(v_1) = v_1 v_2$. It follows that $f(x_{h-1}^R x_h^R) = f^h(v_2^R) f^h(v_1^R) f^{h+1}(v_1^R) = f^h(v_2^R) x_h^R f^{h+1}(v_1^R)$ is a suffix of $z_3 u^R$ (Figure 2). If $|x_h| = |v_1| < |u|$, $f^{h+1}(v_1^R)$ is a proper

suffix of u^R, and so $\emptyset \neq \Sigma[x_h^R] \cap \Sigma[u] \subset \Sigma[z_3] \cap \Sigma[u]$, contradictory to Definition 2.2. The above implies that $h = |z_1|/|u| = (k_2 - k_1)/\nu$ is a positive integer. $\qquad\square$

Example 3.2. Let $w = 12345677612345$ and consider the insertions $\rho(2, 1, 10)$ and $\rho(2, 6, 15)$ into w; also let $w' = 123456652143$ with insertions $\tau(2, 1, 9)$ and $\tau(2, 5, 13)$ into w'.

$$w_1 = \underline{89}1234567768\underline{89}12345 \qquad\qquad w'_1 = \underline{78}12345665\underline{87}2143$$
$$w_2 = 12345\underline{89}67761234\underline{89} \qquad\qquad w'_2 = 1234\underline{78}5665214\underline{87}$$

Note that $w_1 \sim w_2$ with $z_1 = z_3 = 12345$, and $w'_1 \sim w'_2$ with $z_1 z_3 = 12342143 \sim \text{Int}(2, 2)$.

Let $t \in \Sigma_{DOW}$ be in ascending order. We say that t is a *nested sequence of repeat words* if there exist integers $h, \nu \geq 1$ such that $t = x_1 x_2 \cdots x_h x_h x_{h-1} \cdots x_1$, where $|x_i| = \nu$ for all i; in this case we write $t = \text{Nes}(h, \nu)$. Clearly, $x_i \in \Sigma_{SOW}$ for all i and if $\nu = 1$, then $\text{Nes}(h, 1)$ is a return word of size $h \geq 1$.

Proposition 3.3. If $\mathcal{I}_1$ and $\mathcal{I}_2$ are nested insertions, that is $k_1 < k_2 \leq \ell_2 < \ell_1$, then $z_1, z_3 \in \Sigma_{SOW}$ are such that

(1) If $\mathcal{I}_1, \mathcal{I}_2 \in \text{Ret}$, then $z_1 z_3$ is a return word in w.
(2) If $\mathcal{I}_1, \mathcal{I}_2 \in \text{Rep}$, then $z_1 z_3 \sim \text{Nes}(h, \nu)$ where $h = (k_2 - k_1)/\nu$ is a positive integer.

Proof:
We have that $w_1 = u z_1 z_2 z_3 u'_1$ and $w_2 = z_1 u z_2 u'_2 z_3$. As in the proof of Proposition 3.1, it must be that z_1 and z_3 are SOW by Lemma 2.6 because u is SOW; hence $|z_1| = |z_3|$. Also, $z_1 = x_1 x_2 \cdots x_h$, where $f^i(u) = x_i$ for $1 \leq i < h$ (if any such exist) and $x_h = f^h(v_1)$ where v_1 is a prefix of u.

(1) First, assume that $\mathcal{I}_1, \mathcal{I}_2 \in \text{Ret}$ and $u'_1 = u'_2 = u^R$. Because $u^R z_3 = f(z_3 u^R)$, applying Lemma 2.6 symmetrically we have that $z_3 = x_h^R x_{h-1}^R \cdots x_1^R = z_1^R$.
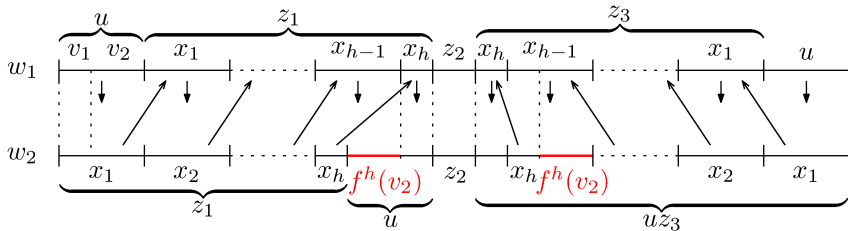

Figure 3: Representation of w_1 and w_2 when $\mathcal{I}_1, \mathcal{I}_2 \in \text{Rep}$ are such that $k_1 < k_2 \leq \ell_2 < \ell_1$.

(2) Suppose $\mathcal{I}_1, \mathcal{I}_2 \in \text{Rep}$ and $u = u'_1 = u'_2$. Because $f(u z_1) = z_1 u$ and $f(z_3 u) = u z_3$, by Lemma 2.7 we have $|u| \leq |z_1| = |z_3|$. Similarly as in Proposition 3.1, we observe that $|x_h| = |u|$. Applying Lemma 2.6 symmetrically (Figure 3) $z_3 = x_h x_{h-1} \cdots x_2 x_1$. It follows that $f(x_h x_{h-1}) = f^{h+1}(v_1) f^h(v_1) f^h(v_2) = f^{h+1}(v_1) x_h f^h(v_2)$ is a prefix of $u z_3$ and if $|x_h| = |v_1| < |u|$, then

$f^{h+1}(v_1)$ is a proper prefix of u, and so $\Sigma[x_h] \cap \Sigma[u] \neq \emptyset$. But then $\emptyset \neq \Sigma[x_h] \cap \Sigma[u] \subset \Sigma[z_3] \cap \Sigma[u]$, contradictory to Definition 2.2. This implies that $h = |z_1|/|u| = (k_2 - k_1)/\nu$ is a positive integer. $\quad\square$

Example 3.4. Let $w = 123456653412$ with the insertions $\rho(2, 1, 13)$ and $\rho(2, 5, 9)$ into w; also let $w' = 12345677654321$ with the insertions $\tau(2, 1, 15)$ and $\tau(2, 6, 10)$ into w'.

$$w_1 = \underline{78}1234566534412\underline{78} \qquad w_1' = \underline{89}1234567765432\underline{1}\underline{98}$$
$$w_2 = 1234\underline{78}5665\underline{78}3412 \qquad w_2' = 12345\underline{89}67769\underline{98}54321$$

Note that $w_1 \sim w_2$ with $z_1 z_3 = 12343412 \sim \mathrm{Nes}(2, 2)$ and $w_1' \sim w_2'$ with $z_1 = 12345$ and $z_3 = z_1^R$.

Proposition 3.5. Let $\nu \in \mathbb{N}$, and suppose one of the following holds:

$$
\begin{array}{llll}
(1) & t = 12\cdots m12\cdots m & (3) & t = \mathrm{Nes}(h, \nu) \\
(2) & t = \mathrm{Int}(h, \nu) & (4) & t = 12\cdots mm\cdots 21
\end{array}
$$

for some $m \geq 1$ or $h \geq 1$, as appropriate. Then there exist two distinct insertions, $\mathfrak{I}_1(\nu, k_1, \ell_1)$ and $\mathfrak{I}_2(\nu, k_2, \ell_2)$, such that $t \star \mathfrak{I}_1(\nu, k_1, \ell_1) \sim t \star \mathfrak{I}_2(\nu, k_2, \ell_2)$. Moreover, the following table holds for this pair of insertions (based on the corresponding case):

	Interleaving	Nested
$\mathfrak{I}_1, \mathfrak{I}_2 \in \mathrm{Rep}$	(1)	(3)
$\mathfrak{I}_1, \mathfrak{I}_2 \in \mathrm{Ret}$	(2)	(4)

Proof:
Let $t \in \Sigma_{DOW}$ correspond to one of the given cases with $|t| = n$. We define a pair of distinct insertions into t as follows so that the table above holds:

$$
\begin{array}{ll}
(1)\text{-}(2) & \mathfrak{I}_1(\nu, 1, \frac{n}{2} + 1) \text{ and } \mathfrak{I}_2(\nu, \frac{n}{2} + 1, n + 1) \\
(3)\text{-}(4) & \mathfrak{I}_1(\nu, 1, n + 1) \text{ and } \mathfrak{I}_2(\nu, \frac{n}{2} + 1, \frac{n}{2} + 1)
\end{array}
$$

In all cases, let $uu' \in \Sigma_{DOW}$ be inserted into t. Observe the following for each case:

(1) $(u12\cdots m)(u12\cdots m) \sim (12\cdots mu)(12\cdots mu)$
(2) $(ux_1 x_2 \cdots x_h)(u^R x_1^R x_2^R \cdots x_h^R) \sim (x_1 x_2 \cdots x_h u)(x_1^R x_2^R \cdots x_h^R u^R)$
(3) $(ux_1 x_2 \cdots x_h)(x_h \cdots x_2 x_1 u) \sim (x_1 x_2 \cdots x_h u)(ux_h \cdots x_2 x_1)$
(4) $(u12\cdots m)(m\cdots 21u^R) \sim (12\cdots mu)(u^R m \cdots 21)$.

