SUBGROUP GENERATED BY TWO DEHN TWISTS ON A NONORIENTABLE SURFACE

MICHAL STUKOW

Abstract. Let \(a\) and \(b\) be two simple closed curves on an orientable surface \(S\) such that their geometric intersection number is greater than 1. It is known that the group generated by corresponding Dehn twists \(t_a\) and \(t_b\) is isomorphic to the free group of rank 2. In this paper we extend this result to the case of a nonorientable surface.

1. Introduction

Let \(N\) be a smooth, nonorientable, compact surface. We will be mainly interested in local properties of \(N\), hence we allow \(N\) to have some boundary components and/or punctures. Let \(\mathcal{H}(N)\) be the group of all diffeomorphisms \(h: N \to N\) such that \(h\) is the identity on each boundary component and \(h\) fix the set of punctures (setwise). By \(\mathcal{M}(N)\) we denote the quotient group of \(\mathcal{H}(N)\) by the subgroup consisting of the maps isotopic to the identity with an isotopy which pointwise fixes the boundary. \(\mathcal{M}(N)\) is called the mapping class group of \(N\). The mapping class group \(\mathcal{M}(S)\) of an orientable surface \(S\) is defined analogously, but we consider only orientation preserving maps. Usually we will use the same letter for a map and its isotopy class.

Important elements of the mapping class group \(\mathcal{M}(S)\) are Dehn twists. Since Dehn twists generate \(\mathcal{M}(S)\), it is very important to have good understanding of possible relations between them. One of the very basic results in this direction is the following theorem.

Theorem 1.1 (Ishida [3]). If \(a\) and \(b\) are simple closed curves on an orientable surface \(S\) such that the geometric intersection number of \(a\) and \(b\) is greater than 1, then the group generated by Dehn twists \(t_a\) and \(t_b\) is free of rank 2.

2000 Mathematics Subject Classification. Primary 57N05; Secondary 20F38, 57M99.

Key words and phrases. Mapping class group, Nonorientable surface, Dehn twist, Free group.

Supported by NCS grant 2012/05/B/ST1/02171.
The main goal of this paper is to extend the above result to the case of a nonorientable surface – see Theorem 8.2.

The paper is organised as follows. In Section 2 we set up some basic notation, Section 3 contains some examples which show how the nonorientable case differs from the orientable one. In Section 4 we recall some language introduced in [4], namely the notion of adjacent and joinable segments. Section 5 is devoted to study properties of curves in the neighbourhood of $a \cup b$. The main theorem of the paper (Theorem 8.2) is proved in Section 8. This proof is based on two propositions (Propositions 6.1 and 7.3) which are proved in Sections 6 and 7.

2. Preliminaries

By a circle on $N$ we mean an oriented simple closed curve which is disjoint from the boundary of $N$. Usually we identify a circle with its image. If two circles $a$ and $b$ intersect, we always assume that they intersect transversely. According to whether a regular neighbourhood of a circle is an annulus or a Möbius strip, we call the circle two-sided or one-sided respectively.

We say that a circle is generic if it bounds neither a disk with fewer than two punctures nor a Möbius strip without punctures. It is known (Corollary 4.5 of [4]) that if $N$ is not a closed Klein bottle, then the circle $a$ is generic if and only if $t_a$ has infinite order in $\mathcal{M}(N)$.

For any two circles $a$ and $b$ we define their geometric intersection number as follows

$$I(a, b) = \inf\{|a' \cap b| : a' \text{ is isotopic to } a\}.$$

We say that circles $a$ and $b$ form a bigon if there exists a disk whose boundary is the union of an arc of $a$ and an arc of $b$. The following proposition provides a very useful tool for checking if two circles are in a minimal position (with respect to $|a \cap b|$).

Proposition 2.1 (Epstein [1]). Let $a$ and $b$ be generic circles on $N$. Then $|a \cap b| = I(a,b)$ if and only if $a$ and $b$ do not form a bigon. \hfill \Box

3. Disappointing examples

Let $a$ and $b$ be two circles in an oriented surface $S$ such that $I(a,b) \geq 2$. The key observation which lead to the conclusion that Dehn twists $t_a$ and $t_b$ generate a free group is the following lemma.
Lemma 3.1 (Lemma 2.3 of [3]). Assume that circles $a, b, c \subset S$ satisfy $I(a, b) \geq 2$. Then for any nonzero integer $k$

$$I(c, a) > I(c, b) \implies I(t^k_a(c), a) < I(t^k_a(c), b).$$

The above lemma allows to apply the so-called ‘Ping Pong Lemma’ (Lemma 8.1) and easily conclude that $\langle t_a, t_b \rangle$ is a free group.

It is known for some time that relations between Dehn twists and geometric intersection numbers became much more complicated if we allow the surface to be nonorientable. Some results in this direction were obtained in [4] but they were too weak to prove a nonorientable version of the above lemma. The main goal of this section is to show that there is a good reason for this, namely Lemma 3.1 is not true on nonorientable surfaces. Moreover, it is possible to find quite general families of counterexamples, hence it seems that there is no ‘easy fix’ for this situation (for a ‘nontrivial fix’ see Propositions 6.1 and 7.3).

Example 3.2. Let $a, b, c$ be two-sided circles indicated in Figure 1 (shaded disks are crosscaps – that is they interiors are to be removed and the boundary points are to be identified by the antipodal map). In particular $I(a, b) = 2$ and $I(c, a) = 8 > I(c, b) = 4$. However, it is straightforward to check that

$$I(t_a(c), a) = 8 > I(t_a(c), b) = 4.$$

The above example can be generalized in the obvious way (by changing $c$) to the example where $I(a, b) = 2$, $I(c, a) = 2n > I(c, b) = n$ and $I(t_a(c), a) = 2n > I(t_a(c), b) = n$, where $n \geq 1$.

