Groups with small Dehn functions and bipartite chord diagrams

A.Yu. Ol’shanskii, M.V. Sapir

Abstract

We introduce a new invariant of bipartite chord diagrams and use it to construct the first examples of groups with Dehn function $n^2 \log n$. Some of these groups have undecidable conjugacy problem. Our groups are multiple HNN extensions of free groups. We show that $n^2 \log n$ is the smallest Dehn function of a multiple HNN extension of a free group with undecidable conjugacy problem.

Contents

1 Introduction 2
2 Why $n^2 \log n$? 3
3 $S$-machines 9
  3.1 $S$-machines as HNN extensions of free groups 9
  3.2 Bands 10
  3.3 Trapezia 12
  3.4 Admissible words and computations 12
  3.5 A slight modification of the $S$-machine from [11] 15
  3.6 The adding $S$-machine 15
  3.7 A composition of $S$ and the adding machine 22
  3.8 Properties of the $S$-machine $S \circ Z$ 23
4 Properties of the group $S \circ Z$ 26
  4.1 A modified length function on $S \circ Z$ 27
  4.2 Combs 28
5 Dispersion of bipartite chord diagrams 29
6 The upper bound of the Dehn function 33
7 The end of the proof 39

References 40

*Both authors were supported in part by the NSF grants DMS 0245600 and DMS 0455881. In addition, the research of the first author was supported in part by the Russian Fund for Basic Research 05-01-00895, the research of the second author was supported in part by the NSF grant DMS 9978802 and the US-Israeli BSF grant 1999298.
1 Introduction

Recall that the \textit{Dehn function} of a finite presentation \langle X \mid R \rangle of a group \(G\) is the smallest function \(f(n)\) such that any word of length at most \(n\) in \(X\) that represents the identity of \(G\) is freely equal to a product of at most \(f(n)\) conjugates of elements of \(R\). The Dehn functions \(f_1, f_2\) of any two finite presentations of the same group \(G\) are equivalent, that is \(f_2(n) \leq Cf_1(Cn) + Cn + C\), \(f_1(n) < Cf_2(Cn) + Cn + C\) for some constant \(C\). As usual, we do not distinguish equivalent functions.

The purpose of this paper is to prove the following statement.

\textbf{Theorem 1.1.} There exist finitely presented multiple HNN extensions of free groups with Dehn function \(n^2 \log n\). Some of these groups have undecidable conjugacy problem. Conversely, if \(d(n)\) is the Dehn function of a multiple HNN extension of a free group and \(\lim_{n \to \infty} \frac{d(n)}{n^2 \log n} = 0\) then the group has decidable conjugacy problem. Here \(\lim^c\) stands for the constructive limit (see Definition 2.1 below).

\textbf{Remark 1.2.} For those unfamiliar with the definition of constructive limit, it is the same as the ordinary \((\epsilon, N)\)-definition, only \(N\) must recursively depend on \(\epsilon\). Theorem 1.1 implies that a multiple HNN extension of a free group whose Dehn function does not exceed, say, \(n^2 \sqrt{\log n}\) or \(n^2 \log \log \log \log n\), must have decidable conjugacy problem.\(^1\)

In particular, this theorem gives the first example of a Dehn function between \(n^2\) and \(n^4\) not of the form \(n^\alpha\), \(\alpha \in \mathbb{R}\) (see [2], [3]). The set of Dehn functions \(\geq n^4\) is known to be large, and contains functions of the form \(n^\alpha\), \(n^\alpha \log n\), \(n^\alpha \log \log n\) for any rational \(\alpha \geq 4\) (and in fact any relatively fast computable \(\alpha > 4\)), and much more complicated functions. An almost complete description of all Dehn functions \(\geq n^4\) has been found in [11]. Unfortunately, the methods from [11] do not give any information about Dehn functions between \(n^2\) and \(n^4\), and methods of [2] and [3] give Dehn functions only of the form \(n^\alpha\). Thus the picture of the class of small Dehn functions is incomplete.

Recall also that it is still unknown if there exists a finitely presented group with Dehn function \(n^2\) and undecidable conjugacy problem. We believe that such groups do not exist and \(n^2 \log n\) is the lowest Dehn function of a group with undecidable conjugacy problem in a large class of groups than the HNN extensions of free groups (see Section 2). Groups with subquadratic Dehn functions are hyperbolic [4], [6], and so they have solvable conjugacy problem [4].

Our groups belong to the class introduced earlier in [11] by the second author. These are the hub-free realizations of \(S\)-machines (in [11] these groups were denoted by \(G^*_N(S)\)). It follows from [11, Lemma 8.1] that the Dehn function of any \(G^*_N(S)\) does not exceed \(n^3\). E. Rips and the second author conjectured that some of these groups have Dehn functions \(n^2 \log n\) and showed how such groups could be constructed. Unfortunately existing methods of finding upper bounds of Dehn functions did not give precise upper bounds of Dehn function of \(G^*_N(S)\). The reason is that \(G^*_N(S)\) does not have distorted cyclic subgroups to apply methods from [2], [3], and van Kampen diagrams over these groups do not have hyperbolic structure which allows one to apply surgeries from [11] or [1].

In this paper, we use a new method of finding the upper bounds of Dehn functions introduced by the first author. The method is based on a new invariant (dispersion) of bipartite chord diagrams naturally associated with van Kampen diagrams over the presentations of our groups. These invariants resemble the invariants of cord diagrams studied in [10, 9] by Polyak and Viro.

\(^1\)The authors are able to show that one cannot replace constructive limits by the ordinary limits in Theorem 1.1. The proof is technically more difficult than Theorem 1.1 and is left out of this paper.
in relation to problems of Arnold, computing Vassiliev invariants of knots, etc. As far as we know, this is the first use of such invariants in geometric group theory, but we are convinced that similar invariants will be applied to other problems including those which are far from Dehn functions. The main idea is that these invariants measure the complexity of a diagram, and allow us to perform surgeries decreasing the complexity.

The paper is constructed as follows. In Section 2, we give a “quasi-proof” of the fact that any finitely presented group with undecidable conjugacy problem has Dehn function at least \( n^2 \log n \), and a complete proof of this statement for multiple HNN extensions of free groups.

In Section 3, we start by introducing general properties of \( S \)-machines viewed as groups (multiple HNN-extensions of free groups). Thus we identify an \( S \)-machine and the corresponding group. In particular, we introduce the standard notions of bands and trapezia from [11].

Then we show how to slow down any \( S \)-machine \( S \circ Z \) so that the space function of the new \( S \)-machine \( S \circ Z \) becomes equivalent to the logarithm of the time function. Later, in Section 6, these properties will translate into the upper bound \( n^2 \log n \) of the Dehn function of \( S \circ Z \).

Several basic properties of the group \( S \circ Z \) are proved in Section 4. In van Kampen diagrams over an \( S \)-machine, there are two types of bands, \( Q \)-bands and \( \theta \)-bands, that start and end on the boundary of the diagram. These bands form a bipartite chord diagram (BCD) since bands of the same type do not intersect and a band of one type intersects a band of another type at most once.

In Section 5, we introduce a new invariant of BCDs, the dispersion, and prove that the dispersion of any BCD is bounded from above by a quadratic polynomial in the number of chords of one of the types. In Section 6, we prove that the area of a van Kampen diagram \( \Delta \) over \( S \circ Z \) with \( |\partial \Delta| \leq n \) does not exceed \( C(n^2 \log n + D) \) where \( D \) is the dispersion of the corresponding BCD.

This gives an upper bound of \( n^2 \log n \) for the Dehn function of \( S \circ Z \). In Section 7, we give the similar lower bound of the Dehn function and show that \( S \circ Z \) has undecidable conjugacy problem provided \( S \) has undecidable halting problem. This completes the proof of Theorem 1.1. Finally we show how to generalize our construction to obtain groups with other unusual Dehn functions between \( n^2 \) and \( n^3 \).

Acknowledgement. The authors are grateful to the referee for many helpful remarks.

2 Why \( n^2 \log n \)?

In this section, we shall give a “quasi-proof” of the following conjecture. Then we show that the conjecture is true for multiple HNN extensions of free groups.

We shall need the well known constructive version of a limit of a sequence of numbers. In fact we are going to use that definition only in the case when the limit is 0.

Definition 2.1. Let \( g: \mathbb{N} \to \mathbb{R} \) be a function. We say that the constructive limit of \( g(n) \) as \( n \to \infty \) is 0 if for every integer \( A > 0 \) there exists \( N = N(A) \) such that for every \( n > N \), \( |f(n)| \leq 1/A \), and the function \( N(A) \) is recursive. In that case we shall write \( \lim_{n \to \infty} g(n) = 0 \).

It is easy to see that \( \lim_{n \to \infty} g(n) = 0 \) if and only if there exists an increasing recursive function \( f(n) \) such that \( g(k) \leq 1/n \) for every \( k \geq f(n) \).

Quasi-Theorem 2.2. Let \( d(n) \) be the Dehn function of a finite group presentation \( P \). Suppose that \( \lim_{n \to \infty} \frac{d(n)}{n^2 \log n} = 0 \). Then \( P \) has decidable conjugacy problem.
Proof. We shall need the following Lemma. We call a van Kampen or annular diagram over a group presentation minimal if it has minimal area among all diagrams over that presentation with the same labels of the contour (contours).

Lemma 2.3. Let \( \Delta \) be a minimal annular diagram with contours \( p, p' \) over a finite group presentation \( P \). Let \( x \) be a shortest path connecting \( p \) and \( p' \). Then the area of \( \Delta \) is at least \( C|x| \log |x| \) for some constant \( C \) depending on \( P \).

Proof. Consider the following construction. Let \( p_0 = p \) (considered as a cyclic path) be the inner contour of the diagram \( \Delta \). Suppose that we have constructed a cyclic path \( p_i \) surrounding the hole of the diagram in \( \Delta \) such that \( p_i \) does not have common vertices with \( p' \). Let \( K_i \) be the annulus bounded by \( p_0 \) and \( p_i \). Let \( M_{i+1} \) be the set of cells of \( \Delta \) outside \( K_i \) that have common vertices with \( p_i \). Then let \( K_{i+1} \) be the minimal annular subdiagram of \( \Delta \) with simple contours that contains \( K_i \) and all cells from \( M_{i+1} \). Let \( p_{i+1} \) be the outer contour of \( K_i \) (the inner contour of \( K_i \) is \( p = p_0 \)).

It follows that every edge of the path \( p_{i+1} \) belongs to the contour of one of the cells of \( M_{i+1} \). Hence every vertex of \( p_{i+1} \) can be connected with a vertex of \( p_i \) by a path such that

1. The length of the path is bounded by a constant,
2. it can be connected with \( p_0 \) by a path of length at most \(^2 \mathcal{O}(i)\) and
3. the number of cells in \( M_{i+1} \) is at least \( \mathcal{O}(|p_{i+1}|) \).

From (1), it follows that the number of subdiagrams \( K_i \) is \( \mathcal{O}(|x|) \). Furthermore, more than a half of the paths \( p_i \) have length at most \( \log_c |x| \) where \( c \) is, say, four times the number of letters in the alphabet of the presentation \( P \). Indeed, otherwise we would have two paths \( p_i \) and \( p_j \), \( i \neq j \) with the same labels, and we could remove the annular subdiagram between \( p_i \) and \( p_j \) reducing the area of \( \Delta \) (that would contradict the minimality of \( \Delta \)).

From (2), it follows that at least half of the subsets \( M_i \) contain at least \( \mathcal{O}(\log |x|) \) cells each. Since these sets do not intersect, the number of cells in \( \Delta \) is at least \( \mathcal{O}(\log |x|) \).

Now the “quasi-proof” of the Quasi-Theorem 2.2 proceeds as follows. Suppose that \( P \) is a finite presentation with undecidable conjugacy problem. Suppose that the constructive limit of \( \frac{d(n)}{n^2 \log n} \) is 0. Then, in particular, \( d(n) \) is bounded from above by a recursive function, and \( P \) has solvable word problem.

Note that if an annular diagram \( \Delta \) with contour labels \( u \) and \( v \) has a simple path \( x \) with label \( t \) connecting the contours, then we can cut \( \Delta \) along \( x \) and obtain a disc van Kampen diagram with boundary label \( t^\pm 1 u t^\mp 1 v^{-1} \). So if \( |t| \) is recursively bounded in terms of \( |u| \) and \( |v| \) (for every \( u \) and \( v \) that are conjugate modulo \( P \)) then the conjugacy of \( u \) and \( v \) can be algorithmically verified.

Pick an increasing recursive function \( f(n) \) with \( \frac{d(3k)}{Ck^2 \log k} < \frac{1}{n} \) for every \( k > f(n) \) where \( C \) is the constant from Lemma 2.3 (as in Definition 2.1). Since the conjugacy problem for \( P \) is undecidable, there exists a minimal annular diagram \( \Delta \) with contours \( p, p' \) such that any path in \( \Delta \) connecting \( p \) and \( p' \) has length at least \( f(|p| + |p'|) \). Let \( n = |p| + |p'| \). Let \( x \) be a shortest path connecting \( p \) and \( p' \). Thus

\[
|x| \geq f(n), \tag{2.1}
\]

and so

\[
\frac{d(3|x|)}{C|x|^2 \log |x|} < \frac{1}{n}, \tag{2.2}
\]

We use the Computer Science “big-O” notation assuming that \( f(n) = \mathcal{O}(g(n)) \) if \( \frac{1}{g(n)} < f(n) < Cg(n) \) for some positive constant \( C \).
Since $x$ is a shortest path connecting $p$ and $p'$, $x$ is simple. Let us cut $\Delta$ along $x$ and obtain a disc diagram $\Gamma$ with boundary label $z u z^{-1} v^{-1}$ where $z$ is the label of $x^{\pm 1}$. By Lemma 2.3, the area of $\Gamma$ is at least $C|x| \log |x|$. 

Now we can take an integer $m$ between $\frac{|x|}{n} - 1$ and $\frac{|x|}{n}$. We attach $m$ copies of $\Gamma$ consecutively to each other along the sides labeled by $z$ to get a van Kampen diagram $\Pi$ with boundary label $z u m z^{-1} v^{-m}$. Notice that $\Pi$ is reduced because it covers $\Delta$ with multiplicity $m$ (after identification of the two $z$-sides of its boundary): for any 2-cell subdiagram $\Sigma$ of $\Pi$ where the cells share an edge, the annular diagram $\Delta$ contains a copy of $\Sigma$, so $\Sigma$ cannot be reducible since $\Delta$ is minimal. The perimeter $r$ of $\Pi$ is between $2|x|$ and $3|x|$, and the area is $m$ times the area of $\Delta$. So, by Lemma 2.3, the area of $\Pi$ is at least $C|x|^2 \log |x|$. By (2.2), we can deduce that the area of $\Pi$ is bigger than $d(r)$. This contradicts the definition of Dehn function of a group presentation. 

\textbf{Remark 2.4.} The only gap in the preceding argument is contained in the last phrase. Even though $\Pi$ is reduced, we cannot guarantee that $\Pi$ has minimal area among all diagrams with the same boundary label, and, in principal, the area of a minimal diagram with this boundary label may be even quadratic in terms of the perimeter $r$. Still we do not know any groups for which this proof does not work. Note that we do not need $\Pi$ to be minimal: only that the minimal diagram with the same boundary label does not have too few cells compared to $\Pi$. Also we have freedom of choosing $u, v, \Delta$ and $x$. We do not need $x$ to be a minimal length path connecting the boundary components of $\Delta$. We only need that the area of $\Delta$ exceeds $O(|x| \log |x|)$ divided by a recursive function in $n$ (depending only on the presentation). In addition, the number $m$ should only be $O\left(\frac{|x|}{|u|+|v|}\right)$. Thus Conjecture 2.2 seems true for a very large class of groups and possibly for all groups.

Let $P$ be the standard presentation of a multiple HNN extension of a free group $F_X$ with stable letters $t_1, \ldots, t_k$ and pairs of finitely generated associated subgroups $A_i = \langle a_{i,1}, \ldots, a_{i,j_i} \rangle, B_i = \langle b_{i,1}, \ldots, b_{i,j_i} \rangle$ given by their free generating sets. So the defining relations of the presentation $P$ are $a_{i,s}^{b_{i,s}^{-1}} = b_{i,s} a_i$, $i = 1, \ldots, k$, $s = 1, \ldots, j_i$. Here and below $a^t$ means $t a t^{-1}$.

As usual when one works with HNN extensions, $t$-bands play significant role (they are also called strips and corridors). We shall give a more general definition of bands in Section 3.2.

For every letter $a$, an $a$-edge in a van Kampen diagram is an edge labeled by $a^{\pm 1}$. A $t_i$-band in a diagram over $P$ is a sequence of cells containing $t_i$-edges, such that every two consecutive cells share a $t_i$-edge. It is well known [5] that in a reduced van Kampen diagram over $P$, there are no $t_i$-annuli, i.e. the first and the last $t_i$-edges of a $t_i$-band cannot coincide. So every maximal $t_i$-band in a diagram must connect two $t_i$-edges on the contour of the diagram. In an annular diagram over $P$, every maximal $t_i$-band either connects two edges belonging to the boundary or is an annulus surrounding the hole of the diagram. The contour of a $t_i$-band has the form $ef^{-1}q$ where $e, f$ are $t_i$-edges, and $e$ and $f$ do not have $t$-edges (these are the sides of the band).