$\quad\square$

3.2. Sequential insertions

In this section we consider the case when both indices of insertion $\mathfrak{I}_1$ into w precede both indices of the insertion $\mathfrak{I}_2$ into w. First we observe the following lemma.

Lemma 3.6. If $k_1 \leq \ell_1 < k_2 \leq \ell_2$, then $k_2 - \ell_1 \geq \nu$.
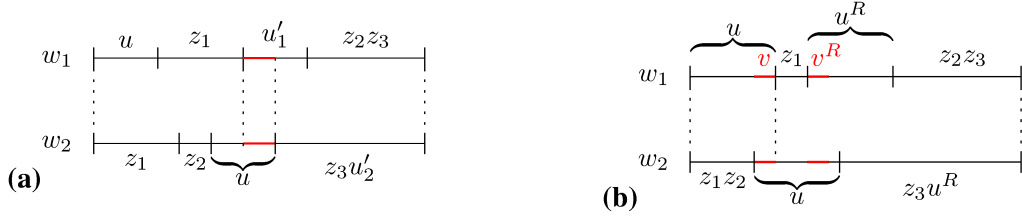
Figure 4: Representation of w_1 and w_2 when $k_1 \leq \ell_1 < k_2 \leq \ell_2$ and $k_2 - \ell_1 < |u|$.

Proof:

We have that $w_1 = u z_1 u_1' z_2 z_3$ and $w_2 = z_1 z_2 u z_3 u_2'$ with $f(w_1) = w_2$.

If $|u| > |z_2| = k_2 - \ell_1$ then a prefix of u_1' maps to suffix of u. Because a prefix of u maps also to prefix of z_1, $u_1' \neq u$, i.e., $u_1' = u^R$. Similarly, because a suffix of $z_2 z_3$ maps to a suffix of u_2', and a prefix of u^R maps to a suffix of u, it can't be $u = u_2'$ (symbols of w are disjoint with those in u). Hence, $u_2' = u^R$. We can exclude the situation in Figure 4a where $f(u)$ is a prefix of $z_1 z_2$. Thus it must be that $|z_1 z_2| < |u|$, and a suffix of u, call it v, maps to a prefix of u. Because $\Sigma[u] \cap \Sigma[z_3] = \emptyset$, it must be that the prefix v^R of u^R maps into a suffix of u. Hence the symbols in a prefix of u equal symbols in a suffix of u, which can only happen if $|z_1| = 0$ contrary to the assumption that $k_1 \neq k_2$. Therefore $|u| \leq |z_2|$. □

The structure of the words that allow distinct sequential insertions yielding equivalent words is described with the following recursive construction.

Definition 3.7. Let $m \geq 0$ and $\nu, j \geq 1$ be integers and let $s_0 = 12 \cdots m 12 \cdots m$ (if $m = 0$, $s_0 = \epsilon$). Define $s_j = s_{j-1} \star \rho(\nu, |s_{j-1}| - m + 1, |s_{j-1}| + 1)$ for $j = 1, 2, \ldots$. The word s_j is called a ρ-*tangled cord at level* j, and is denoted $s_j = T_\rho(\nu, m, j)$.

Let ν divide m and $t_0 = \text{Int}(h, \nu)$ where $h = \frac{m}{\nu}$ (if $m = 0$, $s_0 = \epsilon$). Define $t_j = t_{j-1} \star \tau(\nu, |s_{j-1}| - m + 1, |s_{j-1}| + 1)$ for $j = 1, 2, \ldots$. The word t_j is called a τ-*tangled cord at level* j, and is denoted $t_j = T_\tau(\nu, m, j)$.

In both ρ- and τ-tangled cord words the insertion in the subsequent word is performed at the end of the previous word (inserting u') and m symbols from the last symbol (inserting u).

Example 3.8. A level 3 ρ-tangled cord $w = T_\rho(2, 1, 3)$ is obtained in the following way

$$s_0 = 11 \qquad s_1 = 1\underline{23}1\underline{23} \qquad s_2 = 1231\underline{45}3\underline{45} \qquad s_3 = 1231245341\underline{67}5\underline{67},$$

and a level 2 τ-tangled cord $w = T_\tau(2, 4, 2)$ is obtained with

$$t_0 = 12342143 \qquad t_1 = 1234\underline{56}2143\underline{65} \qquad t_2 = 123456217\underline{78}4365\underline{87}.$$

Recall that $\text{Int}(h, 1)$ for $h \geq 1$ is a repeat word of size h. It follows that $T_\rho(1, m, j) = T_\tau(1, m, j)$ for $m \geq 0$ and $j \geq 1$. Moreover, if $w = T_\sigma(\nu, 0, j)$ for $\sigma \in \{\rho, \tau\}$ and $\nu, j \geq 1$, then for $|v_i| = \nu$,

$$w = \begin{cases} v_1 v_1 \cdots v_j v_j & \text{if } \sigma = \rho \\ v_1 v_1^R \cdots v_j v_j^R & \text{if } \sigma = \tau. \end{cases}$$

As with interleaving and nested insertions, we will use the "image tracking" method given by Lemmas 2.6 and 2.7 when considering sequential insertions. Since both indices $k_1 \leq \ell_1$ of the first insertion precede the indices $k_2 \leq \ell_2$ of the second, we must adapt this method to describe the structure of w in between ℓ_1 and k_2. In Lemma 3.9, we begin by considering the case in which $k_1 = \ell_1 = 1$.

Lemma 3.9. If $\mathcal{I}_1$ and $\mathcal{I}_2$ are sequential insertions such that $k_1 = \ell_1 < k_2 \leq \ell_2$, then $k_2 = \ell_2$ and $k_2 - \ell_1 = 2p\nu$ for some positive integer p. Moreover, if $\mathcal{I}_1, \mathcal{I}_2 \in \text{Rep}$ ($\mathcal{I}_1, \mathcal{I}_2 \in \text{Ret}$ resp.) then $z_1 z_2 z_3 \sim T_\rho(\nu, 0, p)$ ($z_1 z_2 z_3 \sim T_\tau(\nu, 0, p)$ resp.).
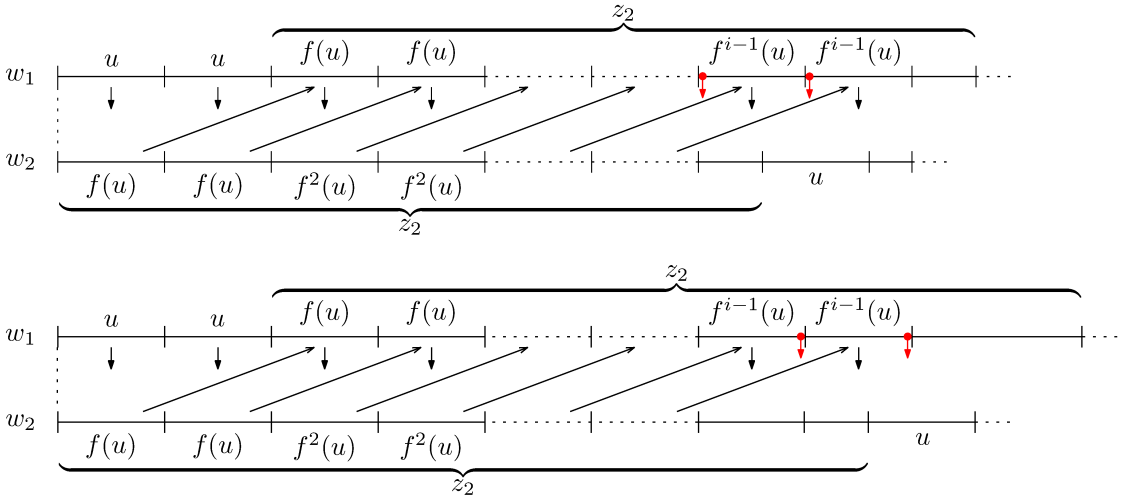

Figure 5: Representation of w_1 and w_2 when $|z_1| = 0$ and $|z_2|$ is not an even multiple of $|u|$.