Example 3.3. Let $a, b, c$ be two-sided circles indicated in Figure 2. In particular $I(a, b) = 8$ and $I(c, a) = 2 > I(c, b) = 1$. However, it is straightforward to check that

$$I(t_a(c), a) = 2 > I(t_a(c), b) = 1.$$
The above example can be generalized in the obvious way (by changing $b$) to the example where $I(a, b) = 2n$, $I(c, a) = 2 > I(c, b) = 1$ and $I(t_a(c), a) = 2 > I(t_a(c), b) = 1$, where $n \geq 1$.

The above examples are very disappointing, because they show that the geometric intersection number is too weak to notice the action of a twist. Moreover, this can happen for arbitrary large complexity (that is for arbitrary large values of $I(a, b)$ and $I(c, a)$).

**Example 3.4.** Let $a, c$ be two-sided circles indicated in Figure 3. Observe

that the action of $t_a$ on $c$ is trivial because $a$ bounds a Möbius strip. This is the case even though $I(a, c) = 4$ (or in general $I(a, c) = 2n$, $n \geq 1$). Such a situation can not happen on oriented surface $S$ — if $I(a, c) > 0$ for some curve $c$ on $S$, then $t_a$ is automatically nontrivial.

4. **Joinable segments of $a$ and $b$**

For the rest of the paper assume that $a$ and $b$ are two generic two-sided circles in a nonorientable surface $N$ such that $|a \cap b| = I(a, b) \geq 2$.

Following [4] by a *segment* of $b$ (with respect to $a$) we mean any unoriented arc $p$ of $b$ satisfying $a \cap p = \partial p$. Similarly we define *oriented segment* of $b$. We call a segment $p$ of $b$ *one-sided* [two-sided] if the union of $p$ and an arc of $a$ connecting $\partial p$ is a one-sided [two-sided] circle. An oriented segment is one-sided [two-sided] if the underlying unoriented segment is one-sided [two-sided].
Oriented segments $PP'$ and $QQ'$ of $b$ are called *adjacent* if both are one-sided and there exists an open disk $\Delta$ on $N$ with the following properties

1. $\partial \Delta$ consists of the segments $PP'$, $QQ'$ of $b$ and the arcs $PQ$, $P'Q'$ of $a$;
2. $\Delta$ is disjoint from $a \cup b$ (Figure 4).

\[\begin{array}{c}
\includegraphics[width=0.4\textwidth]{adjacent_segments}\end{array}\]

**Figure 4.** Adjacent segments of $b$.

Oriented segments $p \neq q$ are called *joinable* if there exist oriented segments $p_1, \ldots, p_k$ such that $p_1 = p$, $p_k = q$ and $p_i$ is adjacent to $p_{i+1}$ for $i = 1, \ldots, k - 1$ (Figure 5).

Unoriented segments are called adjacent [joinable] if they are adjacent [joinable] as oriented segments for some choice of orientations.

In exactly the same way we define segments of $a$ (with respect to $b$) and their properties.

The main reason for the importance of adjacent/joinable segments of $b$ is that they provide natural reductions of the intersection points of $t_a(b)$ and $b$ (Figure 5). In fact, as was observed in [4], these segments are the only source of such reductions.

Let us recall some basic properties of joinable segments.

**Proposition 4.1** (Lemmata 3.4, 3.7 and 3.8 of [4]).

1. Initial [terminal] points of oriented joinable segments of $b$ are on the same side of $a$.
(2) Let \( p \) and \( q \) be oriented segments such that \( q \) begins at the terminal point of \( p \) (note that this includes the case \( q = -p \)). Then \( p \) and \( q \) are not joinable. □

Still following [4], by a double segment of \( b \) we mean an unordered pair of two different oriented segments of \( b \) which have the same initial point. There are exactly \( I(a, b) \) double segments – they correspond to intersection points of \( a \) and \( b \).

Two double segments are called joinable if there exists an oriented segment \( p \) in the first double segment and \( q \) in the other such that \( p \) and \( q \) are joinable.

Studying the action of a twist \( t_a \) on a circle \( b \) it is important to have some obstructions for possible reductions of intersection points between \( t_a(b) \) and \( b \). The basic result in this direction is the following proposition.

**Proposition 4.2** (Lemma 3.9 of [4]). Suppose \( I(a, b) \geq 2 \). Then for each double segment \( P \) there exists a double segment \( Q \neq P \) which is not joinable to \( P \). □

However, it turns out that for our purpose we need slightly stronger result.

**Proposition 4.3.** Suppose \( I(a, b) \geq 3 \). Then for each double segment \( P \) there exist two double segments different from \( P \) which are not joinable to \( P \).

**Proof.** Let \( \{p_1, p_2\} \) constitute a double segment of \( b \) corresponding to the intersection point \( P \) of \( a \) and \( b \), and assume that there exists exactly one double segment of \( b \) not joinable to \( P \) (by Proposition 4.2 there exists at least one such double segment).

Our first assertion is that both \( p_1 \) and \( p_2 \) have the initial and the terminal point on the same side of \( a \).

In fact, since \( I(a, b) > 2 \) there is at least one double segment \( P' = \{p'_1, p'_2\} \) adjacent to \( P \). We can arrange our notation so that \( p_1 \) is adjacent to \( p'_1 \) (Figure 6(i)). By Proposition 4.1 terminal points of \( p_1 \) and \( p'_1 \) are one the same side of \( a \), and segments starting at this terminal points can not be joinable to \( p_1 \). On the other hand, we assumed that one of these segments, say \( u \), is joinable either to \( p_1 \) or \( p_2 \), hence it must be joinable to \( p_2 \). Proposition 4.1 implies that \( u \neq -p_1, u \neq -p'_1 \) and the initial point of \( u \) is on the same side of \( a \) as the initial point of \( p_2 \). This proves that \( p_1 \) starts and terminates on the same side of \( a \) (that is the situation shown in Figure 6(i) is not possible).