The following theorem is a part of Theorem 1.1.

\textbf{Theorem 2.5.} Let $d(n)$ be the Dehn function of $P$ and $\lim_{n \to \infty} \frac{d(n)}{n^2 \log n} = 0$. Then $P$ has decidable conjugacy problem.

\textbf{Proof.} Suppose that the conjugacy problem is undecidable. We use the same notation as in the quasi-proof above.

With every reduced van Kampen diagram $\Psi$ over $P$, one can associate a chord diagram $C(\Psi)$ where the disc is the diagram and chords are the $t$-bands (more precisely, their medians).
Lemma 2.6. Let $Ψ$ and $Ψ'$ be two reduced diagrams with the same boundary, such that $C(Ψ) = C(Ψ')$. Then the areas of $Ψ$ and $Ψ'$ are the same.

Proof. Let $u$ be the common boundary label of $Ψ$ and $Ψ'$, and $C = C(Ψ) = C(Ψ')$. Then there exists a one-to-one correspondence $T \mapsto T'$ between the maximal $t$-bands in $Ψ$ and $Ψ'$. A side of each maximal $t$-band $T$ connects two vertices on $∂Ψ$. The label of the subpath of $∂Ψ$ connecting these vertices is a subword of the label of $∂Ψ$. Note that this subword is the same for the corresponding side of $T'$.

Recall that a pinch is a word of the form $t_iu_i^{-1}$ or $t_i^{-1}v_i$ where $u \in A_i$ and $v \in B_i$. Note that if the area of a van Kampen diagram over $P$ is greater than 0, then its boundary label has at least two pinches since it is equal to 1 modulo $P$.

Lemma 2.7. Suppose the boundary label $l$ of $Ψ$ has only two pinches as a cyclic word. Then $C(Ψ)$ is uniquely determined by $l$.

Proof. The boundary $∂Ψ$ is a product of two paths $pq^{-1}$ where the labels of $p$ and $q^{-1}$ do not have pinches. Hence every $t$-band connects a $t$-edge of $p$ with a $t$-edge of $q$. Since $t$-bands do not intersect, $C(Ψ)$ is reconstructed uniquely.

We say that a word $W$ is cyclically minimal if none of the cyclic shifts of it has pinches. Note that if $W$ is cyclically minimal then any power of $W$ is cyclically minimal as well.

Lemma 2.8. Suppose that a word $W$ is cyclically minimal. Suppose also that a word $U$ has no pinches. Then the word $UWU^{-1}$ has at most two pinches (as a cyclic word).

Proof. Indeed, if this word contains two pinches then $W = W_1W_2W_3$, $U = U_1U_2$ where $U_2W_1$ is a pinch and $W_3U_2^{-1}$ is a pinch. But then $W_3W_1$ is a pinch, and $W$ is not cyclically minimal.

Lemma 2.8 immediately implies

Lemma 2.9. Suppose that the words $W_1$ and $W_2$ are cyclically minimal, and $U$ does not have pinches. Then the word $UW_1U^{-1}W_2^{-1}$ has at most two pinches (as a cyclic word).

Now let us return to the proof of Theorem 2.5. Note that by Remark 2.4, we can do the following operations with $Δ$ and $x$:

- replace $Δ$ by a minimal diagram $Δ_1$ whose boundary labels are equal to the boundary labels of $Δ$ modulo $P$ and have lengths that are recursively bounded in terms of $n = |u| + |v|$, and
- replace $x$ by a path $x'$ connecting the boundary components such that $|x'|/|x|$ is recursively bounded in terms of $|u| + |v|$.
In order to be able to replace \((\Delta, x)\) by \((\Delta_1, x_1)\), one needs to replace the condition (2.1) by the condition \(|x| > f(f_1(n)) f_2(n)\) where \(f_1\) and \(f_2\) are some fixed increasing recursive functions.

Our goal is to choose \(\Delta\) and the path \(x\) so that the diagram \(\Pi\) is minimal.

We may assume that the words \(u\) and \(v\) are cyclically minimal because we can replace all pinches in these (cyclic) words by words without \(t\)'s (this can be done effectively since the word problem is solvable, and the lengths of \(u\) and \(v\) would increase only recursively). Note that a \(t\)-band cannot connect two edges on the same boundary component of \(\Delta\) because otherwise \(u\) or \(v\) would contain pinches (as cyclic words). Thus there are two cases: (1) \(u\) and \(v\) do not contain \(t\)-letters, and (2) \(u\) contains \(t\)-letters (then \(v\) also contains \(t\)-letters).

In the first case the maximal \(t\)-bands in \(\Delta\) form annuli surrounding the hole, the outer side of one annulus is the inner side of the next one. In the second case, the maximal \(t\)-bands are radial, connecting the inner contour with the outer boundary component of \(\Delta\).

Since \(\Pi\) is reduced, Lemmas 2.9, 2.6 and 2.7 tell us that we can claim that \(\Pi\) is minimal provided we can ensure that the label \(z\) of \(x\) does not contain pinches.

**Case I.** Suppose that \(u\) and \(v\) do not have \(t\)-letters. Let \(\tau\) be the number of \(t\)-annuli in \(\Delta\).

Clearly no two sides of these \(t\)-annuli have the same labels (otherwise we could remove the subdiagram bounded by these two sides), which implies as in the proof of Lemma 2.3 that the area of \(\Delta\) is at least \(O(\tau \log \tau)\). We can also assume that \(\tau > f(n)\) where \(f(n)\) is the recursive function from the quasi-proof. Indeed, if \(\tau\) is bounded by a recursive function for every \(\Delta\) then the area of \(\Delta\) is bounded by a recursive function too.

Clearly there exists a path in \(\Delta\) connecting \(p\) and \(p'\) and having length \(O(\tau)\). So we can assume that \(x\) is that path (and not the shortest path connecting \(p\) and \(p'\) as in the quasi-proof).

Suppose that the label \(z\) of \(x\) contains a pinch and \(x'\) is the corresponding subpath of \(x\). Then the first and the last edges of \(x'\) are \(t_j\)-edges belonging to two consecutive \(t_j\)-annuli \(T\) and \(T'\) in \(\Delta\). Since the label \(z'\) of \(x'\) is a pinch, there exists a diagram \(\Sigma\) consisting of one \(t_j\)-band such that the label of \(\partial \Sigma\) is \(z'z''\), and \(z''\) does not contain \(t\)-letters. Therefore we can cut \(\Delta\) along \(x'\), and patch the resulting hole by gluing in a copy of \(\Sigma\) and a copy of the mirror image \(\Sigma'\) of \(\Sigma\) glued together along the part of the boundary labeled by \(z''\). The resulting annular diagram \(\Delta'\) is not reduced. But instead of two \(t_j\)-bands \(T\) and \(T'\), \(\Delta'\) contains one \(t_j\)-band \(T''\) whose set of cells is the union of the sets of cells in \(T\), \(T'\), \(\Sigma\) and \(\Sigma'\). The annulus \(T''\) does not surround the hole of \(\Delta'\). Hence it bounds a disc subdiagram \(\Phi\) of \(\Delta'\). The boundary label of \(\Phi\) does not contain \(t\)-letters. Hence the boundary label of \(\Phi\) must be equal to 1 in the free group. Hence \(\Phi\) can be replaced by a diagram without cells. The new diagram \(\Delta''\) has fewer cells, a contradiction with the minimality of \(\Delta\). Thus \(z\) does not have pinches, and we are done.

**Case 2.** Suppose that \(u\) and \(v\) have \(t\)-letters.

The number \(s\) of maximal \(t\)-bands in \(\Delta\) is bounded by \(\min(|u|,|v|) \leq n = |u| + |v|\). Let us cut \(\Delta\) along a side \(q\) of a \(t_i\)-band. Let \(\Gamma\) be the resulting diagram. We can assume that \(q\) is the shortest among the sides of the \(t\)-bands in \(\Delta\). We can also assume that \(|q|\) is the smallest for all annular diagrams with the same boundary labels and the same area. This implies that there is no diagram with boundary label of the form \(\bar{z}u\bar{z}^{-1}v^{-1}\) whose area does not exceed the area of \(\Gamma\) and \(|z| < |q|\).

The van Kampen diagram \(\Gamma\) has contour \(pq(p')^{-1}(q')^{-1}\) where the labels of \(p\) and \(q\) are (cyclic shifts of) \(u\) and \(v\) respectively, and \(q\) and \(q'\) have the same label \(z\). We shall denote by \(q_-\) and \(q_+\) the initial and terminal vertices of \(q\). Two vertices \(V\) and \(V'\) in \(q\) and \(q'\) are called co-phase if their distances from \(q_-\) and \(q'_-\) (along \(q\) and \(q'\)) respectively are the same. We say that \(V\) is higher (lower) than \(V'\) if the distance from \(V\) to \(q_-\) is bigger (smaller) that the distance from \(V'\) to \(q'_-\).

Let \(T_1, \ldots, T_s\) be all maximal \(t\)-bands in \(\Gamma\) connecting \(p\) and \(p'\), ordered from \(q\) to \(q'\). So \(q\) is
a side of $T_l$, $q'$ is a side of $T_r$. It is easy to prove that the length of each maximal $t$-band in $\Gamma$
is at most $(|q| + n) \exp(Cn)$ for some constant $C$. Indeed, the label $w_i$ of a side of $T_r$ is equal
modulo $P$ to the label of $q$ multiplied on the left and on the right by two words of length at
most $|u| + |v|$. Therefore the length of $w_i$ in the base group cannot exceed $(|q| + n) \exp(C'n)$ for
some constant $C'$: when we reduce a pinch in a word, the length increases by a constant factor.
Since the number of cells in $T_l$ is $O(|w_i|)$, we obtain the desired inequality.

Since $|q| >> |u|$, the area of $\Gamma$ ( = the area of $\Delta_l$) is at most $O(|q| \exp(C(|u| + |v|)))$ for
some constant $C$. For every vertex $V$ on $q$ consider a shortest path $p(V)$ connecting $V$ with a vertex
in $q'$. Note that $p(V)$ does not exceed a constant times $s$. If the vertex $p(V)_+$ is co-phase with
$V$ then we say that $p(V)$ is parallel to $p$. We can assume that $p(V_1)$ does not cross $p(V_2)$ for any
$V_1 \neq V_2$. Indeed, if $p(V_1) = p_1p_2, p(V_2) = p'_1p'_2$ and the end points of $p_1, p'_1$ are the same, then
the lengths of $p_2$ and $p'_2$ are the same (otherwise the path $p(V_1)$ or $p(V_2)$ would not be shortest),
and we can replace $p(V_2)$ by $p'_1p_2$. Thus we can talk about a subdiagram $\Phi(V_1, V_2)$ bounded by
$p(V_1)$ and $p(V_2)$.

Suppose that there exist two vertices $V_1$ and $V_2$ such that $p(V_1)$ and $p(V_2)$ are parallel to $p$
and the labels of the paths $p(V_1)$ and $p(V_2)$ are the same. Then we can remove the subdiagram
$\Phi(V_1, V_2)$ of $\Gamma$. The resulting diaogram $\Gamma'$ will have boundary label of the form $\bar{z}u^{-1}v$ with $\bar{z}$
shorter than $|x|$, a contradiction. Hence the labels of all paths $p(V)$ that are parallel to $p$ are
different. Hence the number $\pi$ of such paths is at least an exponent in $C|u|$ for some constant $C$.
The length $\pi$ is small comparing to $|q|$. These paths cut $\Gamma$ into at most $e(n)$ conjugacy
subdiagrams where $e$ is a recursive function. One of these subdiagrams must have area bigger
than $C|q|/e(n)$ for some constant $C$. Thus we can deal with this “large” subdiagram instead of
$\Gamma$. Hence without loss of generality, we shall assume that $\Gamma$ does not have paths $p(V)$ that are
parallel to $p$ except possibly for $p$ and $p'$.

Now let us number all vertices of $q$ starting with $q_-$: $V_1, V_2, ....$. For each $j = 1, 2, ...,$, let $l(j)$
be the distance from $p(V_j)_+$ to $q_-$. Since the paths $p(V_j)$ do not intersect, and none of the paths
$p(V_j)$ are parallel to $p$ except possibly for $p(q_-)$ and $p(q_+)$, either $l(j) \geq j$ for all $j$ or $j \geq l(j)$
for all $j$ and the inequalities are strict except, possibly, for $V_j = q_-$ and $V_j = q_+$. We can assume
that the first possibility holds because otherwise we can turn $\Gamma$ upside down switching $u$ and $v$.

Let us define a sequence of vertices on $q$ as follows: $Q_1 = V_2$ and for every $j = 2, 3, ...$
let $Q_j = V_{j+2}(Q_{j-1})$. Note that $Q_j$ is co-phase with $p(Q_{j-1})$. Let us define the path $p_j$
as the composition of the subpath $Q_j - Q_{j-1}$ of $p$ and $p(Q_{j-1})$ is parallel to $p$. Then for some
small enough constant $c$ the number of paths $p_j$ of length $\leq c \log |q|$ is at most $\sqrt{|q|}$. Since
the length of $p(Q_{j-1})$ is recursively bounded in terms of $|u|$, we can assume that the number of
$j$’s such that the distance between $Q_{j-1}$ and $Q_j$ along $q$ is smaller than $c \log |q|$ is at most
$\sqrt{|q|}$. Since each $Q_j$ is higher than $Q_{j-1}$, the number of $j$’s for which this distance is not
smaller than $c \log |q|$ is at most $|q|/c \log |q|$. Hence the total number of points $Q_j$ is at most
$O(\sqrt{|q|} + |q|/\log |q|) = O(|q|/\log |q|)$.

Let $q_0$ be the subpath of $q$ connecting $V_1$ and $V_2$ (this path is simply an edge). Since $p(Q_{j-1})$
and $Q_j$ are co-phase, images of paths $q_0$ and all $p(Q_j)$ in $\Delta$ is a path $x$ connecting the boundary
components of $\Delta$. The length of $x$ is at most $O(|q|/\log |q|)$ times a recursive function in $n$ as
required. Thus, by Lemmas 2.6-2.9, it remains only to show that the label $z$ of $x$ does not have
pinches.

For a contradiction, suppose that $z$ contains a pinch $z'$ and $x'$ is a subpath of $x$ whose label
is $z'$. Then $z' = t_j^{\pm 1} u_j t_j^{\mp 1}$. The two $t$-bands intersecting $x'$ are consecutive $t$-bands $T_k$ and
$T_{k+1}$ for some $k$ in the annular diagram $\Delta_i$ (we consider $k$ modulo $s$ so $T_{s+1} = T_0$). Let us connect
$x'_-$ and $x'_+$ with vertices $R, R'$ on the inner contour of $\Delta_i$ along the sides of $T_k$ and
$T_{k+1}$ by paths $q_1, q_2$. The vertices $R, R'$ can be connected by a subpath $\tilde{p}$ of the boundary component
such that $\tilde{p}$ contains the end edges of the $t$-bands $T_k$ and $T_{k+1}$. If this quadrangle surrounds the hole of $\Delta_i$, we can repeat the same construction using another boundary component of $\Delta_i$. The resulting quadrangle won’t surround the hole in that case. Since these cases are similar, we can assume that the initial quadrangle does not surround the hole. Then this quadrangle is a van Kampen diagram over $P$. Since the label of $x'$ is a pinch, it is equal to a word without $t$-letters modulo $P$. Since the labels of $q_1$ and $q_2$ do not contain $t$-letters, the label of $\tilde{p}$ is equal modulo $P$ to a word without $t$-letters. Hence the (cyclic) word $u_i$ contains a pinch, a contradiction. \qed

3 S-machines

3.1 S-machines as HNN extensions of free groups

Probably the easiest way to view an $S$-machine $S$ in the sense of [11] is to consider $S$ as a group that is an HNN extension of a free group $F(Q, Y)$ generated by two sets of letters: state letters $Q = \cup_{i=1}^N Q_i$ and tape letters $Y = \cup_{i=1}^{N-1} Y_i$ where $Q_i$ are disjoint and non-empty. The sets $Q_i$ (resp. $Y_i$) are called parts of $Q$ (resp. $Y$).

We shall follow the tradition of calling state letters $q$-letters and tape letters $a$-letters, even though we shall use $k$ with indexes for state letters and $y$ with indexes for tape letters.

Instead of the set of stable letters we have a collection $\Theta$ of $N$-tuples of $\theta$-letters or rules. The components of $\theta$ are called brothers $\theta_1, \ldots, \theta_N$. In this paper, we always assume that all brothers are different. We set $\theta_{N+1} = \theta_1$, $Y_0 = Y_N = \emptyset$.