Proof:
Suppose that $k_1 = \ell_1 < k_2 \leq \ell_2$. The situation is such that $w_1 = uu'z_2z_3$ and $w_2 = z_2uz_3u'$. Recall that $|z_2| \geq |u| = \nu$ by Lemma 3.6. Assume that $|z_2| = k_2 - \ell_1$ is not an even multiple of $|u|$, and let $u = b_1 \cdots b_\nu$. If $\mathcal{I}_1, \mathcal{I}_2 \in \text{Rep}$, then there exists a positive integer i such that either $f^i(b_1)$ or $f^i(b_\nu)$ belongs to $\Sigma[z_2] \cap \Sigma[u]$ (see Figure 5) contradicting the fact that $\Sigma[w] \cap \Sigma[u] = \emptyset$. Similarly, if $\mathcal{I}_1, \mathcal{I}_2 \in \text{Ret}$, then there exists a positive integer i such that $f^i(b_\nu)f^i(b_\nu) \sqsubseteq u$, or $f^i(b_1) \in \Sigma[z_2] \cap \Sigma[u]$. Then either $u \notin \Sigma_{SOW}$ or $\Sigma[w] \cap \Sigma[u] \neq \emptyset$. It follows that $k_2 - \ell_1 = |z_2| = 2p|u| = 2p\nu$ for some positive integer p such that $f^{p+1}(u) = u$.

We now apply Lemma 2.6 with $s = uu'$, $z = z_2$, and $t = u\,\text{pre}(z_3u', \nu)$. Note that $|z_2| = 2p\nu$ implies that $s = uu' = s_1$, and thus $u\,\text{pre}(z_3u', \nu) = uu'$ (i.e., $z_3 = \epsilon$) because $\Sigma[u] \cap \Sigma[w] = \emptyset$. Thus $z_2 = y_1y_1 \cdots y_py_p$ ($z_2 = y_1y_1^R \cdots y_py_p^R$ resp.) where $y_i = f^i(u)$ for $1 \leq i \leq p$ in the case of two repeat (return resp.) insertions. That is to say, z_2 is equivalent to $T_\rho(\nu, 0, p)$ ($T_\tau(\nu, 0, p)$ resp.).   □

Proposition 3.10. If $\mathcal{I}_1$ and $\mathcal{I}_2$ are sequential insertions such that $k_1 \leq \ell_1 < k_2 \leq \ell_2$, then $k_2 - \ell_1 = 2p\nu$ for some positive integer p and $|z_1| = |z_3| = q$. Moreover, if $\mathcal{I}_1, \mathcal{I}_2 \in \text{Rep}$ then $z_1z_2z_3 \sim T_\rho(\nu, q, p)$, while if $\mathcal{I}_1, \mathcal{I}_2 \in \text{Ret}$ then $z_1z_2z_3 \sim T_\tau(\nu, q, p)$ where ν divides q.

Proof:

The situation $k_1 = \ell_1$ follows from Lemma 3.9, so we assume that $k_1 < \ell_1 < k_2 \leq \ell_2$ with $w_1 = uz_1u'z_2z_3$, $w_2 = z_1z_2uz_3u'$, and recall that $|z_2| \geq |u| = \nu$ by Lemma 3.6.

(1) Suppose $\mathcal{I}_1, \mathcal{I}_2 \in \mathrm{Rep}$; we consider two cases: (1.a) $0 < |z_1| < |u|$ and (1.b) $|z_1| \geq |u|$.

(1.a) Let $0 < |z_1| < |u|$, write $x_1 = z_1$, and let y_1 be the prefix of z_2 such that $|y_1| = |u|$; note that y_1 exists by Lemma 3.6 and that $uz_1 \sim z_1y_1$. Lemma 2.6 implies $u_1 = v_1v_2$ with $v_1 \neq \epsilon$, $z_1 = f(v_1) = x_1$, and $y_1 = f(v_2)f^2(v_1) = f(v_2)f(x_1)$.

Write $v = f(v_2)$ so that $y_1 = vf(x_1)$. Then, $f(uy_1) = x_1vf(y_1) = x_1vf(v)f^2(x_1)$. Note that $x_1v \sqsubseteq z_2$, which implies that y_1x_1v is a prefix of z_2 of length $2|u|$. Moreover, $f(y_1)$ cannot have a proper factor in common with u, otherwise $f(x_1) \cap u \neq \emptyset$ contradicting the fact that $\Sigma[z_2] \cap \Sigma[u] = \emptyset$. Thus, there are two possibilities: either $f(y_1) = u$ or $f(y_1)$ is a proper factor of z_2. In the first situation, $z_2 = y_1x_1v = vf(x_1)x_1v$, and thus $|z_2| = 2|u| = 2\nu$, with $f(v)$ a prefix of u; thus, $z_3 = f(x_1)$ and $|z_3| = |z_1|$. Then $z_1z_2z_3 = x_1y_1x_1vf(x_1) = x_1y_1x_1y_1$. Observe that $z_1z_2z_3 \sim T_\rho(\nu, \ell_1 - k_1, 1)$ and in particular $s_0 \sim x_1x_1$ (where s_0 starts the induction of the tangled cord as in Definition 3.7).
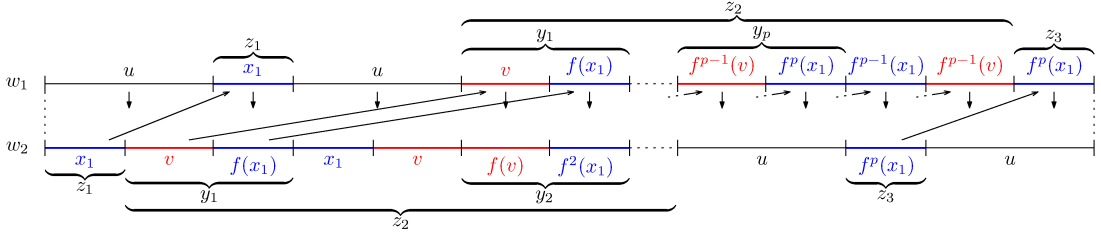

Figure 6: Representation of w_1 and w_2 when $|z_1| < |u|$ and $f(y_1)$ is a proper factor of z_2.