If \( p_2 \) starts and terminates on different sides of \( a \) (Figure 6(ii)) then, by Proposition 4.1 each of two segments starting at the terminal point
of \( p_2 \) is joinable neither to \( p_1 \) nor \( p_2 \). Hence this double segment is not joinable to \( P \). This means that every other double segment is joinable to \( P \). In particular segments \( u_1 \neq -p_1 \) and \( u'_1 \neq -p'_1 \) starting at the terminal points of \( p_1 \) and \( p'_1 \) respectively must be joinable to \( p_2 \). But then segments starting at terminal points of \( u_1 \) and \( u'_1 \) are joinable neither to \( p_1 \) nor \( p_2 \) which is a contradiction. Hence we proved our assertion.

If we denote by \( Q \) and \( R \) the terminal points of \( p_1 \) and \( p_2 \) respectively, then we can assume that the configuration of segments \( p_1 \) and \( p_2 \) is as in Figure 7 (that is \( Q \neq R \) and \( Q \) is between \( P \) and \( R \)). Denote also by \( X \) the intersection point of \( a \) and \( b \) which corresponds to the double segment of \( b \) not adjacent to \( P \).

If we follow the circle \( b \) past the point \( Q \) (Figure 7(i)), by Proposition 4.1, we have in turn segments joinable to \( p_2 \) and \( p_1 \), and this continues until we reach \( X \) (note that we allow here \( X = Q \)). We have exactly the same situation (Figure 7(ii)) if we follow \( b \) past the point \( R \) (we allow here \( X = R \)). But this implies that \( b \) approaches \( X \) from below in an arc \( s_1 \) of \( a \) bounded by \( P \) and \( R \), which does not contain \( Q \) (we allow here \( X = R \)). At the same time \( b \) approaches \( X \) from above in
an arc $s_2$ of $a$ bounded by $P$ and $Q$, which does not contain $R$ (we allow here $X = Q$). Since $s_1 \cap s_2 = \{P\}$, this is impossible. 

5. CURVES IN THE NEIGHBOURHOOD OF $a \cup b$

Fix a regular neighbourhood $N_{a \cup b}$ of $a \cup b$. Topologically $N_{a \cup b}$ is the union of regular neighbourhoods $N_a$ and $N_b$ of $a$ and $b$ respectively. If we define

$$N_{a \setminus b} = N_a \setminus N_b, \quad N_{b \setminus a} = N_b \setminus N_a, \quad N_{a \cap b} = N_a \cap N_b,$$

then

$$N_{a \cup b} = N_a \cup N_b = N_{a \setminus b} \cup N_{b \setminus a} \cup N_{a \cap b},$$

where each of the three sets on the right hand side consists of $I(a, b)$ disjoint disks corresponding to the intersection points of $a$ and $b$ (Figure 8). We will think of these disks as of rectangles with two opposite sides parallel to $a$, and two other parallel to $b$. Observe that rectangles in $N_{a \setminus b}$ and $N_{b \setminus a}$ are in one-to-one correspondence with segments of $a$ and $b$ respectively.

Let $C$ be the family of generic circles on $N$ satisfying the following properties.

1. Each circle in $C$ is contained in $N_{a \cup b}$ and intersects $a \cup b$ transversally.
2. Each intersection point of $c$ and $a \cup b$ is contained in $N_{a \cap b}$.
3. If $c \in C$ and $r$ is one of the rectangles in $N_{a \cup b}$, then each arc of $c \cap r$ has endpoints on two different sides of $r$.

The third condition simply means that $c$ does not turn back when crossing a rectangle. It is obvious that each generic circle contained in $N_{a \cup b}$ is isotopic to a circle in $C$.

Let $c \in C$ and let $r$ be one of the rectangles in $N_{a \cup b}$. If there is an arc of $c$ contained in $r$ which crosses both sides of $r$ parallel to $a$, we say that $r$ contains an arc of $c$ parallel to $b$.
If every rectangle in $N_b$ contains an arc of $c$ parallel to $b$, we say that $c$ \textit{winds around} $b$. Clearly, the sufficient condition for a circle $c \in \mathcal{C}$ to wind around $b$ is that each rectangle in $N_{a \cap b}$ contains an arc of $c$ parallel to $b$.

If $r$ is a rectangle in $N_{b \setminus a}$ and $r$ contains an arc $p$ of $c$ parallel to an arc $q$ of $b$, then we say that $p$ is a \textit{segment} of $c$ (parallel to $b$). Similarly we define segments of $c$ parallel to $a$.

\textbf{Lemma 5.1.} Let $c \in \mathcal{C}$ such that $c$ winds around $b$, and let $a'$ be one of the components of $\partial N_a$. If $|c \cap a| = I(c,a)$ and $\Delta$ is a bigon formed by $c$ and $a'$, then $\Delta \subset N_a$.

\textbf{Proof.} Suppose that there is a bigon $\Delta$ with sides $p \subset a'$ and $q \subset c$ which is not contained in $N_a$. If $\Delta$ and $N_a$ are on the same side of $p$ (Figure 9), we can find a smaller bigon $\Delta' \subset \Delta$ with sides $p' \subset \partial N_a$ and $q' \subset c$ such that $\Delta'$ and $N_a$ are on different sides of $p'$ (this is because $\Delta \setminus N_a \neq \emptyset$). Hence we can assume that $\Delta$ and $N_a$ are on different sides of $p$.

If $c$ intersects the interior of $p$ we can pass to a smaller bigon which is still not contained in $N_a$, hence we can assume that $p \cap c$ consists of two points $A$ and $B$ (note that still $q$ may intersect $a'$).

Let $r_A$ and $r_B$ be rectangles of $N_{a \cap b}$ corresponding to $A$ and $B$ respectively. If $r_A = r_B$, then the segments of $q$ starting at $A$ and $B$ terminate at the same rectangle of $N_{a \cap b}$ (Figure 10). Hence we can pass to a smaller bigon $\Delta' \subset \Delta$ by removing these segments of $q$. Obtained bigon $\Delta'$ is still not contained in $N_a$ since this would imply that $c$ is not in $\mathcal{C}$ ($c$ would need to turn back in one of the rectangles of $N_a$). Hence we can assume that $r_A \neq r_B$.

Let $c_A, c_B \subset N_a$ be arcs of $c$ which start at $A$ and $B$ respectively.