To every $\theta \in \Theta$, we associate two sequences of elements in $F(Q \cup Y)$: $B(\theta) = [U_1, \ldots, U_N]$, $T(\theta) = [V_1, \ldots, V_N]$, and a subset $Y(\theta) = \cup Y_i(\theta)$ of $Y$, where $Y_i(\theta) \subseteq Y_i$.

The words $U_i, V_i$ satisfy the following restriction:

(*) For every $i = 1, \ldots, N$, the words $U_i$ and $V_i$ have the form

$$U_i = v_{i-1}k_iu_i, \quad V_i = v'_{i-1}k'_iu'_i$$

where $k_i, k'_i \in Q_i$, $u_i$ and $u'_i$ are words in the alphabet $Y_{i-1}^{\pm 1}$, $v_{i-1}$ and $v'_{i-1}$ are words in the alphabet $Y_{i-1}^{\pm 1}$.

The generating set $X$ of $S$ consists of all $q$-, $a$- and $\theta$-letters. The relations are:

$$U_i\theta_{i+1} = \theta_iV_i, \quad i = 1, \ldots, s, \quad \theta_ja = a\theta_j$$

for all $a \in Y_j(\theta)$. The first type of relations will be called $(q, \theta)$-relations, the second type $(a, \theta)$-relations.

Sometimes we will denote the rule $\theta$ by $[U_1 \to V_1, \ldots, U_N \to V_N]$. This notation contains all the necessary information about the rule except for the sets $Y_i(\theta)$. In most cases it will be clear what these sets are. In the $S$-machines used in this paper, the sets $Y_i(\theta)$ will be mostly equal to either $Y_i$ or $\emptyset$. By default $Y_i(\theta) = Y_i$.

In order to simplify the notation, we will use the notation $v_{i-1}k_iu_i \to v'^{_{i-1}}k'^{_{i-1}}u'_i$ for a part of a rule when the corresponding $Y_i(\theta)$ is empty (a similar notation has been used in [7]).

Every $S$-rule $\theta = [U_1 \to V_1, \ldots, U_s \to V_s]$ has an inverse $\theta^{-1} = [V_1 \to U_1, \ldots, V_s \to U_s]$; we set $Y_i(\theta^{-1}) = Y_i(\theta)$. We always divide the set of rules $\Theta$ of an $S$-machine into two disjoint parts, $\Theta^+$ and $\Theta^-$ such that for every $\theta \in \Theta^+$, $\theta^{-1} \in \Theta^-$ and for every $\theta \in \Theta^-$, $\theta^{-1} \in \Theta^+$. The rules from $\Theta^+$ (resp. $\Theta^-$) are called positive (resp. negative).
Remark 3.1. 1. Every $S$-machine is indeed an HNN-extension of the free group $F(Y,Q)$ with finitely generated associated subgroups. The stable letters are $\theta_1$ for every $\theta \in \Theta$. We leave it as an exercise to find the associated subgroups.

2. Notice that in [11], a slightly different notation for rules of $S$-machines was used. Instead of, say, the rule

$$[v_0k_1u_1 \rightarrow v'_0k'_1u'_1, v_1k_2u_2 \rightarrow v'_1k'_2u'_2], \tag{3.3}$$

one would use the notation

$$[v_0k_1v_1k_2u_2 \rightarrow v'_0k'_1u'_1v'_1k'_2u'_2]. \tag{3.4}$$

But two relations corresponding to the rule (3.3):

$$\theta_1^{-1}v_0k_1u_1 = v'_0k'_1u'_1\theta_2^{-1}$$

and

$$\theta_2^{-1}v_1k_2u_2 = v'_1k'_2u'_2\theta_3^{-1}$$

are Tietze equivalent to one relation corresponding to (3.4):

$$\theta_1^{-1}v_0k_1v_1k_2u_2 = v'_0k'_1u'_1v'_1k'_2u'_2\theta_3^{-1}$$

since $\theta_2$ is expressible in terms the other generators. Therefore these notations are equivalent.

3.2 Bands

From now on, we shall only consider reduced van Kampen diagrams i.e. diagrams that do not contain cells that have a common edge and are mirror images of each other. Hence all van Kampen diagrams are assumed to be reduced. To study van Kampen diagrams over the group $S$ we shall use bands and trapezia as in [11], [1], etc.

Here we repeat necessary definitions from [11].

Definition 3.2. Let $M$ be a subset of $X$. An $M$-band $B$ is a sequence of cells $\pi_1, \ldots, \pi_n$ in a van Kampen diagram such that

- Each two consecutive cells $\pi_i$ and $\pi_{i+1}$ in this sequence have a common edge $e_i$ labeled by a letter from $M$.
- Each cell $\pi_i$, $i = 1, \ldots, n$ has exactly two $M$-edges, $e_{i-1}$ and $e_i$ (i.e. edges labeled by a letter from $M$).
- If $n = 0$, then $B$ is just an $M$-edge.

The counterclockwise boundary of the subdiagram formed by the cells $\pi_1, \ldots, \pi_n$ of $B$ has the form $e^{-1}q_1fq_2^{-1}$ where $e = e_0$ is an $M$-edge of $\pi_1$, $f = e_n$ is an $M$-edge of $\pi_n$. We call $q_1$ the bottom of $B$ and $q_2$ the top of $B$, denoted $\text{bot}(B)$ and $\text{top}(B)$.

Consider lines $l(\pi_i, e_i)$ and $l(\pi_i, e_{i-1})$ connecting a point inside the cell $\pi_i$ with midpoints of the $M$-edges of $\pi_i$. The broken line formed by the lines $l(\pi_1, e)$, $l(\pi_i, e_i)$, $l(\pi_i, e_{i-1})$, $l(\pi_n, f)$ is called the median of the band $B$. It connects the midpoints of $e$ and $f$ and lies inside the union of $\pi_i$. The $M$-edges $e$ and $f$ are called the start and end edges of the band. If $n = 0$, then the median is the midpoint of $e = f$. 
We say that an $M_1$-band and an $M_2$-band cross if their medians cross. We say that a band is an annulus if its start and end edges coincide. In this case the median of the band is a simple closed curve.

The subdiagram bounded by the median of an annulus in a disc diagram is called the inside diagram of this annulus.

Let $M_1$ and $M_2$ be two disjoint sets of letters, let $(\pi, \pi_1, \ldots, \pi_n, \pi')$ be an $M_1$-band and let $(\pi, \gamma_1, \ldots, \gamma_m, \pi')$ be an $M_2$-band. Suppose that:

- the medians of these bands intersect in two points $A, B$ inside $\pi$ and $\pi'$, and the parts of the medians between $A$ and $B$ form a simple closed curve,

- on the boundary of $\pi$ and on the boundary of $\pi'$ the pairs of $M_1$-edges separate the pairs of $M_2$-edges,

- the start and end edges of these bands are not contained in the region bounded by the medians of the bands.

Then we say that these bands form an $(M_1, M_2)$-annulus and the simple closed curve formed by the parts of medians of these bands is the median of this annulus. For every annulus we define the inside subdiagram of the annulus as the subdiagram bounded by the median of the annulus.

We shall call an $M$-band maximal if it is not contained in any other $M$-band.

As in [11], we can consider $q$-bands where $M$ is one of the sets $Q_i$, $\theta$-bands for every $\theta \in \Theta$, and $a$-bands where $M = \{a\} \subseteq Y$. Every cel of a $q$-band is a $q$-cell by definition. A $q$-cell is also a $(q, \theta)$-cell (and also $(\theta, q)$-cell) since it corresponds to a $(q, \theta)$-relation. The convention is that $a$-bands do not contain $q$-cells, and so they consist of $(a, \theta)$-cells ($= (\theta, a)$-cells) only.

The following lemma has been essentially proved in [11].

**Lemma 3.3.** A reduced van Kampen diagram $\Delta$ over $S$ has no $q$-annuli, $\theta$-annuli, $(q, \theta)$-annuli, $a$-annuli, $(a, \theta)$-annuli.

**Proof.** We assume that $\Delta$ is a counterexample with minimal area. This means in particular that the boundary of $\Delta$ is the boundary component of an annulus $A$, where $A$ has one of the types from the formulation of the lemma.

1. Suppose $A$ is a $q$-annulus. Then it consists of $(q, \theta)$-cells. Hence there is a maximal $\theta$-band $T$ in $\Delta$, whose first cell $\Pi_1$ and the last cell $\Pi_2$ belong to $A$. Being members of the same $q$-band $A$ and $\theta$-band $T$, the cells $\Pi_1$ and $\Pi_2$ cannot be neighbors in $T$ (the diagram is reduced). Hence $T$ and a part of $A$ form a $(\theta, q)$-annulus whose area is smaller than that of $\Delta$. This contradicts the choice of $\Delta$.

2. Suppose $A$ is a $\theta$-annulus. If it contains $q$-cells, then we come to a contradiction as in (1). Otherwise $\Delta$ has no $q$-cells since there is no counter-example of smaller area. The inner part of $A$ has no $\theta$-edges for the same reason. Hence $\Gamma$ has no cells corresponding to the relations of the group $S$. So, on the one hand, the inner label of $A$ is a cyclically reduced non-empty word in $Y$ since $\Delta$ is a reduced diagram, and on the other hand, this word is freely equal to $1$, a contradiction.

3. Suppose $A$ is a $(q, \theta)$-annulus. Then the maximal $q$-band $T$ of $A$ cannot have more than two cells because otherwise $\Delta$ would contain a smaller counterexample as in (2). Hence the length of $T$ is 2, and its cells are mirror copies of each other, a contradiction (we assumed that the diagrams are reduced).

4. Suppose $A$ is an $a$-annulus. Then its boundary labels are words in $\theta$-letter. This leads, as in (1), to a smaller $(a, \theta)$-annulus, a contradiction.

11
3.3 Trapezia

If $W = x_1...x_n$ is a word in an alphabet $X$, $Y$ is another alphabet, and $φ : X → Y \cup \{1\}$ (where 1 is the empty word) is a map, then $φ(W) = φ(x_1)...φ(x_n)$ is called the projection of $W$ onto $Y$. We shall consider the projections of words from $S$ onto $Θ$ (all $θ$-letters map to the corresponding element of $Θ$, all other letters map to 1), and the projection onto the alphabet $\{Q_1,...,Q_n\}$ (every $q$-letter maps to the corresponding $Q$, other letters map to 1).

**Definition 3.4.** The projection of the label of a side of a $q$-band onto the alphabet $Θ^{±1}$ is called the history of the band. The projection of the label of a side of a $θ$-band onto the alphabet $\{Q_1,...,Q_n\}$ is called the base of the band. Similarly we can define the history of a word and the base of a word. The base of a word $W$ is denoted by base$(W)$. It will be convenient instead of letters $Q_1,...,Q_n$, in base words, to use representatives of these sets. For example, if $k_1 ∈ Q_1$, $k_2 ∈ Q_2$, we shall say that the word $k_1ak_2$ has base $k_1k_2$ instead of $Q_1Q_2$.

**Definition 3.5.** Let $Δ$ be a reduced van Kampen diagram which has the contour of the form $p_1^{-1}q_1p_2q_2^{-1}$ where:

- $(TR_1)$ $p_1$ and $p_2$ are sides of $q$-bands,
- $(TR_2)$ $q_1$, $q_2$ are maximal parts of the sides of $θ$-bands such that $φ(q_1)$, $φ(q_2)$ start and end with $q$-letters,
- $(TR_3)$ for every $θ$-band $T$ in $Δ$, the labels of $top(T)$ and $bot(T)$ are reduced.

Then $Δ$ is called a trapezium. The path $q_1$ is called the bottom, the path $q_2$ is called the top of the trapezium, the paths $p_1$ and $p_2$ are called the left and right sides of the trapezium. The history of the $q$-band whose side is $p_1$ is called the history of the trapezium; the length of the history is called the height of the trapezium. The base of $p_1$ is called the base of the trapezium.

**Remark 3.6.** Property $(TR_3)$ is easy to achieve: by folding edges with the same labels having the same initial vertex, one can make the boundary label of any subdiagram in a van Kampen diagram reduced, see [11].

**Remark 3.7.** Notice that the top (bottom) side of a $θ$-band $T$ does not necessarily coincide with the top (bottom) side of the corresponding trapezium of height 1, and is obtained from $top(T)$ (resp. $bot(T)$) by trimming a few first and last $a$-letters. We shall denote the trimmed top and bottom sides of $T$ by $ttop(T)$ and $tbot(T)$.

3.4 Admissible words and computations

Using trapezia, one can now formally define admissible words and application of a rule to a word.

**Definition 3.8.** Let $Δ$ be a trapezium of height 1. Let $θ$ be the element of $Θ^{±1}$ whose representative (one of the brothers) is written on a side of that trapezium. Then the word $W$ written on the bottom of of $Δ$ is called an admissible word for $θ$. The word written on the top of $Δ$ is called the result of application of $θ$ to $W$ and is denoted by $θ \cdot W$. Clearly, $θ \cdot W$ is uniquely
determined by θ and W, and \( θ^{-1} \cdot (θ \cdot W) = W \) (hence the label of the top of the trapezium is an admissible word for \( θ^{-1} \)). We call a word admissible if it is admissible for some \( θ \in Θ \). For every word \( f = f_1 f_2 ... f_n \) in \( Θ \) we define \( f \cdot W \) as \( f_n \cdot (f_2 \cdot (f_1 \cdot W)) ... \). In particular, \( 1 \cdot W = W \) for every word \( W \).

Recall that we assume that all components in every \( N \)-tuple \( θ \) are different. In order for a rule \( θ \) to be applicable to a word \( W = q_1 u_1 q_2 ... u_{n-1} q_n \) where \( q_i \) are from \( Q_{j(i)}^\pm \), \( u_i \) is a word in \( Y^\pm \), the \((q, θ)\)-relations involving \( q_i \) and \( q_{i+1} \) must share a \( θ \)-brother, and that \( θ \)-brother must commute with all letters of \( u_i \) that are not involved in the relations containing \( q_i, q_{i+1} \). This and the definition of admissible word immediately imply the following lemma.

**Lemma 3.9.** Every admissible word of a rule \( θ \) has the form \( q_1 u_1 q_2 ... u_{n-1} q_n \) where for every \( i \) from 1 to \( n \) there exists \( j(i) \) such that \( q_i \in Q_{j(i)}^\pm \), and

- If \( q_i \in Q_{j(i)}^\pm \) then \( u_i \) is a group word in \( Y_{j(i)} \) and \( q_{i+1} \in Q_{j(i)+1} \cup Q_{j(i)}^{-1} \);
- If \( q_i \in Q_{j(i)}^{-1} \) then \( u_i \) is a group word in \( Y_{j(i)-1} \) and \( q_{i+1} \in Q_{j(i)} \cup Q_{j(i)-1}^{-1} \).

By Lemma 3.3, any trapezium \( ∆ \) of height \( h \geq 1 \) can be decomposed into \( θ \)-bands \( T_1, ..., T_h \) connecting the left and the right sides of the trapezium. The word written on the trimmed top side of one of the bands \( T_i \) is the same as the word written on the trimmed bottom side of \( T_{i+1} \), \( i = 1, ..., h \). Therefore with every trapezium \( ∆ \) we can associate a sequence of words \( W_1, W_2, ..., W_{h+1} \) and a sequence of rules \( θ_1, θ_2, ..., θ_h \) such that \( W_2 = θ_1 \cdot W_1, W_3 = θ_2 \cdot W_2, ..., W_{h+1} = θ_h \cdot W_h \).

This pair of sequences will be called a computation of \( S \) connecting \( W_1 \) and \( W_{h+1} \). We shall denote the computation by

\[
W_1 \rightarrow_{θ_1} ... \rightarrow_{θ_{h-1}} W_h \rightarrow_{θ_h} W_{h+1},
\]

or simply

\[
W_1 \rightarrow ... \rightarrow W_h \rightarrow W_{h+1}.
\]

The number \( h \) is called the length of the computation. Since we consider only reduced diagrams, the history of every trapezium is a reduced word. That word \( t = θ_1 θ_2 ... θ_h \) is called the history of computation. The area of the trapezium is called the area of computation. The length of the longest word \( W_i \) in this computation is called its width. It is also convenient to consider empty computations consisting of one word \( W \). The history of an empty computation is the empty word, the start and end words of this computation are equal to \( W \).

Notice that \( W_1 f' = f W_{h+1} \) for some words \( f, f' \) in \( θ \)- and \( a \)-letters whose projections onto \( Θ \) are equal to the history word \( t \). It is easy to see that \( |f| = O(|t|) \).

**Remark 3.10.** One can easily see that the computation \( W \rightarrow W_1 \rightarrow ... \) looks like a computation of a Turing machine with many heads, the \( q \)-letters. Heads can move left and right, change their states, and change \( a \)-letters (which play the role of tape letters) around them.

As for a usual Turing machine, we choose a distinguished stop word \( W \) from \( F(Q, Y) \).

We say that a word \( W \in F(Q, Y) \) is accepted if there exists a computation connecting this word and \( W \).

The following lemma immediately follows from the definition of a computation and Lemma 3.3.