In the second situation, we set $y_2 = f(y_1) = f(v)f^2(x_1) \sqsubseteq z_2$. Similarly as above, we have that $y_1x_1vy_2f(x_1)f(v)$ is a prefix of z_2 of length $4|u|$, and either $f(y_2) = u$ or $f(y_2)$ is a proper factor of z_2. Inductively, there is $p \geq 1$ such that $y_p = f^{p-1}(v)f^p(x_1)$ and $f(y_p) = u$ (see Figure 6). It follows that

$$z_2 = y_1x_1vy_2f(x_1)f(v)\cdots y_pf^{p-1}(x_1)f^{p-1}(v) \text{ with } p = \frac{k_2 - \ell_1}{2\nu}.$$

Because $f^p(v)$ is a prefix of u, $z_3 = f^p(x_1)$ and $|z_3| = |z_1|$. As a consequence,

$$z_1z_2z_3 = x_1y_1x_1vy_2f(x_1)f(v)\cdots y_pf^{p-1}(x_1)f^{p-1}(v)f^p(x_1) = s_p$$

for $p \geq 1$. Observe that $z_1z_2z_3 \sim T_\rho(\nu, \ell_1 - k_1, p)$ because $y_i = f^{i-1}(v)f^i(x_1)$ for $1 \leq i \leq p$. In fact, taking $s_0 \sim x_1x_1$ we have

$$s_p = x_1y_1x_1vy_2f(x_1)f(v)\cdots y_pf^{p-1}(x_1)y_p \sim s_{p-1}\star\rho(\nu, |s_{p-1}|-|z_1|+1, |s_{p-1}|+1) \sim T_\rho(\nu, |z_1|, p).$$

(1.b) Let $|z_1| \geq |u| = \nu$, and set as above y_1 to be the prefix of z_2 such that $|y_1| = |u|$. Because $uz_1 \sim z_1y_1$, by Lemma 2.6 we have that $u = v_1v_2$ with $v_1 \neq \epsilon$, $z_1 = x_1x_2\cdots x_h$, where $f^i(u) = x_i$

for $1 \leq i < h$ (if any such exist) with $x_h = f^h(v_1)$, and $y_1 = f^h(v_2)f^{h+1}(v_1) = f^h(v_2)f(x_h)$. Write $v = f^h(v_2)$ so that $y_1 = vf(x_h)$; if $v_2 = \epsilon$, then $v = \epsilon$ and $y_1 = f(x_h)$ (see Figure 7). Similarly as in the case (1.a), using the same notation, it implies that there is a positive integer $p \geq 1$ such that $f(y_p) = u$.
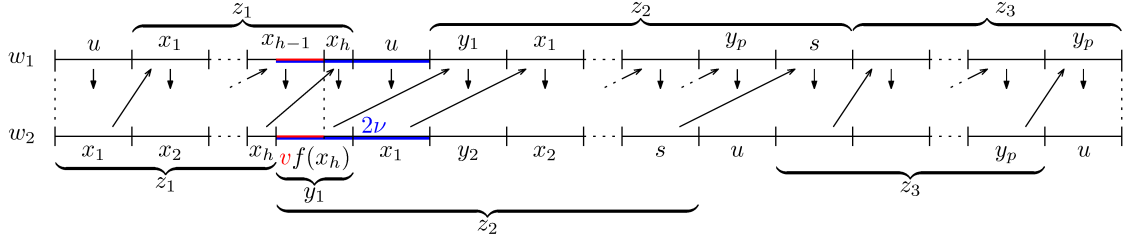

Figure 7: Representation of w_1 and w_2 when $|z_1| \geq |u|$ and $f(y_1)$ is a proper factor of z_2. If $p < h$, then $s = x_p$; otherwise $s = f^{p-h}(x_h v)$.

Suppose first $p < h$; then $z_2 = y_1 x_1 \cdots y_p x_p$, indicating $p = \frac{k_2 - \ell_1}{2\nu}$ because $|y_i| = |x_i|$ for $1 \leq i \leq p < h$ (see Figure 7 where $s = x_p$). Because $y_p x_p$ is a suffix of z_2 and $f(y_p x_p) = u x_{p+1}$, it follows that $x_{p+1} \cdots x_h$ is a prefix of z_3. Note that $f(x_{h-1} x_h) = x_h v f(x_h) = x_h y_1$; so $x_{p+1} \cdots x_h y_1$ is a prefix of z_3. Inductively, we have that $z_3 = x_{p+1} \cdots x_h y_1 \cdots y_p$ because $f^{-1}(u) = y_p$, and so $|z_3| = |z_1|$ since $|y_i| = |u| = \nu$ for all i. Hence,

$$z_1 z_2 z_3 = x_1 \cdots x_h y_1 x_1 \cdots y_p x_p x_{p+1} \cdots x_h y_1 \cdots y_p$$

for $1 \leq p < h$. Observe that $z_1 z_2 z_3 \sim T_\rho(\nu, \ell_1 - k_1, p)$ in this case (i.e., when $1 \leq p < h$). Indeed, following Definition 3.7, taking $s_0 \sim x_1 \cdots x_h x_1 \cdots x_h$ we have that

$$s_1 \sim x_1 \cdots x_h y_1 x_1 \cdots x_h y_1$$
$$s_2 \sim x_1 \cdots x_h y_1 x_1 y_2 x_2 \cdots x_h y_1 y_2$$
$$\vdots$$
$$s_p \sim x_1 \cdots x_h y_1 x_1 y_2 x_2 \cdots y_p x_p x_{p+1} \cdots x_h y_1 \cdots y_p.$$

Now suppose that $p \geq h$, then

$$z_2 = y_1 x_1 y_2 x_2 \cdots x_{h-1} y_h x_h v y_{h+1} f(x_h) f(v) y_{h+2} \cdots y_p f^{p-h}(x_h) f^{p-h}(v),$$

where $y_i = f^{i-1}(v)f^i(x_h)$ for $1 \leq i \leq p$ (see Figure 7, where $s = f^{p-h}(x_h v)$). Because $x_h v$ has length $|u| = \nu$, we again obtain that $|z_2|$ is a multiple of $2|u| = 2\nu$; in particular $|z_2| = 2p\nu$ where $p = \frac{k_2 - \ell_1}{2\nu}$. Because $f^{-1}(u) = y_p$, we have that

$$z_3 = f^{p-h+1}(x_h) \underbrace{f^{p-h+1}(v)f^{p-h+2}(x_h)}_{y_{p-h+2}} \underbrace{f^{p-h+2}(v)f^{p-h+3}(x_h)}_{y_{p-h+3}} \cdots \underbrace{f^{p-1}(v)f^p(x_h)}_{y_p}.$$

Note that $|z_3| = |z_1|$ since $|y_i| = |u| = \nu$ for all i. We observe that $z_1 z_2 z_3 \sim T_\rho(\nu, \ell_1 - k_1, p)$ in the case $p \geq h$ as well. Indeed, taking $s_0, \ldots, s_p$ as in Definition 3.7:

$$s_0 \sim x_1 \cdots x_h x_1 \cdots x_h$$
$$s_1 \sim x_1 \cdots x_h y_1 x_1 \cdots x_h y_1$$
$$\vdots$$
$$s_h \sim x_1 \cdots x_h y_1 x_1 y_2 x_2 \cdots y_h x_h y_1 y_2 \cdots y_h$$
$$s_{h+1} \sim x_1 \cdots x_h y_1 x_1 y_2 x_2 \cdots y_h x_h v y_{h+1} f(x_h) y_2 \cdots y_h y_{h+1}$$
$$\vdots$$
$$s_p \sim z_1 z_2 z_3.$$

(2) Suppose that $\mathfrak{I}_1, \mathfrak{I}_2 \in \mathrm{Ret}$. First, observe that $|z_1| \geq |u|$. On the contrary, suppose that $0 < |z_1| < |u_1|$. Similarly, and with the same notation, as in case (1.a), we have that

$$z_2 = y_1 v^R x_1^R y_2 f(v^R) f(x_1^R) \cdots y_p f^{p-1}(v^R) f^{p-1}(x_1^R) \text{ with } p = \frac{k_2 - \ell_1}{2|u|},$$

where $x_1 = z_1$, v is the nonempty prefix of z_2 such that $x_1 v = f(u)$, $y_i = f^{i-1}(v) f^i(x_1)$ for $1 \leq i \leq p$ and $f(y_p) = f(f^{p-1}(v) f^p(x_1)) = f^p(v) f^{p+1}(x_1) = u$. Note that $f^p(v)$ is a proper prefix of u, hence $f^p(v)^R$ is a proper suffix of u^R. Considering the suffix $y_p f^{p-1}(v^R) f^{p-1}(x_1^R)$ of z_2, we have that $f(y_p f^{p-1}(v^R)) = f(y_p) f^p(v^R) = u f^p(v)^R \sqsubseteq w_2$ (see Figure 6). But $u f^p(v)^R \sqsubseteq w_2$ and $u^R \sqsubseteq w_2$ contradicts the fact that w_2 is a DOW; thus $|z_1| \geq |u|$.

In the rest of the proof we use similar arguments to those used in case (1.b) for repeat insertions (with the same notation).