Recall that we assumed that $c$ winds around $b$, hence $r_A$ and $r_B$ contain arcs of $c$ which are parallel to $b$. This means that $c_A$ either crosses $a$ in $r_A$, or $c_A$ turns in $r_A$ in the direction of $p$ and after running parallel to $p$, $c$ must turn and cross $a$ in $r_B$ – in fact it can not turn...
towards \( p \) and can not cross \( r_B \) since \( r_B \) contains an arc of \( c \) parallel to \( b \) (Figure 11). Similar analysis applied to \( c_B \) shows that the arc of \( c \cap N_a \) containing \( c_B \) must intersect \( a \) – it can do it in \( r_B \), or it can do it in \( r_A \) after running parallel to \( p \). But this implies that \( c \) and \( a \) form a bigon, which is a contradiction. 

\[ \square \]

Let \( c \in C \) and \( p \) be one of the arcs of \( c \cap N_a \). There are 4 different possible configurations of \( p \) (see Figure 12), we refer to them as types A–D. Observe also that if \( c \) winds around \( b \), then arcs of \( c \cap N_a \) of types B–D can only pass through one rectangle in \( N_a \setminus \Delta \) – otherwise \( c \) would intersect itself.

Let \( X_a \) be the set of isotopy classes of circles \( c \) in \( N \), which satisfy the following conditions

1. \( c \in C \),

\[ \]
(2) $I(c, a) = |c \cap a|$, $I(c, b) = |c \cap b|$, 
(3) $I(c, a) < I(c, b)$, 
(4) $c$ winds around $a$.

Similarly we define $X_b$ by requiring (1)–(2) above and additionally

(3') $I(c, b) < I(c, a)$, 
(4') $c$ winds around $b$.

For each double segment $P$, Proposition 4.3 gives two double segments $Q, R$ which are not joinable to $P$. However, we have no guarantee that $Q$ is not joinable to $R$. In fact it may happen that there are only two joinability classes of double segments of $b$. As we will see later, this is an essential difficulty, and in order to deal with such a case we need the following, very technical, lemma.

**Lemma 5.2.** Let $a$ and $b$ be two generic two-sided circles in $N$ such that $I(a, b) \geq 3$, and assume that there are oriented segments $p$ and $q$ of $b$ which start on different sides of $a$, such that each double segment of $b$ contains an oriented segment joinable to $p$ or $q$. Let $c \in X_b$ and $s$ is an arc of $c \cap N_a$ of type $C$ which connects the initial points of oriented segments $p'$ and $q'$ of $c$ joinable to $p$ and $q$ respectively. Then one of the arcs, say $r$, of $c \cap N_a$ following $p'$ or $q'$ (as arcs on $c$) is of type $A$ or $B$. Moreover, the arc of $c \cap N_a$ following $r$ on $c$ is not parallel to $s$ (that is it is not an arc of type $C$ connecting rectangles containing initial points of $p'$ and $q'$).

**Proof.** Suppose first that $p$ starts and terminates on the same side of $a$. In such a case, by Proposition 4.1, segments starting at the terminal points of segments joinable to $p$ must be joinable to $q$. And vice versa, segments starting at terminal points of segments joinable to $q$ must be joinable to $p$. But this implies that $b$ is a circle on an annulus (that is the union of twisted rectangles given by adjacency – see Figure 13(i)). This contradicts the assumption that $I(a, b) \geq 3$.

Hence $p$ starts and ends on two different sides of $a$. But then the segment starting at the terminal point of $p$ can not be joinable to $q$ (because $p$ and $q$ begin on two different sides of $a$), and it can not be joinable to $p$ (by Proposition 4.1). Hence $-p$ must be joinable to $q$ and without loss of generality we can assume that $q = -p$. Moreover, assume that $p$ goes up form $a$ and $q = -p$ goes down from $a$.

Our next claim is that the joinability disk $\Delta$ between the segment $p''$ of $b$ corresponding to $p'$ and any other segment $p$ of $b$ is on the left of $p''$. Suppose to the contrary, that $\Delta$ is on the right of $p''$ and let $\{r, q''\}$ constitute the double segment of $b$ corresponding to $q'$, where $q''$ is a segment joinable to $q$ (Figure 13(ii)). Since $r = p$ or $r$ leaves $N_a$
between $p''$ and $p$, $r$ is joinable to $p$. But this implies that $r$ is joinable to $-q''$ which is a contradiction with Proposition 4.1. Similarly we argue that the joinability disk between the segment $q''$ of $b$ corresponding to $q'$ and any other segment of $b$ is on the right of $q''$. Hence we have the configuration shown in Figure 14(i).

Consider the arc $r$ of $c \cap N_a$ which starts at the terminal point of $q'$. If this arc is of type D, then $r$ must turn to the right in $N_a$ and then leave $N_a$ with a segment of $b$ joinable to $p$. In fact, since $I(a, b) \geq 3$, there is at least one such segment on the left of $r$, and $r$ can not cross it, since $c$ winds around $b$. This yields a bigon $\Delta$ which contradicts Lemma 5.1 (Figure 14(ii)).

Suppose now that $r$ is an arc of type C and let $t$ be an arc of $c \cap N_a$ following $p'$ (Figure 14(i)). If $r$ goes in $N_a$ under $t$, then $c$ after passing through $t$ would wind infinitely many times parallel to $s$. The second
possibility is that \( r \) goes in \( N_a \) between \( s \) and \( t \) – in such a case \( c \) after passing through \( r \) would wind infinitely many times parallel to \( s \).

Hence \( r \) is an arc of type A or B.