**Lemma 3.11.** Let \( ∆ \) be a trapezium with bottom label \( W \) and top label \( W' \). Then there is a unique computation \( W \rightarrow ... \rightarrow W' \) whose history is the history of \( ∆ \).
We shall also need the following lemma.

**Lemma 3.12.** Let \( W_0 \rightarrow_{\theta_1} W_1 \rightarrow \ldots \rightarrow_{\theta_i} W_t \) be a computation of an \( S \)-machine \( S \). Suppose that \( W_i = W_j \) and \( \theta_i \neq \theta_{j+1}^{-1} \) for some \( i, j \), \( 1 \leq i < j < t \). Then

\[
W_0 \rightarrow_{\theta_1} W_1 \rightarrow_{\theta_2} W_2 \ldots \rightarrow_{\theta_i} W_i \rightarrow_{\theta_{i+1}} W_{i+1} \ldots \rightarrow_{\theta_t} W_t
\]

(3.5)

is again a computation of \( S \).

**Proof.** This surgery amounts to removing \( \theta \)-bands number \( i+1, \ldots, j \) (counted from the bottom) in the trapezium \( \Delta \) corresponding to the initial computation. Let us show that the new diagram \( \Delta' \) is reduced. Indeed, pairs of cells in the same \( \theta \)-band in \( \Delta' \) cannot cancel because the same pair of cells existed in \( \Delta \). The cells from different \( \theta \)-bands in \( \Delta' \) cannot cancel because either the same pair of cells exists in \( \Delta \) or one of these cells is in the \( i \)-th \( \theta \)-band of \( \Delta \), and the other one is in the \( j+1 \)-st \( \theta \)-band of \( \Delta \), and these cells cannot cancel in \( \Delta' \) because \( \theta_i \neq \theta_{j+1}^{-1} \) by our assumption. The conditions \((TR_1), (TR_2), (TR_3)\) obviously hold for \( \Delta' \). So \( \Delta' \) is a trapezium, and (3.5) is a computation.

**Lemma 3.13.** Assume that two admissible words \( W \) and \( W' \) are conjugate in the group \( S \). Then there exists a computation \( W \rightarrow \ldots \rightarrow W'' \) of \( S \) where \( W'' \) is a cyclic conjugate of \( W' \) that starts with a \( q \)-letter from the same \( Q_i \) as the first letter of \( W \).

**Proof.** By the van Kampen - Schupp lemma, there is a reduced annular diagram \( \Delta \) whose boundary components \( p \) and \( p' \) are clockwise labeled by \( W \) and \( W' \). It follows from Lemma 3.3 that, for some \( h \geq 0, \Delta \) is a union of concentric \( \theta \)-annuli \( T_1, \ldots, T_h \), where \( T_1 \) is attached to \( p \), \( T_2 \) has a common boundary component with \( T_1, \ldots, T_h \), and \( T_h \) is attached to \( p' \). Let \( k \) be the first letter of \( W \). Then a maximal \( k \)-band connects \( p \) and \( p' \). We may assume that \( h > 0 \). Cutting \( \Delta \) along a side of this \( k \)-band, we get a trapezium. By Lemma 3.11, there is a computation of \( S \) connecting \( W \) with a word \( W'' \) that is a cyclic shift of \( W' \). Since all \( q \)-edges of a \( q \)-band have labels from the same set \( Q_i \), the first letters of \( W \) and \( W' \) are from the same \( Q_i \).

**Remark 3.14.** Suppose that the base of a trapezium \( \Delta \) starts with a \( Q_1 \)-letter and ends with a \( Q_N \)-letter. Then the labels of the sides of the trapezium are the same and do not contain \( a \)-letters: it follows from the agreement that \( Q_{N+1} = Q_1 \), and that in every part of an \( S \)-rule of the form \( v_0kju_1 \rightarrow v'_0 k'_1 u'_1 \) (resp. \( v_{N-1} k_{N} u_{N} \rightarrow v'_{N-1} k'_{N} u'_{N} \)) the words \( v_0, v'_0 \) (resp. \( u_N, u'_N \)) are empty since they are words over empty alphabets \( Y_0 \) and \( Y_{N} \).

**Lemma 3.15.** Suppose that the stop word \( \bar{W} \) starts with a letter from \( Q_1 \) and ends with a letter from \( Q_N \) and has only one \( Q_1 \)-letter. Suppose that the language of accepted words is not recursive. Then the set of words that are conjugates of \( \bar{W} \) in \( S \) is not recursive. Hence \( S \) has undecidable conjugacy problem.

**Proof.** Since the first letter in \( \bar{W} \) is from \( Q_1 \) and the last letter is from \( Q_N \), the left and right sides of any trapezium with the bottom label \( \bar{W} \) are the same by Remark 3.14. Hence if \( \bar{W} \) is accepted, it is a conjugate of \( \bar{W} \). Conversely, if \( \bar{W} \) starts with a \( Q_1 \)-letter and is a conjugate of \( \bar{W} \), then by Lemma 3.13, there exists a computation \( W \rightarrow \ldots \rightarrow \bar{W}' \) where \( \bar{W}' \) is a cyclic conjugate of \( \bar{W} \) starting with a \( Q_1 \)-letter. Since \( \bar{W} \) contains only one \( Q_1 \)-letter, \( \bar{W}' = \bar{W} \), so \( \bar{W} \) is accepted.
3.5 A slight modification of the $S$-machine from [11]

Let $L$ be a recursively enumerable language over an alphabet $X$. Then by [11, Proposition 4.1], there exists an $S$-machine $S$ recognizing $L$ in the sense of Lemma 3.16 below.

For that $S$-machine, $N = 17$, so the set of $q$-letters is partitioned into 17 subsets which will be more convenient to denote by $K_1, ..., K_N$. The elements of $K_i$ will be denoted by $k_i(j)$. The stop word $W$ is $k_1(0)k_2(0)...k_N(0)$. The set $X$ is contained in $Y_1$, and for every positive word $u$ in $X$ we denote $\sigma(u) = k_1(1)uk_2(1)...k_N(1)$. The following lemma is proved in [11].

**Lemma 3.16.** [11, Proposition 4.1] (a) For every $u \in L$ there exists a computation $\sigma(u) \rightarrow_{\theta_1} \cdot \rightarrow \tilde{W}$ consisting of positive words.

(b) For every word $u \notin L$ over $X$ the word $\sigma(u)$ is not accepted.

Let us modify $S$ a little. We add two new sets of state letters $K_{N+1} = \{k_{N+1}\}, K_{N+2} = \{k_{N+2}\}$, and a new state letter $\hat{k}_j$ in every $K_j$ and a new set $Y_{N+1} = \{\alpha\}$ of tape letters. We also add two rules $\eta_0, \eta_1$ of the same form $[k_{N+1} \rightarrow k_{N+1}, k_{N+2} \rightarrow \alpha k_{N+2}, \hat{k}_j \rightarrow \hat{k}_j, j = 1, ..., N]$. Note that now the new number of parts of $Q$ is $N + 2$, so we have to count modulo $N + 2$ and instead of the assumption that $\theta_{N+1} = \theta_1$ we have to assume that $\theta_{N+3} = \theta_1$.

Notice that the new machine admits computations of the form

$$\hat{k}_1 ... \hat{k}_N k_{N+1} k_{N+2} \rightarrow \eta_0 \rightarrow \eta_0 \hat{k}_1 ... \hat{k}_N k_{N+1} \alpha^m k_{N+2} \rightarrow \eta_{1-1} \rightarrow \eta_{1-1} \hat{k}_1 ... \hat{k}_N k_{N+1} k_{N+2}. \quad (3.6)$$

We keep notation $S$ for the new $S$-machine and we keep notation $N$ for the number of parts of the set of $Q$-letters (instead of $N + 2$).

Our goal is to cross-breed $S$ with another $S$-machine in order to slow it down whilst preserving the width of computations.

3.6 The adding $S$-machine

Let $A$ be a finite set of letters. Let the set $A_1$ be a copy of $A$. It will be convenient to denote $A$ by $A_0$. For every letter $a_0 \in A_0$ let $a_1$ denote its copy in $A_1$. Consider the following auxiliary “adding” $S$-machine $Z(A)$.

Its set of state letters is $P_1 \cup P_2 \cup P_3$ where $P_1 = \{L\}, P_2 = \{p(1), p(2), p(3)\}, P_3 = \{R\}$. The set of tape letters is $Y_1 \cup Y_2$ where $Y_1 = A_0 \cup A_1$ and $Y_2 = A_0$.

The machine $Z(A)$ has the following positive rules (there $a$ is an arbitrary letter from $A$). The comments explain the meanings of these rules.

- $r_1(a) = [L \rightarrow L, p(1) \rightarrow a_1^{-1} p(1) a_0, R \rightarrow R]$.  
  *Comment.* The state letter $p(1)$ moves left searching for a letter from $A_0$ and replacing letters from $A_1$ by their copies in $A_0$.

- $r_{12}(a) = [L \rightarrow L, p(1) \rightarrow a_0^{-1} a_1 p(2), R \rightarrow R]$.  
  *Comment.* When the first letter $a_0$ of $A_0$ is found, it is replaced by $a_1$, and $p$ turns into $p(2)$.

- $r_2(a) = [L \rightarrow L, p(2) \rightarrow a_0 p(2) a_0^{-1}, R \rightarrow R]$.  
  *Comment.* The state letter $p(2)$ moves toward $R$. 

15
\( r_{21} = [L \rightarrow L, p(2) \xrightarrow{t} p(1), R \rightarrow R], Y_1(r_{21}) = Y_1, Y_2(r_{21}) = \emptyset. \)

Comment. \( p(2) \) and \( R \) meet, the cycle starts again.

\( r_{13} = [L \xrightarrow{t} L, p(1) \rightarrow p(3), R \rightarrow R], Y_1(r_{13}) = \emptyset, Y_2(r_{13}) = A_0. \)

Comment. If \( p(1) \) never finds a letter from \( A_0 \), the cycle ends, \( p(1) \) turns into \( p(3) \); \( p \) and \( L \) must stay next to each other in order for this rule to be executable.

\( r_3(a) = [L \rightarrow L, p(3) \rightarrow a_0p(3)a_0^{-1}, R \rightarrow R], Y_1(r_3(a)) = Y_2(r_3(a)) = A_0 \)

Comment. The letter \( r_3 \) returns to \( R \).

For every letter \( a \in A \) we set \( r_i(a^{-1}) = r_i(a)^{-1} \) (i = 1, 2, 3).

The natural projection of every admissible word of \( Z(A) \) onto \( A \) takes \( a_1 \) and \( a_0 \) to \( a \), and all other letters to 1. The following lemma is obvious.

Lemma 3.17. Let \( \theta \) be a rule in \( Z(A) \), \( \text{W} = q_1u_1q_2...u_nq_{n+1} \) with base \( q_1q_2...q_{n+1} \) be an admissible word for \( \theta \). Then \( \theta \cdot \text{W} = q'_1\bar{w}_1u_1\bar{w}_2q'_2\bar{w}_2u_2...w_{n+1}q'_{n+1} \) where \( w_j \) and \( \bar{w}_j \) are empty if \( q_j \in \{ L, R \}^{\pm 1} \) and if \( q_j = p \in \{ p(1), p(2), p(3) \} \), then \( w_j = w(p) \) and \( \bar{w}_j = \bar{w}(p) \) determined by the following formulas:

\[
w(p) = \begin{cases} 
    a_1^{-1} & \text{if } \tau_i = r_1(a), p = p(1), \\
    a_0 & \text{if } \tau_i = r_2(a), p = p(2), \\
    a_0 & \text{if } \tau_i = r_3(a), p = p(3), \\
    (a_0 - a_1) & \text{if } \tau_i = r_{12}(a)\epsilon, \epsilon = \pm 1, p = p(1) \text{ if } \epsilon = 1, p = p(2) \text{ if } \epsilon = -1, \\
    \emptyset & \text{if } \tau_i \in \{ r_{13}^{\pm 1}, r_{21}^{\pm 1} \}.
\end{cases}
\]

Finally if \( q_j = p \in \{ p(1)^{-1}, p(2)^{-1}, p(3)^{-1} \} \), then \( w_j = \bar{w}(p^{-1})^{-1} \), \( \bar{w}_j = w(p^{-1})^{-1} \).

Lemma 3.18. Suppose that an admissible word \( \text{W} \) has the form \( \text{LuvR} \) (resp. \( p^{-1}upvR \)) where \( u, v \) are words in \( (A_0 \cup A_1)^{\pm 1} \). Let \( \theta \cdot \text{W} = Lu'p'v'R \) (resp. \( \theta \cdot \text{W} = (p')^{-1}u'p'v'R \)). Then the projections of \( uv \) and \( u'v' \) (resp. \( v^{-1}uv \) and \( (v')^{-1}u'v' \)) onto \( A \) are freely equal.

Remark 3.19. If we replace every letter in \( A_i \) by its index \( i \), then every word in the alphabet \( A_0 \cup A_1 \) turns into a binary number \( b(u) \). If the machine starts with the word \( L_1p(1)R \) where \( u \) is a positive word in \( A_0 \), then \( b(u) = 0 \) and each cycle of the machine amounts to adding 1 to \( b(u) \). After \( 2^{\|u\|} \) cycles the machine stops, the admissible word becomes \( L_1p(3)R \). Let us compute the length of this computation.

Notice that if \( u = (b_1)_{0}...(b_n)_{0} \) for some \( b_i \in A \) and at the beginning of a cycle of the computation the last \( k \) \( a \)-letters in the admissible word are from \( A_1 \) and the \( a \)-letter number \( n - k \) is from \( A_0 \) \((k < n)\), then this cycle of the computation has the following history
(the letter \( p(1) \) moves left searching for the first letter from \( A_0 \) and replacing every \( a \)-letter from \( A_1 \) by the corresponding letter from \( A_0 \); then \((b_{n-k})_0\) is replaced by \((b_{n-k})_1\) and \(p(1)\) is replaced by \( p(2) \); then \( p(2) \) moves right, and after it meets with \( R \), it is replaced by \( p(1) \) again). The length of the cycle is \( 2k + 2 \). The number of cycles of this length is \( 2^n - k \). Therefore the total length of all of these cycles is \( \sum_{k=0}^{n-1} (2k + 2)^{n-k} \). The length of the last cycle is \( 2n + 1 \). Hence the total length of the computation is

\[
\sum_{k=0}^{n-1} (2k + 2)^{n-k-1} + 2n + 1 = 2^n \left( \sum_{k=0}^{n-1} \frac{k + 1}{2k} \right) + 2n + 1 < 6 \cdot 2^n
\]

for every \( n \) since \( \sum_{k=0}^{\infty} \frac{k+1}{2^k} = 4 \) and \( 2 \cdot 2^n \geq 2n + 1 \) for every \( n \geq 0 \). Hence the length of the computation is between \( \frac{2^n}{6} \) and \( 6 \cdot 2^n \).

Informally speaking, the remaining part of the section is devoted to describing all possible computations of the machine \( Z(A) \). In the next section, we shall use that information to describe computation of a composition of the \( S \)-machine from [11] and \( Z(A) \). We first consider the case when the base is \( LpR \). We show that if the machine \( Z(A) \) works without changing the length of an admissible word, then the computations are in some sense unique and are subcomputations of the computations described in Remark 3.19. Then we consider computations where the lengths of words can change. We show (and this is a standard feature of \( S \)-machines used in [11] and other papers) that as soon as the length of an admissible word increases during the computation, it cannot decrease again later, that is there are no trapezia that look like a honey pot (width in the middle is bigger than the width of the bottom and the top). If the base is not normal, say, it is \( p^{-1}pR \), we show that, again, there are no very wide “honey pots”.

**Lemma 3.20.** Suppose \( W \) is a word with base(\( W \)) = \( LpR \). Then there are at most two rules of \( Z(A) \) that are applicable to \( W \) word without changing its length.

**Proof.** Let \( W = LupeR \). Assume that \( \theta \cdot W = W' = Lu'p'v'R \) for a rule \( \theta \). If \( \theta = r_1(a) \) for some \( a \in A^{\pm 1} \) then the only ways equality \( |W| = |W'| \) can occur is when either \( v \) starts with \( a_0 \) or \( u \) ends with \( a_1 \) (by Lemma 3.17). Thus there are at most two choices for \( a \) in that case. A similar statement holds for \( r_2(a) \), \( r_3(a) \). Moreover, since applicability of \( r_1(a) \) determines the value of \( p = p(i) \), rules of only one of these three types can apply to \( W \). For \( \theta = r_{12}(a) \), we have \( |W| = |W'| \) only if \( u \) ends with \( a_0 \) but not with \( a_1^{-1}a_0 \) and \( p = p(1) \). Hence if \( r_{12}(a) \) does not change the length then only one other rule does not change the length (either \( r_1(b) \) where \( b^{-1} \) is the first letter of \( v \) or \( r_{21}^{-1} \) if \( v \) is empty). Similar arguments hold in all other cases.