If $p < h$, then $z_2 = y_1 x_1^R \cdots y_p x_p^R$ with $f(y_p) = u$, which implies that $p = \frac{k_2 - \ell_1}{2\nu}$ because $|y_i| = |x_i|$ for $1 \leq i \leq p < h$. It follows that

$$x_{p+1}^R \cdots x_{h-1}^R v^R x_h^R f(v^R) f(x_h^R) \cdots f^{p-1}(v^R) f^{p-1}(x_h^R) f^p(x_h^R)$$

is a prefix of z_3 (see Figure 7). Assume that ν does not divide $|z_1| = \ell_1 - k_1$ in this situation; then $v \neq \epsilon$. Since $f(y_p) = f(f^{p-1}(v) f^p(x_h)) = f^p(v) f^{p+1}(x_h) = u$, we have that $f^p(v)^R$ is a proper suffix of u^R and $f^{p+1}(x_h)^R$ is a proper prefix of u^R. But then $f(f^{p-1}(v^R) f^{p-1}(x_h^R) f^p(x_h^R)) = f^p(v^R) f^p(x_h^R) f^{p+1}(x_h^R) \sqsubseteq w_2$ with $u^R \sqsubseteq w_2$ contradicts the fact that w_2 is a DOW. Thus,

$$p < h = \left\lceil \frac{\ell_1 - k_1}{\nu} \right\rceil = \frac{\ell_1 - k_1}{\nu}.$$

Then $|z_1| = h\nu$, $v = \epsilon$, $|x_h| = \nu$, and $y_i = f^i(x_h)$ for $1 \leq i \leq p$. Because $f^{-1}(u^R) = y_p^R$, it follows that $z_3 = x_{p+1}^R \cdots x_h^R f(x_h^R) \cdots f^p(x_h^R) = x_{p+1}^R \cdots x_h^R y_1^R \cdots y_p^R$, and so $|z_3| = |z_1| = h\nu$ since $|y_i| = |u|$ for all i. Hence,

$$z_1 z_2 z_3 = x_1 \cdots x_h y_1 x_1^R \cdots y_p x_p^R x_{p+1}^R \cdots x_h^R y_1^R \cdots y_p^R$$

for $1 \leq p < h$. Similarly as in case (1.b) we observe that $z_1 z_2 z_3 \sim T_\tau(\nu, \ell_1 - k_1, p)$ in this situation.

In case $p \geq h$ we have that

$$z_2 = y_1 x_1^R y_2 x_2^R \cdots x_{h-1}^R y_h v^R x_h^R y_{h+1} f(v^R) f(x_h^R) y_{h+2} \cdots y_p f^{p-h}(v^R) f^{p-h}(x_h^R),$$

where $y_i = f^{i-1}(v) f^i(x_h)$ for $1 \leq i \leq p$ and $f(y_p) = u$. Because $|v^R x_h^R| = |u|$, it follows that $|z_2|$ is a multiple of $2|u|$; in particular $|z_2| = 2p\nu$, where $p = \frac{k_2 - \ell_1}{2\nu}$. Moreover, we have that

$$f^{p-h+1}(v^R) f^{p-h+1}(x_h^R) \cdots f^{p-1}(v^R) f^{p-1}(x_h^R) f^p(x_h^R)$$

is a prefix of z_3 (see Figure 7). Similarly as above, $|u| = \nu$ divides $|z_1| = \ell_1 - k_1$. Thus $h = \frac{\ell_1 - k_1}{\nu}$; so $|z_1| = h\nu$, $v = \epsilon$, $|x_h| = |u|$, and $y_i = f^i(x_h)$ for $1 \leq i \leq p$. It then follows that

$$\begin{aligned} z_2 &= y_1 x_1^R y_2 x_2^R \cdots x_{h-1}^R y_h x_h^R y_{h+1} f(x_h^R) y_{h+2} \cdots y_p f^{p-h}(x_h^R) \\ &= y_1 x_1^R y_2 x_2^R \cdots x_{h-1}^R y_h x_h^R y_{h+1} y_1^R y_{h+2} \cdots y_p y_{p-h}^R \\ z_3 &= f^{p-h+1}(x_h^R) f^{p-h+2}(x_h^R) \cdots f^p(x_h^R) = y_{p-h+1}^R y_{p-h+2}^R \cdots y_p^R \end{aligned}$$

because $f^{-1}(u^R) = y_p^R$. Note that $|z_3| = |z_1| = h\nu$ since $|y_i| = |u|$ for all i. Finally we observe that $z_1 z_2 z_3 \sim T_\tau(\nu, \ell_1 - k_1, p)$. □

Example 3.11. Let $w = 12312453467567 = T_\rho(2, 1, 3)$ and consider the insertions $\rho(2, 1, 2)$ and $\rho(2, 14, 15)$ into w; also, let $w' = 123456214365 = T_\tau(2, 4, 1)$ and consider the insertions $\tau(2, 1, 5)$ and $\tau(2, 9, 13)$ into w'. Then

$$\begin{aligned} w_1 &= \underline{89}\underline{18}\underline{9}2312453467567 & w_1' &= \underline{78}1234\underline{8}756214365 \\ w_2 &= 12312453467568\underline{9}789 & w_2' &= 1234562\underline{178}4365\underline{87} \end{aligned}$$

Note that $w_1 \sim w_2$ and $w_1' \sim w_2'$.

Corollary 3.12. For every $T_\sigma(\nu, m, j)$, we have that $T_\sigma(\nu, m, j) \star \mathfrak{I}(\nu, 1, m+1) \sim T_\sigma(\nu, m, j+1)$ where $\sigma \in \{\rho, \tau\}$ and $\mathfrak{I} \in \mathrm{Rep}$ ($\mathfrak{I} \in \mathrm{Ret}$ resp.) if $\sigma = \rho$ ($\sigma = \tau$ resp.).

Proof:
The result follows by using similar arguments as in the proofs of Lemma 3.9 and Proposition 3.10. □

A DOW is a *palindrome* if it is equivalent to its reverse [3]. Let $m \geq 1$ be an integer; observe that the repeat word $12 \cdots m 12 \cdots m$ and the return word $12 \cdots m m \cdots 21$ are palindromes. Let $h, \nu \geq 1$. Observe that $\mathrm{Int}(h, \nu) = x_1 \cdots x_h x_1^R \cdots x_h^R$ is a palindrome. Indeed, consider $f : \Sigma \to \Sigma$ such that $f(x_i) = x_{h-i+1}$ for all i. Similarly, $\mathrm{Nes}(h, \nu) = x_1 \cdots x_h x_h \cdots x_1$ is a palindrome; consider $f : \Sigma \to \Sigma$ such that $f(x_i) = x_i^R$ for all i.

Proposition 3.13. Every $T_\sigma(\nu, m, j)$ is a palindrome where $\sigma \in \{\rho, \tau\}$.

Proof:

Let $\nu \geq 1$, $m \geq 0$, and $\sigma \in \{\rho, \tau\}$. We show that the result holds by inducting on j. Note that $T_\rho(\nu, m, 1)$ is a repeat word of size $m + \nu$ and $T_\tau(\nu, m, 1) = \text{Int}(\frac{m}{\nu} + 1, \nu)$. In either case, $T_\sigma(\nu, m, 1)$ is a palindrome. Next, suppose that $w = T_\sigma(\nu, m, j-1)$ is a palindrome for $j > 1$, and let $n = |w|$. The following shows that $T_\rho(\nu, m, j)$ is a palindrome:

$$
\begin{aligned}
T_\rho(\nu, m, j) \quad &\sim w \star \rho(\nu, 1, m+1) && \text{(by Corollary 3.12)} \\
&\sim w^R \star \rho(\nu, 1, m+1) && \text{(because } w \sim w^R) \\
&\sim (w \star \rho(\nu, n-m+1, n+1))^R \\
&\sim (T_\rho(\nu, m, j))^R && \text{(by the definition of } T_\rho(\nu, m, j)).
\end{aligned}
$$

The case for $T_\tau(\nu, m, j)$ follows equivalently. $\qquad\qquad\square$

Observe that there exist words that are palindromes but are neither ρ-tangled cords nor τ-tangled cords, like 12324143. However, note that 12324143 is equivalent to a cyclic permutation of the tangled cord $T_\rho(1, 1, 3)$.