Finally, let \( u \) be an arc of \( c \cap N_b \setminus a \) following \( r \) and let \( v \) be an arc of \( c \cap N_a \) following \( u \). If \( v \) is an arc of type C parallel to \( s \), then \(-u\) is the segment of \( c \) which runs parallel to \( p' \) or \( q' \). Since \(-u\) terminates on the same side of \( a \) as \( p' \), it must run parallel to \( p' \), hence \( u \) runs parallel to \( q' \). But this implies that \( u \) and \( q' \) start at the same rectangle of \( N_{a \cap b} \), which in turn implies that \( r \) is an arc of type B connecting the rectangles of \( N_{a \cap b} \) with initial points of \( p' \) and \( q' \). But then \( r \) would intersect the segments of \( c \cap N_{a \cap b} \) of type A running in other rectangles of \( N_{a \cap b} \) (there are other rectangles of \( N_{a \cap b} \) since \( I(a, b) \geq 3 \)). Hence \( v \) is not parallel to \( s \). □

6. The case of \( I(a, b) \geq 3 \)

The main goal of this section is to prove the following.

**Proposition 6.1.** Let \( a \) and \( b \) be two generic two-sided circles in \( N \) such that \( I(a, b) \geq 3 \). Then for any integer \( k \neq 0 \) we have

\[
\tau_a^k(X_b) \subseteq X_a \quad \text{and} \quad \tau_b^k(X_a) \subseteq X_b.
\]

**Proof.** Of course it is enough to prove that \( \tau_a^k(X_b) \subseteq X_a \). There is no canonical choice for the orientation of the neighbourhood \( N_a \), but in our figures, we will assume that \( \tau_a^k \) twists to the right.

**Construction of \( \tau_a^k(c) \).** Fix a circle \( c \in X_b \). It is enough to prove that \( \tau_a^k(c) \in X_a \).

We begin by constructing the circle \( d = \tau_a^k(c) \). Outside \( N_a \) and on each arc of \( d \cap N_a \) of type D, \( d \) is equal to \( c \). For each arc of \( c \cap N_a \) of types A–C, \( d \) circles \( |k| \) times around \( a \). In particular \( d \) winds around \( a \) and

\[
I(d, a) = |d \cap a| = |c \cap a| = I(c, a)
\]

\[
|d \cap b| = I(c, a) \cdot I(a, b) \cdot |k|.
\]

Now the problem is that \( d \) may not be an element of \( C \) and \( d \) does not need to be in a minimal position with respect to \( b \).

Before we start reducing \( d \), observe that if an arc of \( d \) enters \( N_a \) and then turn to the left, then after passing through one rectangle in \( N_{a \setminus b} \) it must turn back or leave \( N_a \) through the same side of \( N_a \) as it entered \( N_a \) (Figure [15]). In fact, arcs of \( d \) turning to the left in \( N_a \) came from arcs of \( c \cap N_a \) of types C and D, and we observed before that such arcs can pass through only one rectangle in \( N_{a \setminus b} \).

**Reduction of type I.** Suppose that one of the rectangles \( r \) in \( N_{a \setminus b} \) contains an arc \( p \) of \( d \) such that the endpoints of \( p \) are on the same
side of $r$ ($d$ turns back in $r$). Clearly this can not happen for $r$ being one of the rectangles in $N_{b\setminus a}$ (because in such a rectangle $d$ coincides with $c$) or $r$ being a rectangle in $N_{a\setminus b}$ (by construction $d$ runs parallel to $a$ in each such rectangle). Hence $r$ must be a rectangle in $N_{a\cap b}$ and $p$ must intersect the $b$-side of $r$ (otherwise $c$ would not be an element of $C$). Hence we have the situation illustrated in Figure 16 and we can replace $d$ with the circle $d'$ shown in the same figure. In such a case we

say that we reduced $d$ by a \textit{reduction of type I}.

Observe that reductions of type I correspond to arcs of $c \cap N_a$ of type C, hence on each arc of $d \cap N_a$ there is at most one reduction of type I.

Let $d_1$ be a circle obtained from $d$ by performing all possible reductions of type I. In particular $d_1 \in C$. Observe also that the only arcs of $d_1 \cap N_a$ which turn to the left after entering $N_a$ are arcs corresponding to (in fact equal to) arcs of $c \cap N_a$ of type D.

We now argue that $d_1$ winds around $a$. In fact, if we fix a rectangle $r$ in $N_{a\cap b}$ corresponding to a double segment $P$ of $b$, and $r'$ is another rectangle in $N_{a\cap b}$, then (since $c$ winds around $b$) $r'$ contains an arc $q$ of $c$ parallel to $b$ (Figure 17). Now $t^k_a(q)$ is strictly monotone in $N_a$ with respect to $a$, hence this arc does not admit a reduction of type I. In particular, $t^k_a(q)$ gives an arc of $d_1$ which is parallel to $a$ in $r$.

\textit{Reduction of type II.} Suppose that there are arcs $p$ and $q$ of $b$ and $d_1$ respectively such that

- $p$ and $q$ form a bigon with interior disjoint from $b \cup d_1$,
- $p \setminus N_{a\cap b}$ is a two-sided segment of $b$.
In such a case we can remove the bigon formed by $p$ and $q$ (Figure 18) and we say, that we reduced $d_1$ to $d_1'$ by a **reduction of type II**.

Let us describe the possible reductions of type II in more details. Let $A$ and $B$ be vertices of the bigon $\Delta$ formed by $p$ and $q$, and let $r$ be the rectangle of $N_{a\cap b}$ containing $A$. Since we assume that $p$ corresponds to a two-sided segment of $b$, at one endpoint of $q$, say $B$, $d_1$ turns to the left as it enters $N_a$. We claim that $p$ and $q$ enter $N_a$ in the same rectangle of $N_{a\cap b}$. Suppose to the contrary that $q$ after entering $N_a$ gives an arc $s$ of $d_1 \cap N_a$ which meets $p$ after passing through a rectangle of $N_{a\cap b}$ (Figure 19). Since $s$ turns to the left after entering $N_a$, this arc must be an arc of type D, hence after crossing $p$ it must turn left and follow an arc $t$ of $d_1$ which runs parallel to $p$ in a rectangle of $N_{b\cap a}$. But this means that $q \cup s \cup t$ together with an arc in $\partial N_a$ bound a bigon not
contained in \( N_a \) – this contradicts Lemma 5.1. Hence we proved that in fact \( q \) enters \( N_a \) in the same rectangles of \( N_{a \cap b} \) in which it intersects \( p \).