**Lemma 3.21.** For every admissible word \( W \) with base(\( W \)) = \( LpR \), every rule \( \theta \) applicable to \( W \), and every natural number \( t > 1 \), there is at most one computation \( W \rightarrow_{\theta} W_1 \rightarrow ... \rightarrow W_t \) of length \( t \) where the lengths of the words are all the same.

**Proof.** Let \( W \rightarrow_{\theta} W_1 \rightarrow_{\theta_1} W_2 \rightarrow ... \rightarrow_{\theta_{t-1}} W_{t-1} \) be a computation where all words have the same length. By Lemma 3.20, \( W_2 \) is completely determined by \( W_1 \) (because the history of the computation is a reduced word, and so \( \theta_1 \neq \theta^{-1} \)), \( W_3 \) is completely determined by \( W_2 \), etc.

For every word \( f \) and every \( i \leq |f| \) we denote the prefix of \( f \) of length \( i \) by \( f[i] \).
Lemma 3.22. Let \( W = L v p u R \), where \( p \in \{ p(1), p(2) \} \), \( u \) is a word in \( A_0^{\pm 1} \), and \( v \) is a word in \((A_0 \cup A_i)^{\pm 1}\). Suppose that a non-empty computation

\[
W = W_1 \rightarrow \cdots \rightarrow f \cdot W
\]

is such that all words \( W_s \) in the computation have the same length, \( f \cdot W = L v' p'u R, p' \in \{ p(1), p(2) \} \), and either

1. \( p = p(2) \) and \( f[1] = r_2(a) \) or
2. \( p = p(1) \) and \( f[1] = r_1(a)^{-1} \)

where \( a \) is the first letter of \( u \). Suppose also that \( f[j] \cdot W \) does not have the form \( L v'' p'' u R \) for every \( 1 < j < |f| \). Then \( u \) is a positive word, \( v' = v \) and \( p' = p(1) \) in case (1) and \( p' = p(2) \) in case (2). Moreover, under the above conditions, the computation (3.10) is unique.

Proof. We shall consider only the case when \( p = p(2) \). The other cases are similar. Let \( u = b_1 b_2 \cdots b_n \), \( b_i \in A_0^{\pm 1} \). Suppose that \( u \) is not positive, \( u = u[k]b^{-1} b_{k+2} b_n \), where \( b_{k+2}, \ldots, b_n \in A \) (it could happen that \( b_{k+2} \cdots b_n \) is empty, i.e. \( n = k + 1 \)).

Let \( f[1] = \theta \). Since all words in the computation have the same length, we can conclude, by Lemma 3.21, that for every \( j \) between 1 and \( |f| \) there exists exactly one computation of length \( j \) starting with \( W \rightarrow \theta \cdot W \). The history of that computation must be \( f[j] \).

Using Remark 3.19 and (3.9) one can find a computation \( W \rightarrow \theta \cdot W \rightarrow \cdots \) with history \( g \) of the form

\[
g = r_2(b_1) \cdots r_2(b_n) r_2 r_1(b_n) r_2(b_n) r_2 r_1(b_n) \cdots r_1(b_{k+2}),
\]

and \( g \cdot W = L v u b^{-1} p(1) b_{k+2} b_n R \). Therefore either \( g \) is a prefix of \( f \) or \( f \) is a prefix of \( g \).

Note that \( f \) cannot be a prefix of \( g \) because \( f \cdot W = L v' p'u R \) and there is no \( j \) such that \( g[j] \cdot W \) has that form. Hence \( g \) is a prefix of \( f \), i.e. \( g = f[j] \) for some \( j \). Moreover \( j < |f| \).

Notice that since \( b \in A_0 \), there is no rule except for \( r_1(b_{k+2})^{-1} \) (if \( n \geq k + 2 \)) or \( r_{21}^{-1} \) (if \( n = k + 1 \)) which can be applied to that word without increasing its length. This contradicts the assumption that \( f \) is reduced. Thus \( u \) is positive.

Now we can consider the computation with the history of the form

\[
h = r_2(b_1) \cdots r_2(b_n) r_2 r_1(b_n) r_2(b_n) r_2 r_1(b_n) \cdots r_1(b_{k+1})
\]

(see (3.9) again) such that \( h \cdot W = L v(1) u R \) and words \( h[i] \cdot W, i < |h| \), do not have the form \( L v'' p'' u R \). Since no words \( f[i] \cdot W \) have that form by assumption, we can conclude that \( f = h \). This completes the proof. \( \square \)

Now we shall find out what happens when the lengths of the words change during a computation.

Lemma 3.23. Let \( W \), be an admissible word with \( \text{base}(W) = L p R \). Let \( \theta \in \{ r_1(a)^{\pm 1}, r_2(a)^{\pm 1}, r_3(a)^{\pm 1} \} \). Suppose that \( |\theta \cdot W| > |W| \). Then for every computation \( W \rightarrow \theta W_1 \rightarrow \theta W_2 \) we have \( |W_2| > |W_1| \). Moreover, if \( \theta' \) is not of the form \( r_1(a)^{\pm 1} \), then the \( p \)-letters in \( W_2 \) and \( W_1 \) are the same and \( W_2 \) does not have a 2-letter subword of the form \( p R \).

Proof. Suppose that \( \theta = r_1(a) \) (the other cases are similar). If \( |\tau \cdot W| > |W| \), then \( \tau \cdot W = L u a^{-1} p(1) a_1 v R \) where the right hand side is a reduced word. It is easy to see that the only rule that can apply to \( \tau \cdot W \) without increasing the length is \( r_1^{-1} \), and the only type of rules that can apply are \( r_1(b) \) and \( r_{12}(b) \). This immediately implies both statements of the lemma. \( \square \)
Lemma 3.24. Let \( W = L_{vp} u R \) and \( \text{base}(W) = L p R \). Suppose that \(|\theta \cdot W| > |W|\). Then for every computation \( W \rightarrow_\theta W_1 \rightarrow W_2 \rightarrow \ldots \rightarrow f \cdot W \), we have \(|W_i| > |W|\) for every \( i \geq 1 \).

Proof. By contradiction, suppose that there exists a computation \( W \rightarrow_\theta W_1 \rightarrow \ldots \rightarrow f \cdot W \) such that \(|f \cdot W| \leq |W|\). Consider such a computation with the smallest \(|f| = t\) and smallest \(|u|\) for all such computations of length \( t\). Then \(|W| < |W_1| = \ldots = |W_{t-1}| > |W_t|\). By Lemma 3.23, \( \theta = r_{12}(a)^{\pm 1}\) for some \( a \).

We have \( \theta \cdot W = L_{vp}(a_0^{-1} a)^{\pm 1} p(m) u R \) for some \( \epsilon \in \{-1, 1\}, m = (\epsilon + 3)/2 \), and some words \( u, v \) where all letters in \( u \) belong to \( A_0^{\pm 1}\). We shall assume that \( \epsilon = 1\) (the other case is similar). So \( m = 2\).

Suppose that the letter \( a_1 \) inserted by \( f[1]\) is not touched during the computation \( W_1 \rightarrow \ldots \rightarrow W_{t-1}\) (i.e. it is not cancelled with a letter inserted by one of the rules of this computation).

Since the letters of \( u \) do not disappear during the computation \( W_1 \rightarrow \ldots \rightarrow W_{t-1}\) (they may only change indices from 0 to 1 by Lemma 3.18), the word \( W_{t-1}\) has the form \( L_{vp}(a_0^{-1} a_1 u_1 p u_2)\) where \( u_2\) is a suffix of \( u \) and \( u_1\) is obtained from the corresponding prefix of \( u \) by changing indices of the letters. Applying Lemma 3.23 to the inverse computation \( W_t \rightarrow W_{t-1} \rightarrow \ldots \rightarrow W \) we conclude that the last rule \( \theta' \) in \( f \) is \( r_{12}(b)^{\pm 1}\) for some \( b \). Hence \( W_t = L_{vp} p' u_2 R \). Since \(|W_t| < |W_{t-1}|\) and \( f \) is a minimal counterexample, we conclude that \( u_2\) must be equal to \( u\). So \( u_1\) must be empty (otherwise the computation \( W_t \rightarrow W_{t-1} \rightarrow \ldots \rightarrow W \) would be a smaller counterexample since it has the same length as \( W_1 \rightarrow \ldots \rightarrow W_t\), but starts with \( W_t = L_{vp} p' u_2 R\) with \(|u_2| < |u|\), and either \( p = p(2)\), and the last rule of \( f \) is \( r_{12}(a)^{-1}\) or \( p = p(1)\), the first letter of \( u \) is \( a_0^{-1}\), and the last rule of \( f \) is \( r_{1}(a)^{-1}\). By Lemma 3.22, both cases are impossible.

Therefore we can assume that \( a_1 \) is touched by the computation. Hence, by Lemma 3.22 applied to the computation \( W_1 \rightarrow \ldots \rightarrow W_{t-1}, u \) is a positive word, and if \( s \) is the number of the rule that touches \( a_1\), we have \( s \leq t - 1\), \( f[s] \cdot W = L_{vp}^{-1} p(1) a_0 R\) and the rule number \( s \) in \( f \) is \( r_1(a)\). Since \( u \) is positive, the word \( a_0 u\) is reduced. But then every rule of \( Z(A)\) except \( r_1(a)^{-1}\) would increase the length of the word, a contradiction. This completes the proof of the lemma.

Lemma 3.25. Let \( \text{base}(W) = L p R \). Then for every computation \( W = W_0 \rightarrow W_1 \rightarrow \cdots \rightarrow W_t = f \cdot W \) of the \( S \)-machine \( Z(A)\):

1. \(|W_i| \leq \max(|W_0|, |f \cdot W|), i = 0, \ldots, t,\)

2. If \( W = L_{up} R \) where \( p = p(1)\) (resp. \( p = p(3)\), \( f \cdot W \) contains \( p(3) R \) (resp. \( p(1) R \)) and all \( a\)-letters in \( W, f \cdot W \) are from \( A_0^{\pm 1}\), then the length of \( f \) is between \( 2^{|u|}\) and \( 6 \cdot 2^{|u|}\), \( u \) is a positive word, and all words in the computation have the same length.

Proof. 1. Immediately follows from from Lemma 3.24.

2. We consider the case when \( W = L_{up}(1) R\), the other case is similar. If \( u \) is empty, the statement is obvious. So assume that \( u \) is not empty. By Lemmas 3.24 and 3.18, all words in the computation have the same length.

The letter \( p(3)\) can occur only after rule \( r_{13}\) is executed. The admissible word \( f[i] \cdot W\) to which \( r_{13}\) is applied must have the form \( L_{p(1)} u R\). Indeed, the \( a\)-letters in \( f[i] \cdot W\) must be from \( A\) (by the definition of \( r_{13}\)), and the projection of \( W\) and \( f[i] \cdot W\) onto \( A\) must coincide by Lemma 3.18. The word \( f[i-1] \cdot W\) in the computation must be \( L a_1 p(1) u' R\) where \( a_0\) is the first letter of \( u\), \( a_0 u' = u\). Hence during the computation the first letter of \( u\) must change the index from 0 to 1. Hence for some \( i_2 < i_1\), we have \( f[i_2] \cdot W = L a_0 p(2) u R\) and \( u[1]\) is a positive letter. Applying now Lemma 3.22 to the smallest initial part of the subcomputation
If \( f[i_2] \cdot W, ..., f[i - 1] \cdot W \) satisfying the conditions of this lemma, we conclude that \( u \) is a positive word.

The first letter in \( f \) is either \( r_{12}(a) \) or \( r_{21}^{-1} \). Suppose first that \( f[1] = r_{21} \).

By Remark 3.19 there exists a computation connecting \( \text{Lup}(1)R \) and \( \text{Lup}(3)R \). It has the history \( g = r_{12}(b_1)r_{21}r_1(b_1)r_{12}(b_2)r_2(b_1) \cdot r_{13}(b_2)r_3(b_1) \) where \( u = b_s \ldots b_2 b_1, b_1 \in A_0 \). Hence, by Lemma 3.21, either \( g \) is a prefix of \( f \) or \( f \) is a prefix of \( g \). But note that there is only one rule applicable to the word \( \text{Lup}(3)R \) that does not increase its length (namely \( r_{3}(b_1)^{-1} \)). Hence \( f = g \) and so \( |f| \) is between \( 2 |u| \) and \( 6 \cdot 2 |u| \).

Now let \( f[1] = r_{21}^{-1} \). Then, as in the previous paragraph, we deduce that \( f \) must start with \( f[s + 1] = r_{21}^{-1}r_2(b_1^{-1})r_2(b_2^{-1}) \ldots r_2(b_s^{-1}) \). But then \( f[s + 1] \cdot W = \text{Lup}(2)uR \) and there is no rule except \( r_2(b_s) \) that can be applied to this word without increasing the length. The \( s + 2 \)-nd rule in \( f \) cannot be \( r_2(b_s) \) since \( f \) is reduced. Thus this case is impossible.

**Lemma 3.26.** Let \( W = W_0 = (p(0))^{-1}u_0p(0) \) be an admissible word with \( \text{base}(W) = p^{-1}p \) \( (p(0) \in \{p(1), p(2), p(3)\}) \). Let \( W \rightarrow_\theta W_1 \rightarrow_\theta W_2 \ldots \rightarrow_\theta W_t \) be a computation. Let \( W_j = (p(j))^{-1}u_jp(j) \), \( u_j = w_j^{-1}w_{j-1}w_j \) where \( w_j \) is defined in Lemma 3.17, \( j = 1, 2, \ldots \). Suppose that none of \( p(j) \) is equal to \( p(3) \). Then:

1. The word \( \theta_1 \theta_2 \ldots \) is a subword of the word of the form \( r_1 \ldots r_1x_1r_2 \ldots r_2y_1 \ldots r_1x_2 \ldots \), where \( r_i \) stands for any \( r_i(a), \) \( x_j \in \{r_{12}(a), r_{21}^{-1} \mid a \in A \}, \) \( y_j \in \{r_{12}(a)^{-1}, r_{21}^{-1} \mid a \in A \} \).

2. If none of the rules \( \theta_1, \theta_2, \ldots \) is \( r_{21}^{-1} \), then none of the words \( w_1, w_2, \ldots, w_t \) is empty, and the product \( w_1 \ldots w_t \) is a freely reduced word.

3. The word \( \theta_1 \ldots \theta_t \) is completely determined by the rule \( \theta_1 \) and the word \( w_1 w_2 \ldots w_t \).

**Proof.** 1. The first statement is obvious.

2. Statement 1 and Lemma 3.17 imply that the sequence \( w_1, w_2, \ldots, w_t \) is a subsequence of the sequence

\[
a_1(1), \ldots, a_1(t_1), (a_0(t_1 + 1)^{-1}a_1(t_1 + 1))^{\epsilon_1}, a_0(t_1 + 2), \ldots, a_0(t_2),
(a_0(t_2 + 1)^{-1}a_1(t_2 + 1))^{\delta_1}, a_1(t_2 + 2), \ldots
\]

where \( a_i(j) \in A_i^{\pm 1}, \epsilon_j \in \{0, 1\}, \delta_j \in \{0, -1\} \), and we set \( v^0 = 1 \) for every word \( v \). Moreover if \( r_{21}^{-1} \) does not appear in the computation, none of the \( \delta_j \) and \( \epsilon_j \) are equal to 0. Hence \( w_1 w_2 \ldots w_t \) is freely reduced.

3. The third statement follows from the form of the sequence (3.11). Indeed, if \( \theta_1 = r_1(a) \), then \( w_1 = a_1 \in A_1^{\pm 1} \), and \( \theta_2 \) is completely determined by the next one or two letters of the word \( w_1 w_2 \ldots w_t \): if the second letter is \( b_1 \), then \( \theta_2 = r_1(b) \); if it is \( b_0^{-1} \) and the third letter is \( b_1 \), then \( \theta_2 = r_{12}(b) \); if it is \( b_0 \) and the third letter either does not exist or has index 0, then \( \theta_2 = r_{21} \). Similarly for other choices of \( \theta_1 \), the second and the third letter of \( w_1 w_2 \ldots w_2 \) completely determine \( \theta_2 \). Now we can complete the proof by induction on \( t \).

**Lemma 3.27.** Suppose \( \text{base}(W) \in \{\text{LpR}, p^{-1}pR\} \), both \( W \) and \( f \cdot W \) contain \( p(1)R \) (resp. \( p(3)R \)) and all a-letters in \( W, f \cdot W \) are from \( A_0 \). Then \( f \) is empty.

**Proof.** The statement can be proved in the same way as part 2 of Lemma 3.25 provided \( \text{base}(W) = \text{LpR} \).