Definition 3.14. Let $w \in \Sigma_{DOW}$, and let uu' be a repeat (return resp.) word in w. We say that uu' is a *maximal* repeat (return resp.) word in w if for any repeat (return resp.) word vv' in w such that $u \sqsubseteq v$, we have that $u = v$.

Example 3.15. Consider $T_\rho(2, 3, 2) = s_2$ and $T_\tau(2, 4, 1) = s_1'$.

$$
\begin{aligned}
s_0 &= 123123 & s_0' &= 12342143 \\
s_1 &= s_0 \star \rho(2, 4, 7) = 1234512345 & s_1' &= s_0' \star \tau(2, 5, 9) = 123456214365 \\
s_2 &= s_1 \star \rho(2, 8, 11) = 123451267345 67 &&
\end{aligned}
$$

Note that $T_\rho(2, 3, 2)$ contains four maximal repeat words: 33, 1212, 4545, and 6767. On the other hand, $T_\tau(2, 4, 1)$ contains three maximal return words: 1221, 3443 and 5665.

As the previous example suggests, every τ-tangled cord ends with a maximal return word of size ν. The same property holds for ρ-tangled cord whenever $j > 1$.

Proposition 3.16. The following hold.

(1) If $T_\rho(\nu_1, m_1, j_1) \sim T_\rho(\nu_2, m_2, j_2)$ and $j_1 > 1$, then $(\nu_1, m_1, j_1) = (\nu_2, m_2, j_2)$.

(2) If $T_\tau(\nu_1, m_1, j_1) \sim T_\tau(\nu_2, m_2, j_2)$, then $(\nu_1, m_1, j_1) = (\nu_2, m_2, j_2)$.

Proof:

(1) Let $T_\rho(\nu_1, m_1, j_1)$ and $T_\rho(\nu_2, m_2, j_2)$ be as given with $j_1 > 1$. Because these DOWs are equivalent and thus have equal lengths, note that $m_1 + \nu_1 j_1 = m_2 + \nu_2 j_2$. Moreover, $T_\rho(\nu_1, m_1, j_1)$ ends with a maximal repeat word of size ν_1. If $j_2 = 1$, then $T_\rho(\nu_2, m_2, j_2)$ is a repeat word of size $m_2 + \nu_2$ which cannot be equivalent to $T_\rho(\nu_1, m_1, j_1)$; thus, $j_2 > 1$. It follows that $T_\rho(\nu_2, m_2, j_2)$ ends with a maximal repeat word of size ν_2; so $\nu_1 = \nu_2$. By setting $\nu = \nu_1$, we get $|m_1 - m_2| = \nu|j_1 - j_2|$. We

now distinguish between the following two cases:

$$\text{(i) } m_1, m_2 > 0 \qquad \text{(ii) } m_1 = 0 \text{ or } m_2 = 0.$$

(1.i) Suppose $m_1, m_2 > 0$. Then $\text{pre}(T_\rho(\nu, m_1, j_1), m_1 + \nu + 1) = 12 \cdots m_1 v_1 1$ and $\text{pre}(T_\rho(\nu, m_2, j_2), m_2 + \nu + 1) = 12 \cdots m_2 \hat{v}_1 1$, with $|v_1| = |\hat{v}_1| = \nu$. It follows that $m_1 = m_2$ because $T_\rho(\nu, m_1, j_1) \sim T_\rho(\nu, m_2, j_2)$.

(1.ii) Suppose that $m_1 = 0$ but $m_2 \neq 0$. Then $\text{pre}(T_\rho(\nu, m_1, j_1), 2\nu) = v_1 v_1$ while $\text{pre}(T_\rho(\nu, m_2, j_2), m_2 + \nu) = 12 \cdots m_2 \hat{v}_1$, with $|v_1| = |\hat{v}_1| = \nu$. But this contradicts the assumption that $T_\tau(\nu, m_1, j_1) \sim T_\tau(\nu, m_2, j_2)$. So $m_1 = 0$ implies $m_2 = 0$. The same argument can be applied to show that $m_2 = 0$ implies $m_1 = 0$.

In both cases above, $m_1 = m_2$ and so $j_1 = j_2$. We can apply a similar argument to prove (2). $\qquad \square$

3.3.   Repeat vs return insertions

In this section we show that two distinct nontrivial insertions which yield equivalent words must both be of the same type, that is, both repeat insertions or both return insertions. We note that the proof for this result follows the same methodology, notation, and very similar arguments to those presented in Sections 3.1 and 3.2.

Theorem 3.17. Let $\mathfrak{I}_1(\nu, k_1, \ell_1)$ and $\mathfrak{I}_2(\nu, k_2, \ell_2)$ be two distinct nontrivial insertions into $w \in \Sigma_{DOW}$. If $w_1 = w \star \mathfrak{I}_1(\nu, k_1, \ell_1) \sim w \star \mathfrak{I}_2(\nu, k_2, \ell_2) = w_2$, then either $\mathfrak{I}_1, \mathfrak{I}_2 \in \text{Rep}$ or $\mathfrak{I}_1, \mathfrak{I}_2 \in \text{Ret}$.

Proof:
Let $w = z_1 z_2 z_3$, as in Sections 3.1 and 3.2 be of length n, and $w_1 \sim w_2$ with equivalence map f. Because $\mathfrak{I}_1$ and $\mathfrak{I}_2$ are nontrivial, $\nu \geq 2$. We prove the result by contradiction.
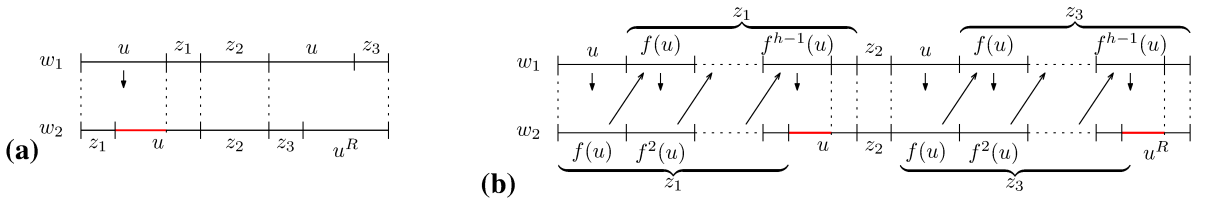

Figure 8: Representation of w_1 and w_2 when $\mathfrak{I}_1 \in \text{Rep}, \mathfrak{I}_2 \in \text{Ret}$; **(a)** $0 < |z_1| < |u|$, or **(b)** $|z_1| \geq |u|$.

Case of interleaving or nested insertions. Suppose $k_1 < k_2 \leq \ell_1 < \ell_2$ and without loss of generality we assume that $\mathfrak{I}_1 \in \text{Rep}$ and $\mathfrak{I}_2 \in \text{Ret}$. If $|z_1| < |u|$, then $f(u) \cap u \neq \emptyset$ (see Figure 8a). Because uu is a repeat word in w_1, $f(u)f(u)$ is a repeat word in w_2; so by Lemma 2.8 we have $|f(u)| = 1$ or $|u| = 1$ because uu^R is return word in w_2. Thus $|u| = 1$ implying that the insertions must be trivial. If $|z_1| \geq |u|$, then observe that $f(u)$ is a prefix of z_1. But then $f^2(u)$ is either a factor of z_1 or $f^2(u) \cap u \neq \emptyset$. Inductively, for some $h > 1$ we have that $f^i(u)$ is a factor of z_1 for $1 \leq i < h$ and $f^h(u) \cap u \neq \emptyset$ (see Figure 8b). Consider the second occurrence of u in w_1; we have that $f^i(u)$

is a factor of z_3 for $1 \leq i < h$ and $f^h(u) \cap u^R \neq \emptyset$. But $f^i(u)f^i(u)$ is a repeat word in w_2 of size $|u|$ for every $1 \leq i \leq h$; by Lemma 2.8, $f^h(u) \cap u \neq \emptyset$ implies that $|f^h(u)| = 1$ or $|u| = 1$ because uu^R is a return word in w_2. Thus, $|u| = 1$ contradicting $\nu \geq 2$. The case of nested insertions, that is, $k_1 < k_2 \leq \ell_2 < \ell_1$, follows by similar arguments.