In particular, if \( d'_1 \) and \( d_1 \) differ by the reduction of type II corresponding to the bigon formed by arcs \( p \) and \( q \) of \( b \) and \( d_1 \) respectively, then \( p \) and \( q \) are in the same rectangle of \( N_{b \setminus a} \) and \( q \) is parallel to \( p \). This means that \( d_1 \) and \( d'_1 \) intersect rectangles in \( N_{a \cup b} \) in exactly the same way, hence \( d'_1 \in C \) and \( d'_1 \) winds around \( a \).

Let \( d_2 \) be the circle obtained from \( d_1 \) by performing all possible reductions of type II. As we observed, \( d_2 \in C \) and \( d_2 \) winds around \( a \).

Reduction of type III. Suppose that there exist arcs \( p \) and \( q \) of \( b \) and \( d_2 \) respectively such that

- \( p \) and \( q \) form a bigon with interior disjoint from \( b \cup d_2 \),
- \( p \setminus N_{a \cap b} \) is an one-sided segment of \( b \).

In such a case we can remove the bigon formed by \( p \) and \( q \) – see Figure 20. We say that we reduced \( d_2 \) to \( d'_2 \) by a reduction of type III.

![Figure 20. Reduction of type III.](image)

Let try to understand reductions of type III in more details.

Observe first that \( d'_2 \in C \). In fact, if \( d'_2 \) turns back in one of the rectangles \( r \) of \( N_{a \cup b} \), then \( r \) must be a rectangle which contains one of the vertices of the bigon which led to the reduction (in all other rectangles we either didn’t change anything, or \( d'_2 \) runs parallel to \( b \) in them). But then we have the situation shown in Figure 21 which would imply that \( c \) and an arc of \( \partial N_a \) form a bigon which is not contained
in \( N_a \), and this contradicts Lemma 5.1. Hence \( d'_2 \in \mathcal{C} \) (in particular \( d'_2 \) does not admit reductions of type I).

Moreover, reduction of type III does not create any new arcs of \( d'_2 \cap N_a \) which turn to the left after entering \( N_a \) and this reduction does not change the segments of \( d_2 \) which run parallel to two-sided segments of \( b \), hence \( d'_2 \) does not admit reductions of type II.

Let \( d_3 \) be the circle obtained from \( d_2 \) by performing all possible reductions of type III. As we already observed, \( d_3 \in \mathcal{C} \) and \( d_3 \) does not admit reductions of types I–III.

Let us argue that \( d_3 \) winds around \( a \). Fix a double segment \( P \) of \( b \) and the corresponding rectangle \( r_P \) in \( N_a \cap b \). By Proposition 4.2 there exists a double segment \( Q \) of \( b \) which is not adjacent to \( P \). Since \( c \) winds around \( b \), the rectangle \( r_Q \) of \( N_a \cap b \) corresponding to \( Q \) contains an arc of \( c \cap N_a \) which is parallel to \( b \) (hence is of type A). This arc yields an arc \( s \) of \( d_3 \cap N_a \) which is parallel to \( a \) in every rectangle of \( N_a \cap b \) except \( r_Q \) if \( k = 1 \). Observe that reductions of type I and II does not change the initial and terminal rectangles of \( s \), hence \( s \) as an arc of \( d_2 \) is still parallel to \( a \) in \( r_P \). Moreover, reduction of type III can change the initial and terminal rectangles of \( s \) only to the ones adjacent to \( Q \), hence \( s \) as an arc of \( d_3 \) is still parallel to \( a \) in \( r_P \). This proves that \( d_3 \) winds around \( a \).

**Bigons formed by \( d_3 \) and \( b \).** Let us prove that \( d_3 \) and \( b \) are in the minimal position, so they do not form any bigon. Suppose on the contrary, that \( \Delta \) is a bigon with vertices \( A \) and \( B \) bounded by arcs \( p \) and \( q \) of \( d_3 \) and \( b \) respectively. By taking the inner most bigon, we can assume that the interior of \( \Delta \) is disjoint from \( d_3 \cup b \).

Since \( d_3 \) winds around \( a \), each rectangle of \( N_a \cap b \) contains an intersection point of \( d_3 \) and \( b \), hence \( q \) is either an arc of \( b \) in a single rectangle \( r_{A,B} \) in \( N_a \cap b \), or \( q \) is a segment of \( b \) connecting two different rectangles \( r_A \) and \( r_B \) of \( N_a \cap b \). In the second case \( d_3 \) would admit a reduction of type II or III (depending on whether \( q \) is two-sided or not), hence let concentrate on the first possibility. If \( \Delta \) is contained in \( N_a \), then \( d_3 \) admits a reduction of type I, which is not possible. Hence \( \Delta \) is not contained in \( N_a \).

Let \( p' \) be a subarc of \( p \) which is obtained from \( p \) by removing the arcs contained in \( N_a \) which connect \( A \) and \( B \) with the boundary of \( N_a \) (Figure 22). We claim that the arc \( p' \cap d_3 \) is in fact an arc of \( c \). In order to prove this claim it is enough to show that each arc of \( p' \cap N_a \) is an arc of type D. In fact, if \( p' \) enters a rectangle \( r \) of \( N_a \cap b \), then it can not intersect \( a \) in this rectangle (since \( d_3 \) winds around \( a \)), it can not intersect \( b \), and it can not turn back. Hence it must turn to the
Figure 22. Bigon $\Delta$ between $b$ and $d_3$.

left/right and then leave $N_a$ in the next rectangle of $N_{a\cap b}$. Therefore $p'$ is in fact an arc of $c$ and the existence of $\Delta$ contradicts Lemma 5.1.

Counting intersection points between $d_3$ and $b$. In order to finish the proof we need to show that $I(d_3, b) > I(d_3, a)$. The idea is to show that for each intersection point of $d_3 \cap a$ there are associated intersection points of $d_3 \cap b$.