Let \( \text{base}(W) = p^{-1}pR, W = p(1)^{-1}up(1)R = f \cdot W \) where all letters in \( u \) are from \( A_0^{\pm 1} \) (the case of \( p(3) \) is similar). By Lemma 3.26, part 1, \( u = w^{-1}uw \) where \( w = w_1 w_2 \ldots \) where \( w_j \)'s are determined by the formulas from Lemma 3.17, the product \( w_1 w_2 \ldots \) considered as a

20
word in the alphabet $(A_0 \cup A_1)^{\pm 1} \cup \{1\}$ (i.e. we do not throw away the empty factors) does not contain subwords $xx^{-1}, x1x^{-1}$ or $11$. By Lemma 3.17, $\ldots \bar{w}_2 \bar{w}_1 = 1$. Then, by Lemma 3.17, since all letters of $u$ are from $A_0$, all rules in $f$ are from $\{r_2(a), r_3(a), r_{12}^{\pm 1}, r_{13}^{\pm 1}\}$. But in that case $\bar{w}_j = w_j^{-1}$ by Lemma 3.17. Therefore the product $\ldots \bar{w}_2 \bar{w}_1$ is reduced. Since that product is 1, $u_1, u_2, \ldots, \bar{w}_1, \bar{w}_2, \ldots$ are empty words. Since the factorization $u_1u_2\ldots$ does not contain 11, we conclude that either $f$ is empty or $f = r_{21}^{\pm 1}$ (by Lemma 3.17). The second option is clearly impossible. 

We shall need the following general statement from [8].

**Lemma 3.28 ([8], Lemma 8.1).** For arbitrary elements $u, v, w$ of $F$ and any integer $t \geq 0$, the length of an arbitrary product $w^i w v^j$ in $F$ is not greater than $2(|u| + |v| + |w|) + |w^i w v^j|$ provided $0 \leq j \leq t$.

**Lemma 3.29.** Suppose that one of the following conditions for an admissible word $W$ of $Z(A)$ is satisfied (there $p = \{p(1), p(2), p(3)\}$):

1. $W$ does not contain a $p$-letter.
2. $\text{base}(W) = Lp p^{-1}$;
3. $\text{base}(W) = pp^{-1} p$;
4. $\text{base}(W) = p^{-1} p R$;
5. $\text{base}(W) = Lp R$.

Then the width of any computation

$$W = W_0 \rightarrow_{\theta_0} W_1 \rightarrow_{\theta_1} \ldots \rightarrow_{\theta_{t-1}} W_t$$

is at most $C \max(|W|, |W_t|)$ for some constant $C$.

**Proof.** 1 is obvious: the length of the admissible word does not change during the computation.

2. Let $\text{base}(W) = Lpp^{-1}$. Let $W = W_0 \rightarrow_{\theta_0} W_1 \rightarrow_{\theta_1} \ldots \rightarrow_{\theta_{t-1}} W_t$ be a computation. Then $W_i = L u_i q_i v_i q_i^{-1}$, where $q_i \in \{p(1), p(2), p(3)\}$. It is easy to see that for each $i = 0, \ldots, t - 1$, $u_{i+1} = u_i w_i, v_{i+1} = \bar{w}_i^{-1} v_i \bar{w}_i$ (equalities are in the free group) where $w_i$ and $\bar{w}_i$ are determined by Lemma 3.17 (see (3.7), (3.8)).

Therefore $\bar{w}_i$ is a letter from $A_0^{\pm 1}$ or 1, and the projection of $w_i$ onto $A_0$ is freely equal to $\bar{w}_i$. In particular, $|\bar{w}_i| \leq |w_i|$. Since there is no $R$ between $p$ and $p^{-1}$, none of the rules $r_{13}$ are $r_{21}$.

If $\theta_i = r_{13}$, then $u_i = \emptyset$, and $w_0 \ldots w_{i-1} = u_0^{-1}$. Therefore $|v_i| \leq 2|u_0| + |v_0| < 3 \max(|u_0|, |v_0|)$. Hence it is enough to assume that rules $r_{13}$ do not occur during the computation and prove that in that case, say, $|W_i| < 7 \max(|W_0|, |W_t|)$.

We can assume that all rules $\theta_i$ are in $\{r_1(a), r_2(a), r_{12}(a)^{\pm 1}\}$ (the case when all $\theta_i$ are of the form $r_3$ is similar but easier). Then the history of the computation is a subword of a word of the following form (we write $r_i, r_{12}$ instead of $r_i(a)$ and $r_{12}(a)$):

$$r_1 \ldots r_1 r_{12} r_2 \ldots r_2 r_{12}^{-1} r_1 \ldots$$

(3.12)

Since the history is reduced, by Lemma 3.26, $w_0 w_1 \ldots w_{i-1},$ is a freely reduced word for every $i$. In addition $|\bar{w}_0 \bar{w}_1 \ldots \bar{w}_i| \leq |w_0 w_1 \ldots w_i|$. Hence $3i + |u_0| + |v_0| + 3 \geq |W_i| > i - |u_0|$ for every $i$. Therefore $|W_i| \leq 3i + |W_0| \leq 3t + |W_0| < |W_0| + 3|W_i| + 3|u_0| < 4|W_0| + 3|W_t| \leq 7 \max(|W_0|, |W_t|)$. 

21
3. Let base($W$) = $pp^{-1}p$. Let $W = W_0 \to_{\theta_0} W_1 \to_{\theta_1} \ldots \to_{\theta_{t-1}} W_t$ be a computation, $W_i = q_1 u_i q_1^{-1} v_i q_i$, where $q_i \in \{p(1), p(2), p(3)\}$.

Let $W_j$ for some $j$ be a longest word in the computation. Then we can assume without loss of generality that $|W_0|$ is shorter than any of the words $W_1, \ldots, W_j$, and $W_t$ is shorter than any of the words among $W_j, \ldots, W_t$. So assume that $|W_j| < |W_j| > |W_0|$.

Since the base of $W$ does not contain $L$ between $p^{-1}$ and $p$ or $R$ between $p$ and $p^{-1}$, rules $r_{21}^{-1}, r_{13}^{-1}$ are excluded, so $\tau \in \{r_1(a), r_2(a), r_3(a), r_{12}(a)^{\pm 1}\}$. As in part 2, we shall assume that $r_3$ does not appear in the computation.

By Lemma 3.17, $u_{i+1} = \bar{w}_i u_i \bar{w}_i^{-1}$, $v_i + 1 = v_i^{-1} v_i w_i$ where $w_i = w(q_i)$ and \(\bar{w}_i = \bar{w}(q_i)\) are determined by formulas (3.7), (3.8).

As in part 2, the history $h = \theta_0 \ldots \theta_{t-1}$ of the computation is a subword of the word of the form (3.12), the product $w_0 w_1 w_2 \ldots w_{t-1}$ is a freely reduced word, and $|\bar{w}_j| \leq |w_j|$ for every $j$.

Therefore if $|v_{i+1}| > |v_i|$ for some $i$, then $|v_i| < |v_i| < |v_i|$ and $W_t$ cannot be shorter than $W_{t-1}$: indeed $|u_{t-1}| - |u_t| \leq 2$ by (3.8) from Lemma 3.17 and $|v_t| - |u_{t-1}| \geq 2$ since $v_t$ is a conjugate of $v_{t-1}$. Similarly if $|v_{i+1}| < |v_i|$, then $W_0$ cannot be shorter than $W_1$. Hence $|v_0| = |v_1| = \ldots = |v_t|$, i.e. all $v_i$ are cyclic shifts of $v$. This implies that the word $w_0 \ldots w_{t-1}$ is periodic with period $d \leq |v|$. By Lemma 3.26, the history $h$ of the computation is determined by its first letter $\theta_1$ and the word $w = w_1 w_2 \ldots$. Every letter in $w$ is contained in one of the words $w_i$, so it corresponds to one of $\theta_i$. Consider the letters in $w$ number $1, d+1, 2d+1, \ldots$. These letters are the same since the word $w$ is periodic with period $d$. Let $D$ be the total number of rules in $Z(A)$. Then among the first $D + 1$ rules corresponding to these letters, there are two equal rules. Since none of the words $w_i$ contain two same letters, we can deduce that the word $h$ is periodic with period $d_1 \leq (D + 1)d$.

Therefore, by Lemma 3.17, the sequence $\bar{w}_0, \ldots, \bar{w}_{t-1}$ is periodic with the same period $d_1$. Let $z = \bar{w}_{t-1} \ldots \bar{w}_0$. Then $u_j = z_j z^s u_0 z^{-s} z_j^{-1}$, $u_t = z_t z^{s'} u_0 z^{-s'} z_t^{-1}$ for some words $z_j, z_t$ of length at most $d_1$ and some $s, s'$. Now we can apply Lemma 3.28 and deduce that $|u_j| \leq 2(2d_1 + |u_0|) + |u_t| + 4d_1 \leq C_1(|u_0| + |u_0| + |u_t|)$ for some constant $C_1$. Therefore $|W_j| \leq C(|W_0| + |W_t|)$ for some constant $C$ as required.

4. We can assume that $W$ is a shortest word in the computation. Using notation similar to Case 3, $W_i = p^{-1} v_i p_i R$ and $v_i + 1 = v_i^{-1} v_i w_i$, $u_i + 1 = \bar{w}_i u_i$. As in Case 3, we can assume that $|v_0| = |v_1| = \ldots$ which implies, as before, that the history of the computation is periodic with period $d_1 \leq (D + 1)|v|$. The proof can be completed as in Case 3, using Lemma 3.28.

Case 5 follows from Lemma 3.25.

3.7 A composition of $S$ and the adding machine

Now let us define a “composition” $S \circ Z$ of $S$ and $Z(A)$. Essentially we insert a $p$-letter between any two consecutive $k$-letters in admissible words of $S$, and treat any subword $k_i \ldots p \ldots k_{i+1}$ as an admissible word for $Z(A)$.

First, for every $i = 0, \ldots, N - 1$, we make two copies of the alphabet $Y_i$ of $S$ ($i = 1, \ldots, N - 1$): $Y_{i, 0} = Y_i$ and $Y_{i, 1}$. The set of state letters of the new machine is

\[ K_1 \cup P_1 \cup K_2 \cup P_2 \cup \ldots \cup P_{N-1} \cup K_N \]

where $P_i = \{ p_i, p_i(\theta, 1), p_i(\theta, 2), p_i(\theta, 3) \mid \theta \in \Theta \}$, $i = 1, \ldots, N - 1$. We shall denote the components of this union by $Q_1, \ldots, Q_{2N-1}$.

The set of state letters is

\[ \hat{Y} = (Y_{1, 0} \cup Y_{1, 1}) \cup Y_{1, 0} \cup (Y_{2, 0} \cup Y_{2, 1}) \cup Y_{2, 0} \cup \ldots \cup (Y_{N-1, 0} \cup Y_{N-1, 1}) \cup Y_{N-1, 0}; \]
the components of this union will be denoted by \( \bar{Y}_{1}, ..., \bar{Y}_{2N-2} \).

The set of positive rules \( \bar{\Theta} \) of \( S \circ Z \) is a union of the set of suitably modified positive rules of \( S \) and \( 2(N - 1)|\Theta| \) copies \( \Z_{i}(\theta)^{+} \) \( (\theta \in \Theta, i = 1, ..., N) \) of positive rules of the machine \( Z(Y_{i}) \) (also suitably modified).

More precisely, every positive rule \( \theta \in \Theta^{+} \) of the form

\[
[k_{1}u_{1} \rightarrow k'_{1}u'_{1}, v_{1}k_{2}u_{2} \rightarrow v'_{1}k'_{2}u'_{2}, ..., v_{N-1}k_{N} \rightarrow v'_{N-1}k'_{N}]
\]

where \( k_{i}, k'_{i} \in K_{i}, u_{i} \) and \( v_{i} \) are words in \( Y \), is replaced by

\[
\bar{\theta} = [k_{1}u_{1} \rightarrow k'_{1}u'_{1}, v_{1}p_{1} \rightarrow v'_{1}p_{1}(\theta, 1), k_{2}u_{2} \rightarrow k'_{2}u'_{2}, ..., v_{N-1}p_{N-1} \rightarrow v'_{N-1}p_{N-1}(\theta, 1), k_{N} \rightarrow k'_{N}]
\]

with \( \bar{Y}_{2i-1}(\theta) = Y_{i,0}(\theta) \) and \( \bar{Y}_{2i} = \emptyset \) for every \( i \).

Thus each modified rule from \( \Theta \) turns on \( N - 1 \) copies of the machine \( Z(A) \) (for different \( A \)'s).

Each machine \( Z_{i}(\theta) \) is a copy of the machine \( Z(Y_{i}) \) where every rule \( \tau = [U_{1} \rightarrow V_{1}, U_{2} \rightarrow V_{2}, U_{3} \rightarrow V_{3}] \) is replaced by the rule of the form

\[
\bar{\tau}_{i}(\theta) = \\
\begin{bmatrix}
\bar{U}_{1} \rightarrow \bar{V}_{1}, \bar{U}_{2} \rightarrow \bar{V}_{2}, \bar{U}_{3} \rightarrow \bar{V}_{3}, \\
k'_{j} \rightarrow k'_{j}, p_{j}(\theta, 3) \rightarrow p_{j}(\theta, 3), j = 1, ..., i - 1, \\
p_{s}(\theta, 1) \rightarrow p_{s}(\theta, 1), k'_{s+1} \rightarrow k'_{s+1}, s = i + 1, ..., N - 1
\end{bmatrix}
\]

where \( \bar{U}_{1}, \bar{U}_{2}, \bar{U}_{3}, \bar{V}_{1}, \bar{V}_{2}, \bar{V}_{3} \) are obtained from \( U_{1}, U_{2}, U_{3}, V_{1}, V_{2}, V_{3} \), respectively, by replacing \( p(j) \) with \( p_{i}(\theta, j) \), \( L \) with \( k'_{j} \) and \( R \) with \( k'_{j+1} \), and for \( s \neq i, \bar{Y}_{2s-1}(\bar{\tau}_{i}(\theta)) = Y_{i,0} \).

In addition, we need the following transition rule \( \zeta(\theta) \) that returns all \( p \)-letters to their original form.

\[
[k'_{i} \rightarrow k_{i}, p_{j}(\theta, 3) \rightarrow p_{j}, i = 1, ..., N, j = 1, ..., N - 1].
\]

Thus while the machine \( Z_{i}(\theta) \) works all other machines \( Z_{j}(\cdot), j \neq i, \) must stay idle (their state letters do not change and do not move away from the corresponding \( k \)-letters). After the machine \( Z_{i}(\theta) \) finishes (i.e. the state letter \( p_{i}(\theta, 3) \) appears next to \( k_{i+1} \)), the next machine \( Z_{i+1}(\theta) \) starts working. After all \( p \)-letters have the form \( p_{j}(\theta, 3) \) we can apply \( \zeta(\theta) \) and turn all \( p_{j}(\tau, 3) \) into \( p_{j} \). Thus in order to simulate a computation of \( S \) consisting of a sequence of applications of rules \( \theta_{1}, \theta_{2}, ..., \theta_{s} \), we first apply all rules corresponding to \( \theta_{1} \), then all rules corresponding to \( \theta_{2} \), then all rules corresponding to \( \theta_{3} \), etc.

The modified rules \( \theta \) of \( S \) will be called \textit{basic rules}. We shall call basic rules, transition rules \( \zeta(\theta) \), and their inverses \textit{main rules}.

### 3.8 Properties of the \( S \)-machine \( S \circ Z \)

\textit{Notation}. For every word \( W \), we shall denote the number of \( a \)-letters in \( W \) by \( |W|_{a} \).

Every word in the alphabet \( \{Q_{1}, ..., Q_{2N-1}\} \) is called a \textit{base word}. Let \( B \) be a finite set of base words.

\textbf{Definition 3.30}. We call a base word \( B \)-\textit{covered} if

- it is covered by bases from \( B \) (i.e. every letter belongs to a subword from \( B \)),
it starts and ends with the same $q$-letter $x$.

We call a base word $w$ \textit{B-tight} if it has the form $uxvx$ where $xvx$ is a $B$-covered word, $w$ does not contain any other $B$-covered subwords. A base word is called \textit{B-narrow} if it does not contain $B$-covered subwords.

\textbf{Lemma 3.31.} There exists a finite set of base words $B$ such that

(*) the length of every $B$-narrow base is smaller than a constant $K_0$,

(**) for every admissible word $W$ with base from $B$ the width of every computation $W \rightarrow W_1 \rightarrow \ldots \rightarrow W_t$ does not exceed $C(|W| + |W_t| + \log_2 t)$ for some constant $C$.

\textbf{Proof.} Let $B$ be the set of bases of the form $(k_i q_i k_{i+1})^{\pm 1}, (k_i q_i^{-1} q_{i+1})^{\pm 1}, (k_i k_{i+1}^{-1}, k_i^{-1} k_i, (q_i q_i^{-1} q_i)^{\pm 1}$ ($q_i = \{p_i(\theta, 1), p_i(\theta, 2), p_i(\theta, 3), p_i\}$). Lemma 3.9 implies that condition (*) holds for $B$. Indeed, base words of the form $(q_i q_i^{-1} q_i)^{\pm 1}$ and all base words that start and end on the same $k$-letters and do not contain $(q_i q_i^{-1} q_i)^{\pm 1}$ as a subword are $B$-covered, and every word of length at least $6N$ contains one of these $B$-covered words.