Sequential insertions. Suppose that $k_1 \leq \ell_1 < k_2 \leq \ell_2$ and without loss of generality assume that $\mathfrak{I}_1 \in \mathrm{Rep}$ and $\mathfrak{I}_2 \in \mathrm{Ret}$. As in Lemma 3.9 and Proposition 3.10, we consider the cases: (I) $k_1 = \ell_1$ and (II) $k_1 \neq \ell_1$. In both cases recall that $|z_2| \geq |u|$ by Lemma 3.6.
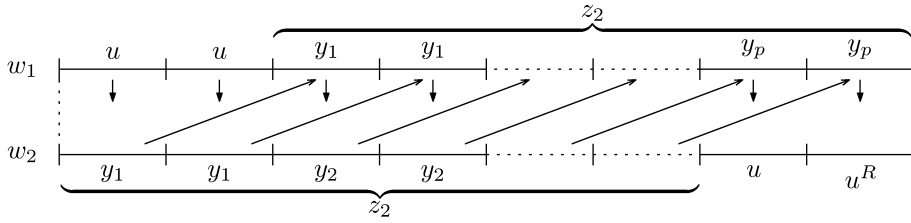

Figure 9: Representation of w_1 and w_2 when $\mathfrak{I}_1 \in \mathrm{Rep}$, $\mathfrak{I}_2 \in \mathrm{Ret}$, and $k_1 = \ell_1$.

(I) Let $k_1 = \ell_1$; i.e., $z_1 = \epsilon$. As in proof of Lemma 3.9 for repeat insertions, we have that $z_2 = y_1 y_1 \cdots y_p y_p$ where $y_i = f^i(u)$ for $1 \leq i \leq p$, $f(y_p) = u$ and $p = \frac{k_2 - \ell_1}{2\nu}$ (see Figure 9). It follows that $f(y_p)f(y_p)$ is a repeat word in w_2 of size $|u|$. Then, by Lemma 2.8, $f(y_p) = u$ implies that $|f(y_p)| = 1$ or $|u| = 1$, and the insertions must be trivial.

(II) Let $k_1 \neq \ell_1$, so that $z_1 \neq \epsilon$. If $0 < |z_1| < |u|$, by the proof of Proposition 3.10 for case (1.a), we have that
$$z_2 = y_1 x_1 v y_2 f(x_1) f(v) \cdots y_p f^{p-1}(x_1) f^{p-1}(v) \text{ with } p = \frac{k_2 - \ell_1}{2\nu},$$
where $x_1 = z_1$, v is the nonempty prefix of z_2 such that $x_1 v = f(u)$, $y_i = f^{i-1}(v)f^i(x_1)$ for $1 \leq i \leq p$ and $f(y_p) = f(f^{p-1}(v)f^p(x_1)) = f^p(v)f^{p+1}(x_1) = u$ (see Figure 6). It follows that $f(y_p f^{p-1}(x_1)f^{p-1}(v)) = uf^p(x_1)f^p(v)$, where $f^p(x_1)$ is a proper suffix of y_p and $f^p(v)$ is a proper prefix of u. Because $u^R \sqsubseteq w_2$, this contradicts the fact that w_2 is a DOW.

So assume that $|z_1| \geq |u|$, and let y_1 be the prefix of z_2 such that $|y_1| = |u|$ (exists by Lemma 3.6). By applying arguments similar to the ones used in the proof of Proposition 3.10 for case (1.b) (with the same notation), we obtain that there exists a positive integer $p \geq 1$ such that $f(y_p) = u$. If $p < h$, then $z_2 = y_1 x_1 \cdots y_p x_p$ and $z_3 = x_{p+1} \cdots x_h y_1 \cdots y_p$ (see Figure 7). On the other hand, for $p \geq h$ we have
$$z_2 = y_1 x_1 y_2 x_2 \cdots x_{h-1} y_h x_h v y_{h+1} f(x_h) f(v) y_{h+2} \cdots y_p f^{p-h}(x_h) f^{p-h}(v)$$
and $z_3 = f^{p-h+1}(x_h)y_{p-h+2} \cdots y_p$ (see Figure 7). In both cases, $p = \frac{k_2 - \ell_1}{2\nu}$ and $f(y_p)f(y_p)$ is a repeat word in w_2 of size $|u|$. Then, by Lemma 2.8, $f(y_p) = u$ implies that $|f(y_p)| = 1$ or $|u| = 1$ contradicting the fact that the insertions are nontrivial. $\qquad\square$

4. Summary

In this section, we summarize the results from Section 3. First, we collect the main results for Section 3, generalizing appropriately. Let $w \in \Sigma_{DOW}$ be in ascending order of length n and $\mathfrak{I}_i(\nu, k_i, \ell_i)$ for $i = 1, 2$ be two distinct insertions into w. Assume that $k_1 < k_2$ without loss of generality. Write $w = z_0 z_1 z_2 z_3 z_4$ and assume that the four locations of insertions (two from each $\mathfrak{I}_1$ and $\mathfrak{I}_2$) appear between words z_i and z_{i+1} for $i = 0, \ldots, 3$. In particular $|z_0| = k_1 - 1$. In the following result we summarize Propositions 3.1, 3.3, and 3.10. Observe that, by Theorem 3.17, the classification below is exhaustive.

Theorem 4.1. Let $w \in \Sigma_{DOW}$, and let $\mathfrak{I}_i(\nu, k_i, \ell_i)$ for $i = 1, 2$ be two distinct insertions into w. Let $w_i = w \star \mathfrak{I}(\nu, k_i, \ell_i)$ for $i = 1, 2$. Write $w = z_0 z_1 z_2 z_3 z_4$ using the notation above. If $w_1 \sim w_2$, then the following hold:

	$\mathfrak{I}_1, \mathfrak{I}_2 \in \mathrm{Rep}$	$\mathfrak{I}_1, \mathfrak{I}_2 \in \mathrm{Ret}$
Interleaving $k_1 < k_2 \leq \ell_1 < \ell_2$	$z_1 z_3$ is a repeat word	$z_1 z_3 \sim \mathrm{Int}\left(\frac{|z_1|}{\nu}, \nu\right)$
Nested $k_1 < k_2 \leq \ell_2 < \ell_1$	$z_1 z_3 \sim \mathrm{Nes}\left(\frac{|z_1|}{\nu}, \nu\right)$	$z_1 z_3$ is a return word
Sequential $k_1 \leq \ell_1 < k_2 \leq \ell_2$	$z_1 z_2 z_3 \sim T_\rho\left(\nu, |z_1|, \frac{|z_2|}{2\nu}\right)$	$z_1 z_2 z_3 \sim T_\tau\left(\nu, |z_1|, \frac{|z_2|}{2\nu}\right)$

Remark 4.2. Let $w, w_1, w_2 \in \Sigma_{DOW}$ be as in Theorem 4.1; then the structure of the words w_1 and w_2 obtained by the insertions satisfies the following properties (uu' is the inserted repeat or return word):

	$\mathfrak{I}_1, \mathfrak{I}_2 \in \mathrm{Rep}$	$\mathfrak{I}_1, \mathfrak{I}_2 \in \mathrm{Ret}$
Interleaving	$z_1 u z_3 u$ is a repeat word	$z_1 u z_3 u^R \sim \mathrm{Int}\left(\frac{|z_1|}{\nu} + 1, \nu\right)$
Nested	$z_1 u u z_3 \sim \mathrm{Nes}\left(\frac{|z_1|}{\nu} + 1, \nu\right)$	$z_1 u u^R z_3$ is a return word
Sequential	$z_1 z_2 u z_3 u \sim T_\rho\left(\nu, |z_1|, \frac{|z_2|}{2\nu} + 1\right)$	$z_1 z_2 u z_3 u^R \sim T_\tau\left(\nu, |z_1|, \frac{|z_2|}{2\nu} + 1\right)$

We generalize Proposition 3.5 and Corollary 3.12 with the following theorem.