Observe that these arcs of $d \cap N_a$ which intersect $a$ are in 1-1 correspondence to arcs of $c \cap N_a$ intersecting $a$ (hence arcs of types A–C). Moreover, all reductions we performed on $d$ preserved this bijection – the reason is that during reductions we did not create any new arcs intersecting $a$, and we did not remove any of the existing ones.

If $p$ is an arc of $c \cap N_a$ of type A or B, and $p'$ is the arc of $d \cap N_a$ corresponding to $p$, then $p'$ may admit only reductions of type II and III.

Recall from the analysis made after the definition of reduction of type II that this kind of reduction does not change the rectangles of $N_{a\cap b}$ in which $p'$ leaves $N_a$. Hence $p'$ as an arc of $d_2$ still intersects at least two double segments of $b$.

The situation with reductions of type III is more complicated, since $p'$ may admit multiple such reductions (Figure 23). However, by Proposition 4.3, after possible reductions $p'$ as an arc of $d_3$ still intersects at least two double segments of $b$, hence there are at least 2 intersection points of $d_3 \cap b$ corresponding to $p$ (Figure 23).

Observe also that since $c$ winds around $b$ there are some arcs of $c \cap N_a$ of type A, hence for some intersection points of $c \cap a$ we get at least two intersection points of $d_3 \cap b$.

Finally, consider an arc of $c \cap N_a$ of type C. Such an arc admits a reduction of type I, and then it can admit further reductions of types II and III. If every such arc yields an arc of $d_3$ which intersects $b$, then we are done. Hence assume that for some arc $s$ of type C of $c \cap N_a$ the corresponding arc $s'$ of $d_3 \cap N_a$ does not intersect $b$.

Let $p$ and $q$ be segments of $b$ corresponding to endpoints of $s$. If there is a double segment $P$ of $b$ which does contain neither a segment
joinable to $p$ nor a segment joinable to $q$, then $s'$ would intersect $P$. Hence we are exactly in the situation described in Lemma 5.2. In particular, for each arc of $d_3 \cap N_a$ which does not intersect $b$ we can associate an arc of $c \cap N_a$ of type A or B, hence also an arc of $d_3 \cap N_a$ which crosses $b$ in 2 points. Moreover, by the second part of Lemma 5.2, this association is injective. This means that $I(d_3, b) \geq I(d_3, a)$. In fact the inequality is strict, because there are arcs of $c \cap N_a$ of type A, which were not used in the above association – for example arcs parallel to $b$ in rectangles of $N_{a\cap b}$ different from rectangles containing initial points of $p$ and $q$. □

7. THE CASE OF $I(a, b) = 2$ WITH $N_{a\cup b}$ NONORIENTABLE

It is not hard to check that Proposition 6.1 is not true if $I(a, b) = 2$ and $N_{a\cup b}$ is nonorientable. Hence we need slightly more sophisticated analysis in that case.

The case in question is very special because of the following proposition.

**Proposition 7.1.** If $a$ and $b$ are two generic two-sided circles in a surface $N$ such that $|a \cap b| = I(a, b) = 2$ and $N_{a\cup b}$ is nonorientable, then no component of $N \setminus (a \cup b)$ is a disk.

**Proof.** Observe that $N_{a\cup b}$ is a Klein bottle with two boundary components (Figure 24(i)). Hence if one of the components of $N \setminus (a \cup b)$ is a disk, then one of the circles $a$ or $b$ bounds a Möbius strip which is a contradiction. □

For a circle $c \in \mathcal{C}$ define

$J(c, a) =$ number of connected components of $c \setminus N_a$

$J(c, b) =$ number of connected components of $c \setminus N_b$.

**Proposition 7.2.** Let $a$ and $b$ are two generic two-sided circles in a surface $N$ such that $|a \cap b| = I(a, b) = 2$ and $N_{a\cup b}$ is nonorientable.

![Figure 23. Reduction of an arc of type A/B.](image-url)
If $c, c' \in \mathcal{C}$ such that $c$ is isotopic to $c'$, then $J(c, a) = J(c', a)$ and $J(c, b) = J(c', b)$.

**Proof.** Suppose first that $|c \cap c'| > 0$ and let $\Delta$ be a bigon formed by $c$ and $c'$. Since the boundary of $N_{a \cup b}$ is disjoint from $c \cup c'$, if a component of $N \setminus N_{a \cup b}$ intersects $\Delta$, then this component must be a disk. By Proposition 11, this is not possible, hence $\Delta$ is contained in $N_{a \cup b}$. Moreover, we can assume that the vertices of $\Delta$ are in the interior of rectangles of $N_{a \cup b}$.

Fix a rectangle $r$ in $N_{a \backslash b} \cup N_{b \backslash a}$ and let $\Delta_r$ be a connected component of $\Delta \cap r$. Since $\Delta \subset N_{a \cup b}$ and $c, c'$ do not turn back in any of the rectangles of $N_{a \cup b}$, $\Delta_r$ must be either a rectangle, a triangle or a bigon with two sides being arcs of $c$ and $c'$ (Figure 25). In any case, if we remove the bigon $\Delta$, that is if we replace $c'$ with the circle $c''$ isotopic to $c'$ which is obtained by pushing $c'$ across $\Delta$ (Figure 26(i)), then

$$J(c'', a) = J(c', a) \text{ and } J(c'', b) = J(c', b).$$

Therefore we can assume that $c$ and $c'$ are disjoint. This means that there exists an annulus $M$ in $N$ with boundary curves $c$ and $c'$.

If a component of $N \setminus N_{a \cup b}$ intersects $M$, then this component must be an annulus with boundary curves isotopic to $c$ and $c'$. In such a
case $N$ is a nonorientable surface of genus 4 (Figure 24(ii)) and

$$J(c, a) = J(c', a) = 2 \quad \text{and} \quad J(c, b) = J(c', b) = 2.$$ 

Finally, assume that $M$ is contained in $N_{a \cup b}$. As in the case of a bigon formed by $c$ and $c'$, if $r$ is a rectangle in $N_{a \cup b}$, then $M_r$ must be a rectangle with two sides being arcs of $c$ and $c'$ which connect opposite sides of $r$ (Figure 26(ii)). This implies that $M_r$ gives in $r$ exactly one arc of $c$ and one arc of $c'$. This means that $J(c, a) = J(c', a)$ and $J(c, b) = J(c', b)$. □

For two generic two-sided circles $a, b$ in $N$ such that $|a \cap b| = I(a, b) = 2$ we define $\tilde{X}_a$ to be the set of isotopy classes of circles in $N$ which satisfy the following conditions

1. $c \in \mathcal{C}$,
2. $J(c, a) < J(c, b)$,
3. $c$ winds around $a$.