Let $W$ be an admissible word with one of the bases from $B$. By passing to $W^{-1}$ if necessary, we can assume that base($W$) has one of the forms $k_i q_i k_{i+1}$, $q_i^{-1} q_{i+1}$, $k_i k_{i+1}^{-1}$, $k_i^{-1} k_i$, $q_i q_i^{-1} q_i$.

Let $f = f_1 f_2 f_3$ where $f_1, f_3$ contain no basic and transition rules or their inverses, $f_2$ starts and ends with a basic or transition rule or its inverse (one or more of these subwords may be empty).

\textbf{Case 1.} Suppose that $f = f_1$, so $f$ does not contain main rules.

Notice that we can represent $f$ as a product $z_1 z_2 \ldots z_k$ where each $z_i$ is a history of computation of one of the machines $Z_{j(i)}(\theta)$ for some $\theta$ and $j(i)$ depending on $i$, $j(i) \neq j(i + 1)$. Since $f^{\pm 1}$ does not contain basic and transition rules, all $\theta$’s are the same. By Lemma 3.25, 3.29, if base($W$) $\not\in \{k_i p_{k_{i+1}}, p_{-1} p_{k_{i+1}}\}$, then $s = 1$, the width of the computation does not exceed $C \max(|W|, |f \cdot W|)$ for some constant $C$ and no basic or transition rule can apply to $f \cdot W$.

If $W = k_i u p v k_{i+1}$ or $W = p^{-1} u p v k_{i+1}$, then each subcomputation corresponding to $z_j$ either does not affect the admissible word, or it is essentially a computation of $Z(Y_i)$ (with $L$ replaced by $k_i$, $R$ replaced by $k_{i+1}$, etc.). In the first case the width is equal to the length of $W$. In the second case, the width of the subcomputation corresponding to $z_j$ does not exceed $C$ times the sum of lengths of the beginning and ending words of the computation (by Lemma 3.29). Moreover in the second case the end word of the subcomputation, except possibly for $f \cdot W$, has the form $k_i w p' k_{i+1}$ (resp. $(p')^{-1} w p' k_{i+1}$) for some word $w$ in $Y_{i,0}^{\pm 1}$. By Lemma 3.18, the words $w$ and $u w$ (or $w$ and $v^{-1} u v$) must have the same projections onto $Y_i$. Therefore all $w$’s have the same length which does not exceed $\max(|W|, |f \cdot W|)$ (resp. $2 \max(|W|, |f \cdot W|)$). Therefore the width of the whole computation does not exceed $C(|W| + |f \cdot W|)$. Moreover we can conclude that if $f \cdot W$ contains a 2-letter subword $p R$, then $|f \cdot W| \leq |W|$.

\textbf{Case 2.} Suppose that $f_2$ is of length 1, i.e. it contains exactly one main rule. Then the base of $W$ is $k_i q_i k_{i+1}$ or $q_i^{-1} q_{i+1}$. We can repeat the argument from the last paragraph of Case 1, applied to the subcomputation corresponding to $f_1$ and $f_3^{-1}$. Note that a main rule $f_2^{\pm 1}$ must be applicable to $f_1 \cdot W$ and to $f_3^{-1} \cdot (f \cdot W)$, hence these words must contain subwords of the form $p k_{i+1}$ and all $a$-letters must belong to $Y_{i,0}$. Hence the lengths of the end words of these subcomputations cannot be bigger than $2|W|$ or $2|f \cdot W|$ by Lemma 3.18.

\textbf{Case 3.} Suppose that $f_2$ contains at least two main rules.

Again base($W$) must belong to $\{k_i q_i k_{i+1}, q_i^{-1} q_{i+1}\}$ for some $i$. By the argument in Case 1, $|f_1 \cdot W| \leq 2|W|$, $|f_1 f_2 \cdot W| \leq 2|f \cdot W|$, so we can assume that $f_1$ and $f_3$ are empty, thus $f$ starts
and ends with a main rule. We can assume that this computation has minimal length among all computations connecting $W$ and $f \cdot W$ and having the same width.

Let $\tau g \tau'$ be a subword of $f$ where $\tau$, $\tau'$ are main rules and $g$ does not contain main rules. Then either $g$ is empty or $g = g_1...g_m$ and each $g_j$ is a non-empty product of rules of one of the machines $Z_{s_j}(\theta_j)$, $s_j \neq s_{j+1}$.

Suppose that $g$ is not empty and $s_j = i$ for some $j$. In this case we shall call the subword $\tau g \tau'$ active. Let $f = f'g_jf''$ for some $f', f''$. Let $W' = f' \cdot W$, $W'' = g_j \cdot W'$. Then both $W'$ and $W''$ must contain subwords $pk_{i+1}$ where $p \in \{p_i(.,1),p_i(.,3)\}$ and all a-letters of $W', W''$ are in $Y_{i,0}$. Since $g_j$ is not empty, by Lemma 3.27, either $W'$ contains $p_i(.,1)$ and $W''$ contains $p_i(.,3)$, or $W'$ contains $p_i(.,3)$ and $W''$ contains $p_i(.,1)$. Since rule $r_{13}^{\pm 1}$ cannot be applicable to admissible words of $Z(A)$ of the form $p^{-1}pR$, we conclude that $\text{base}(W') = k_iq_ik_{i+1}$. By Lemma 3.25, the length of $g_j$ is at least $2|W'|-3$ (here $|W'|-3$ is the number of a-letters in $W'$), and the lengths of the words in the subcomputation $W' \rightarrow ... \rightarrow W''$ are the same.

If none of $s_j$ are equal to $i$, then we call the subword $\tau g \tau'$ passive. In that case all rules of $g$ fix admissible words with base $\text{base}(W)$. Therefore all words in the subcomputation corresponding to $g$ are the same. From the definition of main rules, then either $\tau' = \tau^{-1}$, or $\tau$ is the inverse of a basic rule and $\tau'$ is another basic rule, or $\tau$ is a transition rule and $\tau'$ is the inverse of another transition rule. If $\tau' \neq \tau^{-1}$, $g$ is non-empty, and the subcomputation corresponding to $g$ does not contain the longest word in the computation, we can use Lemma 3.12 and remove the subcomputation corresponding to $g$ and obtain a shorter computation of the same width connecting $W$ and $f \cdot W$. That would contradict the assumption that $f$ is the shortest possible. Hence we can assume that if $\tau g \tau'$ is passive and $g$ is not empty, and the subcomputation corresponding to $g$ does not contain the longest word in the computation, then $\tau' = \tau^{-1}$.

Notice also that we have proved, that in our computation, $|W_j| = |W_{j+1}|$ unless the $j+1$-st rule of the computation is a main rule.

Let $J = |f| \cdot W|$ be the width of the computation $W \rightarrow ... \rightarrow f \cdot W$. We can assume (by the remark made in the previous paragraph) that the $t$-th rule in $f$ is a main rule. For every admissible word $U$ with $\text{base}(U) = \text{base}(W)$ and every basic rule $\tau$ the difference $|U| - |\tau \cdot U|$ is bounded from above by a constant $c$. We can assume that $J - |W| > 4c$. Let $t_0$ be such that $J - |f|t_0| \cdot W| \geq 4c$ and $t - t_0$ is minimal possible. Then the suffix of $f|t|$ obtained by removing $f|t_0|$ has the form $\tilde{f} = \tau_1g_1\tau_2g_2...\tau_jg_j\tau_{j+1}$ such that

- $j \geq 3$,
- all $\tau_l$ are main rules, all $g_l$ do not contain main rules,
- all admissible words in the subcomputation corresponding to $\tilde{f}$ are of length between $J - 4c$ and $J$,
- one of the admissible words in that subcomputation has length at most $J - 4c$.

If one of the subwords $\tau_lg_l\tau_{l+1}$ in $\tilde{f}$ is active then the length of $f$ is at least $2^{J-4c-3}$. So $J \leq \log_2 |f| + 4c + 3$ and (***) holds.

Thus we can assume that all subwords $\tau_lg_l\tau_{l+1}$ are passive. But then, as we proved before, for every $l$ either $g_l$ is empty or $\tau_{l+1} = \tau_l^{-1}$. Moreover if $g_l$ is empty then $\tau_{l+1}$ cannot be the inverse of a basic rule or a transition rule. Hence $g_{l+1}$ cannot be empty. Therefore the lengths of admissible words in the subcomputation corresponding to $f$ can differ by at most $2c$, a contradiction with the fact that one of these words has length $J$ and the length of another one is at most $J - 4c$. \qed
From now on we shall fix the set $B$ of base words from Lemma 3.31, and a constant $K$ such that
\[ K > 2K_0. \quad (3.13) \]

4 Properties of the group $S \circ Z$

Let $\Theta_+$ be the set of positive rules of $S \circ Z$.

We redefine the map $\sigma$ and the stop word $\tilde{W}$ (see Section 3.5) as follows:
\[ \sigma'(v) = k_1(1)v p(1)k_2(1)p(2)k_3(1)...p(N - 1)k_N(1) \]
for every word $v$ over $X$,
\[ \tilde{W}'' = k_1(0)p(1)k_2(0)p(2)k_3(0)...p(N - 1)k_N(0). \]

Recall that $L$ is the language recognized by $S$.

**Lemma 4.1.** Let $W_1 = \sigma'(v)$ and $W_1 \to_{\theta_1} ... \to_{\theta_{i-1}} W_i$ be any computation of $S \circ Z$. Let $\theta_{i_1}, ..., \theta_{i_s}$ be all basic rules or their inverses in that computation. Let $\bar{\theta}_{i_j}$ be the rule of $S$ corresponding to $\theta_{i_j}$. For each $j$ let $W_j$ be the (natural) projection of $W_j$ on the alphabet of $q$- and $a$-letters of $S$. Then there exists a computation $\tilde{W}_1 \to ... \to \tilde{W}_t$ of the machine $S$, whose history is the reduced form of the word $\bar{\theta}_{i_1}...\bar{\theta}_{i_s}$.

**Proof.** This immediately follows from Lemmas 3.18, 3.12 and the definition of $S \circ Z$. \qed

**Lemma 4.2.** A word $v$ over $X$ belongs to $L$ if and only if $\sigma'(v)$ and $\tilde{W}'$ are conjugate in $S \circ Z$.

**Proof.** Suppose that $\sigma'(v)$ is a conjugate of $\tilde{W}'$. Then, by Lemma 3.13, there exists a computation $\sigma'(v) \to ... \to \tilde{W}'$ of $S \circ Z$. Then, by Lemma 4.1, there exists a computation $\bar{\sigma}'(v) \to ... \to \tilde{W}'$. Note that $\bar{\sigma}'(v) = \sigma(v)$, $\tilde{W}' = \tilde{W}$. Hence $v$ is accepted by $S$. Hence $v \in L$.

Conversely suppose that $v \in L$. Then, by Lemma 3.16, there exists a computation of $S$ connecting $\sigma(v)$ with $\tilde{W}$ and consisting of positive words. Then, by Lemma 3.22 (applied several times), there exists a computation of $S \circ Z$ connecting $\sigma'(v)$ and $\tilde{W}'$, so by Remark 3.14, $\sigma'(v)$ and $\tilde{W}'$ are conjugate in $S \circ Z$. \qed

Let $L$ be the maximal length of a defining relation of $S \circ Z$.

**Lemma 4.3.** Let $T$ be a $\theta$-band with base of length $b$. Let $l_a$ be the number of $a$-edges in the top path $\text{top}(T)$. Then the length of $T$ is between $l_a - (L - 1)b$ and $l_a + (L + 1)b$.

**Proof.** Every $a$-letter in the label of $\text{top}(T)$ labels an edge of a $(a, \theta)$-cell or on a $(q, \theta)$-cell. Every $(q, \theta)$-cell in $T$ has at most $L$ $a$-edges lying on $\text{top}(T)$, and every $(a, \theta)$-cell has at most one $a$-edge lying on $\text{top}(T)$. At most $L$ of these $a$-edges can belong to the same $(q, \theta)$-cell. Hence $|T| \geq l_a - Lb + b$.

The number of $(a, \theta)$-cells having no common $a$-edges with $\text{top}(T)$, does not exceed $Lb$ because at least one of the two $a$-edges of this cell is glued to an $a$-edge of a $(q, \theta)$-cell. This proves the inequality $|T| \leq l_a + Lb + b$. \qed

**Lemma 4.4.** Let $\Delta$ be a trapezium of height $h \geq 1$ whose base is $B$-tight and $B$-covered. Then the area of $\Delta$ does not exceed $Ch(|W'_a| + |W''_a| + log h + 1)$, where $W, W'$ are the labels of its top and bottom, respectively for some constant $C$. 

26
Proof. Notice that the number of letters in the base of $\Delta$ does not exceed $K_0 + 1$. Since $\text{base}(W)$ is covered by bases from $B$, $\Delta$ is covered by subtrapezia whose width, by Lemmas 3.31 and 4.3, does not exceed a constant times $|W|_a + |W'|_a + \log h + 1$, height does not exceed $h$, and the number of these subtrapezia does not exceed a constant (since $B$ is finite and from every cover of base($W$) by words from $B$, we can find a subcover which covers every letter of base($W$) at most a constant number of times). Hence the width of $\Delta$ does not exceed $C(|W|_a + |W'|_a + \log h + 1)$ for some constant $C$. Since, by Lemma 4.3, the area of $\Delta$ does not exceed a constant times the height times the width, the statement of the lemma follows.

4.1 A modified length function on $S \circ Z$

Let us modify the length function on the group $S \circ Z$. The standard length of a word (path) will be called its combinatorial length. From now on we use the word length for the modified length. As before let $L$ be the maximal length of a defining relation of $S \circ Z$. We set the length of every $q$-letter equal 1, and the length of every $a$-letter equal a small enough number $\delta$ so that

$$2 - 4L\delta - 4K\delta > \delta.$$  \hspace{1cm} (4.14)

We also set to 1 the length of every word of length $\leq L$ which contains exactly one $\theta$-letter and no $q$-letters (such words are called $(\theta, a)$-syllables). The length of a decomposition of an arbitrary word in a product of letters and $(\theta, a)$-syllables is the sum of the lengths of the factors. The length of a word $w$ is the smallest length of such decompositions. The length of a path in a diagram is the length of its label. The perimeter $|\partial \Delta|$ of a van Kampen diagram is similarly defined by cyclic decompositions of the boundary $\partial \Delta$. It follows from this definition that for any product $s = s_1s_2$ of two paths in a van Kampen diagram, we have

$$|s_1| + |s_2| \geq |s| > |s_1| + |s_2| - L\delta$$  \hspace{1cm} (4.15)

A maximal $\theta$-band of a van Kampen diagram $\Delta$ is called a rim band if its top or its bottom side lies on the contour $\partial \Delta$.

**Lemma 4.5.** Let $\Delta$ be a van Kampen diagram whose rim band $T$ has base with at most $K$ letters. Denote by $\Delta'$ the subdiagram $\Delta \setminus T$. Then $|\partial \Delta| - |\partial \Delta'| \geq \delta$.

**Proof.** Let $s$ be the top side of $T$ and $s \subset \partial \Delta$. Note that by our assumptions the difference between the number of $a$-edges in the bottom $s'$ of $T$ and the number of $a$-edges for $s$ cannot be greater than $LK$. Hence $|s'| - |s| \leq LK\delta$. However, $\Delta'$ is obtained by cutting off $T$ along $s'$, and its boundary contains two $\theta$-edges fewer than $\Delta$. Hence we have $|s_0| - |s'_0| \geq 2 - 2L\delta$ by (4.15), for the complements $s_0$ and $s'_0$ of $s$ and $s'$, respectively, in the boundaries $\partial \Delta$ and $\partial \Delta'$. Finally,

$$|\partial \Delta| - |\partial \Delta'| \geq 2 - 2L\delta - LK\delta - 2L\delta > \delta$$

by (4.14) and (4.15).  \hspace{1cm} $\square$

The definition of length has also the following obvious consequence.

**Lemma 4.6.** Let $s$ be a path in a diagram $\Delta$ having $c \theta$-edges and $d$ $a$-edges. Then

(a) $|s| \geq \max(c, c + (d - Lc)\delta)$;

(b) $|s| = c$ if $s$ is a top or a bottom of a $q$-band.
4.2 Combs

**Definition 4.7.** We say that a reduced diagram $\Gamma$ is a comb if it has a maximal $q$-band $Q$ (the handle of the comb), such that

$(C_1)$ $\text{bot}(Q)$ is a part of $\partial \Gamma$, and every maximal $\theta$-band of $\Gamma$ ends at a cell in $Q$.

If in addition the following properties hold:

$(C_2)$ one of the maximal $\theta$-bands $T$ in $\Gamma$ has a $B$-tight base and
$(C_3)$ other maximal $\theta$-bands in $\Gamma$ have $B$-tight or $B$-narrow bases

then the comb is called tight.

The number of cells in the handle $Q$ is the length of the comb, and the maximal length of the bases of the $\theta$-bands of a comb is called the basic width of the comb.

![Figure 1: Comb.](image)

Notice that every trapezium is a comb.