Theorem 4.3. Let $w \in \Sigma_{DOW}$ and $\nu \in \mathbb{N}$. Suppose $z_1 z_2 z_3 \sqsubseteq w$ for some $z_1, z_3 \in \Sigma_{SOW}$ and $z_2 \in \Sigma^*$. If one of the following (1) - (6) holds, then there exist two distinct insertions, $\mathfrak{I}_1(\nu, k_1, \ell_1)$ and $\mathfrak{I}_2(\nu, k_2, \ell_2)$, such that $w \star \mathfrak{I}_1(\nu, k_1, \ell_1) \sim w \star \mathfrak{I}_2(\nu, k_2, \ell_2)$.

(1) $z_1 z_3$ is a repeat word in w ($\mathfrak{I}_1, \mathfrak{I}_2 \in \mathrm{Rep}$ and interleaving insertions)

(2) $z_1 z_3 \sim \mathrm{Int}(h, \nu)$ for some $h \geq 1$ ($\mathfrak{I}_1, \mathfrak{I}_2 \in \mathrm{Ret}$ and interleaving insertions)

(3) $z_1 z_3 \sim \mathrm{Nes}(h, \nu)$ for some $h \geq 1$ ($\mathfrak{I}_1, \mathfrak{I}_2 \in \mathrm{Rep}$ and nested insertions)

(4) $z_1 z_3$ is a return word in w ($\mathfrak{I}_1, \mathfrak{I}_2 \in \mathrm{Ret}$ and nested insertions)

(5) $z_1 z_2 z_3 \sim T_\rho(\nu, m, j)$ ($\mathfrak{I}_1, \mathfrak{I}_2 \in \mathrm{Rep}$ and sequential insertions)
       for some $m \geq 0$ and $j \geq 1$

(6) $z_1 z_2 z_3 \sim T_\tau(\nu, m, j)$ ($\mathfrak{I}_1, \mathfrak{I}_2 \in \mathrm{Ret}$ and sequential insertions)
       for some $m \geq 0$ and $j \geq 1$

Remark 4.4. Observe that 11 is the class representative for trivial repeat and return words in Σ^*, and

$$\mathrm{Int}(1, 1) = \mathrm{Nes}(1, 1) = T_\rho(1, 0, 1) = T_\tau(1, 0, 1) = 11.$$

By Theorem 4.3, it follows that for any $w \in \Sigma_{DOW} \setminus \{\epsilon\}$, any $a \in \Sigma[w]$ can serve as $z_1 = a$ and $z_3 = a$, and for any $\nu \in \mathbb{N}$, there exist a pair of interleaving (nested resp.) insertions into w, $\rho(\nu, k_1, \ell_1)$ and $\rho(\nu, k_2, \ell_2)$ ($\tau(\nu, k_1, \ell_1)$ and $\tau(\nu, k_2, \ell_2)$ resp.), which yield equivalent words. Moreover, if $aa \sqsubseteq w$, then there is also a pair of sequential and trivial insertions into w which yield equivalent words.

5.   Conclusion

In this paper, we characterized the structure of a DOW w which allows two distinct insertions to yield equivalent words. In particular, we proved that a nontrivial repeat insertion and a nontrivial return insertion into w cannot produce equivalent resulting words. We summarized these results in Section 4; with them, we can consider a "word graph" in which vertices are equivalence classes of DOWs (with respect to $\sim$) and directed edges exist from $[w]_\sim$ to $[w']_\sim$ if $w' \sim w \star \mathfrak{I}(\nu, k, \ell)$ for some insertion $\mathfrak{I}(\nu, k, \ell)$. Our results characterize when two insertions define the same edge in this graph. It is also natural to define the distance between $[w]_\sim$ and $[w']_\sim$ as the length of a shortest path from $[w]_\sim$ to $[w']_\sim$ in the word graph. The notion of determining the distance from the empty word $[\epsilon]_\sim$ to $[w]_\sim$ may describe the complexity of a particular DNA rearrangement processes in the ciliate *Oxytricha trifallax* [5, 11]. It may also be of interested to compare different paths between two given equivalence classes and to characterize the subgraphs which may appear in such a word graph.

References

[1] Arredondo RC. Properties of graphs used to model DNA recombination. Graduate Theses and Dissertations, University of South Florida. 2014.

[2] Braun J, Nabergall L, Neme R, Landweber LF, Saito M, Jonoska N. Russian doll genes and complex chromosome rearrangements in *Oxytricha trifallax*. *G3: Genes, Genomes, Genetics*. 2018; **8**(5):1669–1674. doi:10.1534/g3.118.200176.

[3] Burns J, Dolzhenko E, Jonoska N, Muche T, Saito M. Four-regular graphs with rigid vertices associated to DNA recombination. *Discrete Applied Mathematics*. 2013; **161**(10-11):1378–1394. doi:10.1016/j.dam.2013.01.003.

[4] Burns J, Jonoska N, Saito M. Genus ranges of chord diagrams. *Journal of Knot Theory and its Ramifications*. 2015; **24**(4):1550022, doi:10.1142/S0218216515500224.

[5] Burns J, Kukushkin D, Chen X, Landweber LF, Saito M, Jonoska N. Recurring patterns among scrambled genes in the encrypted genome of the ciliate *Oxytricha trifallax*. *Journal of Theoretical Biology*. 2016; **410**:171–180. doi:10.1016/j.jtbi.2016.08.038.

[6] Chang WJ, Bryson PD, Liang H, Shin MK, Landweber LF. The evolutionary origin of a complex scrambled gene. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; **102**(42):15149–15154. doi:10.1073/pnas.0507682102.

[7] Choffrut C, Karhumäki J. Combinatorics of words. *Handbook of Formal Languages*, Springer Berlin/Heidelberg. 1997; 329–438. doi:10.1007/978-3-642-59136-5_6.

[8] Courcelle B. Circle graphs and monadic second-order logic. *Journal of Applied Logic*. 2008; **6**(3):416–442. doi:10.1016/j.jal.2007.05.001.

[9] Ehrenfeucht A, Harju T, Petre I, Prescott DM, Rozenberg G. Computation in living cells: gene assembly in ciliates. *Natural Computing Series*, Springer Berlin/Heidelberg. 2003. doi:10.1007/978-3-662-06371-2.

[10] Gibson A. Homotopy invariants of Gauss words. *Mathematische Annalen*. 2011; **349**(4):871–887. doi:10.1007/s00208-010-0536-0.

[11] Jonoska N, Nabergall L, Saito M. Patterns and distances in words related to DNA rearrangement. *Fundamenta Informaticae*. 2017; **154**(1-4):225–238. doi:10.3233/FI-2017-1563.

[12] Levenshtein V. Binary codes capable of correcting spurious insertions and deletions of ones. *Russian Problemy Peredachi Informatsii*. 1965; **1**:12–25.

[13] Lothaire M. Algebraic combinatorics on words. Cambridge University Press. 2002. doi:10.1017/CBO9781107326019.

[14] Lyndon RC, Schützenberger MP. The equation $a^M = b^N c^P$ in a free group. *Michigan Math. J.* 1962; **9**(4):289–298. doi:10.1307/mmj/1028998766.

[15] Needleman S, Wunsch C. A general method applicable to the search for similarities in the amino acid sequences of two proteins. *Journal of Molecular Biology*. 1970; **48**(3):444–453. doi:10.1016/0022-2836(70)90057-4.

[16] Sankoff D, Kruskal J. Time Warps, String Edits, and Macromolecules. *The Theory and Practice of Sequence Comparison, Reading*. 1983.

[17] Shtylla B, Traldi L, Zulli L. On the realization of double occurrence words. *Discrete Mathematics*. 2009; **309**(6):1769–1773. doi:10.1016/j.disc.2008.02.035.

[18] Stein PR. On a class of linked diagrams, I. Enumeration. *Journal of Combinatorial Theory, Series A*. 1978; **24**(3):357–366. doi:10.1016/0097-3165(78)90065-1.

[19] Turaev V. Virtual strings, *Annales de l'Institut Fourier*. 2004; **54**(7):2455–2525. doi:10.5802/aif.2086.