Similarly, we define $\tilde{X}_b$ by requiring (1) above and additionally

1'. $J(c, b) < J(c, a)$,
2'. $c$ winds around $b$.

As an analogue of Proposition 6.1 we have

**Proposition 7.3.** Let $a$ and $b$ be two generic circles in $N$ such that $I(a, b) = 2$ and $N_{a \cup b}$ is nonorientable. Then for any integer $k \neq 0$ we have

$$t^k_a(\tilde{X}_b) \subseteq \tilde{X}_a \quad \text{and} \quad t^k_b(\tilde{X}_a) \subseteq \tilde{X}_b.$$ 

**Proof.** As in the proof of Proposition 6.1 we concentrate on the inclusion $t^k_b(\tilde{X}_a) \subseteq \tilde{X}_a$. We begin by constructing the circle $d = t^k_a(c)$ and as before we assume that $t^k_a$ twists to the right in $N_a$. We perform reductions of type I on $d$, and as the result we get a circle $d_1 \in \mathcal{C}$ which winds around $a$. Observe that by Proposition 7.2 it is enough to show that $J(d_1, b) > J(d_1, a)$ (we do not need to care about reductions of types II–III).

\[ \begin{array}{ll}
(i) & \includegraphics[width=0.2\textwidth]{figure1.png} \\
(ii) & \includegraphics[width=0.2\textwidth]{figure2.png}
\end{array} \]

**Figure 26.** Arcs of $c$ and $c'$ in a rectangle $r$. 
Let \( n_A, n_B, n_C, n_D \) be numbers of arcs of \( c \cap N_a \) of types A, B, C, D respectively (Figure 12). In particular
\[
J(d_1, a) = n_A + n_B + n_C + n_D.
\]
In order to determine the number \( J(d_1, b) \), suppose first that \( |k| = 1 \). Each arc of \( c \cap N_a \) of type A gives an arc of \( d_1 \) which goes once around \( a \), hence gives \( I(a, b) = 2 \) in \( J(d_1, b) \). An arc of \( c \cap N_a \) of type B gives \( I(a, b) + 1 = 3 \) in \( J(d_1, b) \), and an arc of type C gives \( I(a, b) - 1 = 1 \). An arc of \( c \cap N_a \) of type D does not change after the twist and gives 1 in \( J(d_1, b) \). Finally observe that if \( |k| > 1 \), then for each arc of \( c \cap N_a \) of types A–C we have additionally \((|k| - 1) \cdot I(a, b) = 2(|k| - 1)\) arcs of \( d_1 \setminus N_b \). Hence we proved the following formula
\[
J(d_1, b) = 2n_A + 3n_B + n_C + n_D + 2(|k| - 1) \cdot I(a, c).
\]
Since \( c \) winds around \( b \), we have \( n_A > 0 \), hence in fact \( J(d_1, b) > J(d_1, a) \).

Remark 7.4. The proof of Proposition 7.3 can be repeated with little changes when \( I(a, b) \geq 2 \) and no component of \( N \setminus N_{a,b} \) is a disk or an annulus (so for example if \( N = N_{a,b} \)). However, if there are disks in the complement of \( N_{a,b} \), then Proposition 7.2 is not true and the situation becomes much more difficult.

8. Twists generating a free group

Recall the so called 'Ping Pong Lemma' (see for example Lemma 3.15 of [2]).

Lemma 8.1. Suppose that a group \( G \) acts on a set \( X \), and \( X_1, X_2 \subseteq X \) are nonempty and disjoint. Let \( g_1, g_2 \in G \) such that for every nonzero integer \( k \),
\[
g_1^k(X_2) \subseteq X_1 \quad \text{and} \quad g_2^k(X_1) \subseteq X_2.
\]
Then the group generated by \( g_1 \) and \( g_2 \) is a free group of rank 2.

Theorem 8.2. Let \( a \) and \( b \) be two generic two-sided circles in a nonorientable surface \( N \). If \( I(a, b) \geq 2 \) then the group generated by \( t_a \) and \( t_b \) is isomorphic to the free group of rank 2.

Proof. If \( I(a, b) = 2 \) and \( N_{a,b} \) is orientable, then we can repeat Ishida’s proof [3] without any changes.

Observe that sets \( X_a, X_b, \bar{X}_a, \bar{X}_b \) defined in Sections 6 and 7 satisfy
\[
X_a \cap X_b = \emptyset, \quad \bar{X}_a \cap \bar{X}_b = \emptyset,
\]
\[
a \in X_a, \bar{X}_a, \quad b \in X_b, \bar{X}_b.
\]
Hence, if $I(a, b) = 2$ and $N_{a,b}$ is nonorientable or $I(a, b) \geq 3$, then the theorem follows from Lemma 8.1 and Propositions 7.3 and 6.1. □

References

[1] D. B. A. Epstein. Curves on 2–manifolds and isotopies. Acta Math., 115:83–107, 1966.
[2] B. Farb and D. Margalit. A Primer on Mapping Class Groups, volume 49 of Princeton Mathematical Series. Princeton Univ. Press, 2011.
[3] A. Ishida. The structure of subgroups of mapping class groups generated by two Dehn twists. Proc. Japan Acad. Ser. A Math. Sci., 72(10):240–241, 1996.
[4] M. Stukow. Dehn twists on nonorientable surfaces. Fund. Math., 189:117–147, 2006.

Institute of Mathematics, University of Gdańsk, Wita Stwosza 57, 80-952 Gdańsk, Poland
E-mail address: trojkat@mat.ug.edu.pl