**Definition 4.8.** Let $Q$ be a maximal $q$-band of $\Delta$. Then $\text{bot}(Q)$ divides $\Delta$ into two parts. The part containing $Q$ is called the top part of $\Delta$ with respect to $Q$. The other part is called the bottom part of $\Delta$ with respect to $Q$.

**Lemma 4.9.** Let $\Delta$ be a reduced diagram with non-zero area. Assume that every rim band of $\Delta$ has base of length at least $K$. Then there exists a maximal $q$-band $Q$ in $\Delta$ such that the top part $\Gamma$ of $\Delta$ with respect to $Q$ is a tight comb.

**Proof.** Let $T_0$ be a rim band of $\Delta$. Its base $w$ is of length at least $K$, and therefore $w$ has disjoint prefix and suffix of lengths $K_0$ since $K > 2K_0$ by (3.13). The prefix of this base word must have its own $B$-tight subprefix $w_1$, by Lemma 3.31, part 1, and the definition of $B$-tight words. A $q$-edge of $T_0$ corresponding to the last $q$-letter of the $w_1$ is the start edge of a maximal $q$-band $Q'$ which bounds a subdiagram $\Gamma'$ containing a band $T$ (a subband of $T_0$) satisfying property $(C_2)$. It is useful to note that a minimal suffix $w_2$ of $w$, such that $w_2^{-1}$ is tight, allows us to construct another band $Q''$ and a subdiagram $\Gamma''$ which satisfies $(C_2)$ and has no cells in common with $\Gamma'$.

Thus, there are $Q$ and $\Gamma$ satisfying $(C_2)$. Let us choose such a pair with minimal area($\Gamma$). Assume that there is a $\theta$-band in $\Gamma$ which does not cross $Q$. Then there must exist a rim band $T_1$ which does not cross $Q$ in $\Gamma$. Hence one can apply the construction from the previous paragraph to $T_1$ and construct two bands $Q_1$ and $Q_2$ and two disjoint subdiagrams $\Gamma_1$ and $\Gamma_2$ satisfying the requirement $(C_2)$ for $\Gamma$. Since $\Gamma_1$ and $\Gamma_2$ are disjoint, one of them, say $\Gamma_1$, is inside $\Gamma$. But
the area of $\Gamma_1$ is smaller than the area of $\Gamma$, and we come to a contradiction. Hence $\Gamma$ is a comb and condition $(C_1)$ is satisfied.

Assume that the base of a maximal $\theta$-band $\mathcal{T}$ of $\Gamma$ is not narrow. Then it has a $\mathcal{B}$-tight proper prefix (we may assume that $\mathcal{T}$ terminates on $Q$), and again one obtain a $q$-band $Q'$ in $\Gamma$, which provides us with a smaller subdiagram $\Gamma'$ of $\Delta$, satisfying $(C_2)$, a contradiction. Hence $\Gamma$ satisfies property $(C_3)$ as well.

**Lemma 4.10.** Let $l$ and $b$ be the length and the basic width of a comb $\Gamma$ and let $\mathcal{T}_1, \ldots, \mathcal{T}_l$ be consecutive $\theta$-bands of $\Gamma$ (as in Figure 1). We can assume that the base of a maximal $\theta$-band $T$ of $\Gamma$ is not narrow. Then it has a $B$-tight proper prefix (we may assume that $T$ terminates on $Q$), and again one obtain a $q$-band $Q'$ in $\Gamma$, which provides us with a smaller subdiagram $\Gamma'$ of $\Delta$, satisfying $(C_2)$, a contradiction. Hence $\Gamma$ satisfies property $(C_3)$ as well.

**Proof.** For every $i$, $\text{top}(\mathcal{T}_i)$ and $\text{bot}(\mathcal{T}_{i+1})$ can have an initial segment in common. Let $\alpha_i$ and $\alpha'_i$ be the numbers of the $a$-edges in $\text{bot}(\mathcal{T}_i)$ and $\text{top}(\mathcal{T}_i)$ respectively that belong to the boundary of $\Gamma$.

Let $n_i$ be the length of $\mathcal{T}_i$. It follows from Lemma 4.3 that $|n_{i+1} - n_i| \leq 2Cb + \alpha'_i + \alpha_{i+1}$. Since $n_1 \leq Cb + \alpha_1$ by the same lemma, we have for any $i$,

$$n_i \leq (Cb + \alpha) + 2C(i - 1)b + \alpha'_1 + \alpha_2 + \cdots + \alpha'_{i-1} + \alpha_i \leq 2Clb + 2\alpha_i$$

This inequality provides us with a required upper bound for the area $\sum_{i=1}^l n_i$. Finally,

$$\alpha - \alpha_1 \geq \sum_{i=1}^{l-1} \alpha'_i + \sum_{i=2}^l \alpha_i + \alpha'_i \geq \sum_{i=1}^{l-1} |n_{i+1} - n_i| - 2Cb(l - 1) + (n_l - Cb)$$

$$\geq \sum_{i=1}^{l-1} (n_i - n_{i+1}) - 2Cb(l - 1) + (n_l - Cb) \geq n_1 - 3Cb \geq \alpha_1 - 4Clb.$$

$$\square$$

### 5 Dispersion of bipartite chord diagrams

**Definition 5.1.** Recall that a chord diagram is a system of chords in a disc such that the intersecting point of any two chords is in the interior of the disc. We can consider the intersection graph of a chord diagram whose vertices are chords and two chords are connected when they intersect. If that graph is bipartite, the chord diagram is called *bipartite*. We shall always consider bipartite chord diagrams (BCD for short) with fixed subdivision into two parts $\mathcal{T}$ and
Q, so chords from one of the parts do not intersect. The intersection points of T- and Q-chords are the nodes of the BCD.

For an integer $K \geq 1$, we fix a $K$-tuple $\alpha$ of numbers $0 < \alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_K = 1$. For every node $o$ on a T-chord $T$ and choice of the left-right direction on $T$, we assign the weight $\alpha_l$ where $l$ is the minimum of $K$ and the number of nodes on $T$ to the left of $o$ (including $o$). Thus the weight of a node $o$ is smaller if it is closer to the boundary (according to the chosen direction on the T-chord containing $o$). Note that we assign two weights to each node $o$ according to the two possible directions on a chord containing $o$.

Let $C$ be a Q-chord. Let $(o_1, o_2)$ be a pair of nodes from $C$. Fixing an orientation from $o_1$ to $o_2$ on $C$, we determine a left-to-right orientation on every T-chord crossing $C$. The node $o_i$ ($i = 1, 2$) lies on some T-chord $T_i$. Let $o'_i$ be the next node to the left of $o_i$ on $T_i$ or the intersection point of $T_i$ and $\partial D$ if there are no nodes on $T_i$ between $o_i$ and $\partial D$ to the left of $o_i$ (according to the chosen direction of $T_i$). If either both $o'_i$ lie on the same Q-chord or both lie on $\partial D$, then we call the pair $(o_1, o_2)$ good. Otherwise we call the pair $(o_1, o_2)$ bad. By definition, the weight of the pair $(o_1, o_2)$ is the product of their weights (corresponding to the direction from $o_1$ to $o_2$ on the Q-chord). We shall set the weight of every good pair to 0. Note that a pair $(o_1, o_2)$ may be good while the pair $(o_2, o_1)$ is bad.

![Figure 3: (o1, o2) is a bad pair, (o2, o1) is a good pair.](image)

The $\alpha$-dispersion $D_\alpha(C)$ of the system of T-chords on a Q-chord $C$ is the sum of weights of all bad pairs of nodes $(o_1, o_2)$ such that $o_1$ and $o_2$ lie on $C$. For example, let $1$ be the 1-tuple of numbers (1). Then $D_1(C)= 1$ for Figure 3.

**Definition 5.2.** The sum of dispersions $D_\alpha(C)$ over all Q-chords $C$ is called the $\alpha$-dispersion $D_\alpha(G)$ of the BCD $G$.

Clearly $D_1(G)$ is the number of bad pairs of $G$, and $D_\alpha(G) \leq D_1(G)$ for every $\alpha$.

We need a quadratic upper bound for the $\alpha$-dispersion $D$ of a BCD $G$ in terms of the number of T-chords in $G$. 

30
Lemma 5.3. Let \( r \) be the number of \( T \)-chords of a BCD \( G \) on a disc \( D \). Then the \( 1 \)-dispersion \( D_1(G) \) of the BCD \( G \) does not exceed \( r^2 - r \).

Proof. We will induct on the number of \( T \)-chords in \( G \). If the set \( T \) is empty, then \( D_1(G) = 0 \), and the statement is obviously true.

Thus, we may assume that there is a chord \( T \) in \( T \). We can assume that \( T \) is close to the boundary that is there are no \( T \)-chords in one of the half-discs obtained by cutting \( D \) along \( T \). We denote by \( C_1, \ldots, C_l \) all the \( Q \)-chords intersected by \( T \), where \( l \geq 0 \). We enumerate and orient the \( Q \)-chords so that \( C_2 \) is to the right of \( C_1 \), \( C_3 \) is to the right of \( C_2 \), etc., and there are no nodes on \( C_k \) above \( o_k = C_k \cap T \), for each \( k = 1, \ldots, l \).

Let \( X_k \) be the set of \( T \)-chords \( T' \) of \( G \) such that the pair \((T \cap C_k, o_k)\) is bad. Denote by \( L_k \) the number of chords in \( X_k \). Thus \( L_k \) is the number of bad pairs of the form \((o, o_k)\).

Note that a chord \( T' \) from \( X_{k+1} \) cannot intersect \( C_k \) because otherwise the pair \((T' \cap C_k, o_k)\) would be good. Hence the sets \( X_k \) are pairwise disjoint, and \( \sum L_k \leq r - 1 \) because \( T \) does not belong to any \( X_k \).

Similarly let \( R_k \) be the number of bad pairs of the form \((o_k, o)\). Then as above, the sum of all \( R_k \) does not exceed \( r - 1 \). Hence

\[
\sum L_k + \sum R_k \leq 2(r - 1) \quad (5.16)
\]

Consider the BCD \( G_0 \) obtained from \( G \) by deleting the \( T \)-chord \( T \). Notice that every bad pair \((o', o'')\) in \( G_0 \) is also bad in \( G \) and the difference \( D_1(G) - D_1(G_0) \) between the number of bad pairs for \( G \) and the number of bad pairs for \( G_0 \) is the number of bad pairs of the forms \((o, o_k)\) and \((o_k, o)\) which does not exceed \( 2(r - 1) \) by (5.16). By the inductive assumption, \( D_1(G_0) \) does not exceed \( (r - 1)^2 - (r - 1) \). Hence \( D_1(G) \) does not exceed

\[
(r - 1)^2 - (r - 1) + 2(r - 1) = r^2 - r.
\]
Remark 5.4. The estimate $r^2 - r$ in Lemma 5.3 is optimal: for every $r$ one can easily construct (using the proof of Lemma 5.3) a BCD with $r$ T-chords whose 1-dispersion is exactly $r^2 - r$.

Lemma 5.5. Let a BCD $G_0$ be obtained by a deleting (a) a T-chord or (b) a Q-chord $C$ of a BCD $G$. Then the $\alpha$-dispersion of $G_0$ does not exceed the $\alpha$-dispersion of $G$.

Proof. (a) If we remove a T-chord, then every bad pair in $G_0$ is a bad pair in $G$ having the same weight. This implies part (a).

(b) Suppose $C$ is a Q-chord. Then the weight of any pair $(o_1, o_2)$ in $G_0$ cannot exceed the weight of that pair in $G$. Indeed, if the neighbor $o'_i$ of $o_i$ in $G$ did not belong to $C$, then the weight of that neighbor is the same or smaller in $G_0$ (since that neighbor can only become closer to the boundary). If $o'_i$ belongs to $C$ and its weight in $G$ is $\alpha_k$, then the weight of the neighbor of $o_i$ in $G_0$ cannot be bigger than $\alpha_k$ (because the neighbor of $o_i$ in $G_0$ is closer to the boundary than the neighbor of $o_i$ in $G$).

Let $X_0$ be the set of bad pairs in $G_0$. Then $X_0$ is the union of the set $Y_1$ of all bad pairs of $G$ which do not belong to $C$ (the weight of that pair in $G_0$ does not exceed its weight in $G$, we noted that in the previous paragraph) and the set $Y_2$ of good pairs $(o_1, o_2)$ of $G$ whose neighbors $o'_1, o'_2$ belong to $C$ but the pair $(o'_1, o'_2)$ is bad in $G$ (the weight of $(o_1, o_2)$ in $G_0$ equals the weight of $(o'_1, o'_2)$ in $G$). Since $(o'_1, o'_2)$ is uniquely determined by $(o_1, o_2)$, the total weight of the pairs from $X_0 = Y_1 \cup Y_2$ in $G_0$ does not exceed the total weight of bad pairs from $C$ in $G$. Hence $D_\alpha(G_0) \leq D_\alpha(G)$.

Definition 5.6. We say that a Q-chord $C'$ is close to a Q-chord $C$ of a BCD $G$ if every chord crossing $C'$ also crosses $C$.

From now on let

$$\alpha = \left(\frac{1}{K}, \frac{2}{K}, ..., 1\right).$$

We shall write $D(G)$ instead of $D_\alpha(G)$ and call it simply dispersion of the grading $G$.

Lemma 5.7. Suppose that a Q-chord $C'$ is close to a Q-chord $C$ in a BCD $G$ of a disc $D$. Let $D_0$ be one of the two subdiscs of the disc $D$ divided by $C$ that contains $C'$. Suppose that the
number of nodes from $D_0$ on any $T$-chord in $G$ is at most $K$. Denote the numbers of nodes on $C$ and $C'$ by $l$ and $l'$, respectively. Let $G_0$ be the BCD obtained from $G$ by removing $C'$. Then

$$D(G_0) \leq D(G) - \frac{1}{K^2} l'(l - l'). \quad (5.17)$$

Proof. Let $V_1$ be the set of nodes on $C$ that belong to $T$-chords intersecting $C'$, $V_2$ be the set of other nodes on $C$. Then $|V_1| = l'$, $|V_2| = l - l'$, and the number of pairs $(o, o')$ of nodes on $C$ such that $C'$ is to the left of $C$ according to the direction $o \to o'$ and either $(o \in V_1$ and $o' \in V_2)$ or $(o \in V_2$ and $o' \in V_1)$, is $l'(l - l')$.

Consider one of such pairs, $(o, o')$. Let $o = o_0, o_1, o_2, \ldots$ be a sequence of nodes from $D_0$ on the $T$-chord $T$ passing through $o$. Similarly consider a sequence of nodes $o' = o'_0, o'_1, o'_2, \ldots$ from $D_0$ on a $T$-chord $T'$. Since one of the nodes $o$ or $o'$ is in $V_1$ and another one in $V_2$, there exists $i = 0, 1, \ldots$ such that $(o_i, o'_i)$ belongs to some $Q$-chord $C_1 \neq C'$ and is bad.

The weight in $G_0$ of the node from the pair $(o_i, o'_i)$ that belongs to a $T$-chord that does not cross $C'$, is the same as its weight in $G$ and is at least $\frac{1}{K}$. The weight of the other node from that pair decreases by at least $\frac{1}{K}$ when we pass from $G$ to $G_0$ since by the assumption the number of nodes on every $T$-chord in $D_0$ is at most $K$ and we removed $C'$ that was to the left of $C_i$. Hence the weight of the pair decreases by at least $\frac{1}{K^2}$. Since the number of such pairs $(o_i, o'_i)$ is $l'(l - l')$, the total weight of such pairs decreases by at least $\frac{1}{K^2} l'(l - l')$.

Now as in Lemma 5.5 let $X_0$ be the set of bad pairs in $G_0$. Then $X_0$ is the union of the set $Y_1$ of all bad pairs of $G$ which do not belong to $C'$ and the set $Y_2$ of good pairs $(o_1, o_2)$ of $G$ whose neighbors $o'_1, o'_2$ belong to $C'$ but the pair $(o'_1, o'_2)$ is bad in $G$ (the weight of $(o_1, o_2)$ in $G_0$ equals the weight of $(o'_1, o'_2)$ in $G$). As in the proof of Lemma 5.5 the weight of any pair of $Y_1$ in $G_0$ does not exceed its weight in $G$. Since all pairs $(o_i, o'_i)$ that we considered in the previous paragraph belong to $Y_1$, we get the inequality (5.17). \square

6 The upper bound of the Dehn function

Lemma 3.3 implies that we can associate a bipartite chord diagram $G$ to any reduced diagram $\Delta$, where $T$-chords are the medians of the maximal $\theta$-bands, and $Q$-chords are the medians of $q$-bands.

Remark 6.1. If $\Delta$ is not a topological disc then $\partial \Delta$ can be transformed into a circle by an arbitrary small deformation, and so the topological structure of $G$ is well-defined.

The dispersion of this BCD is called the dispersion $D = D(\Delta)$ of the diagram $\Delta$.

In the following lemma, we estimate the area of a van Kampen diagram over $S \subset Z$ in terms of its perimeter and dispersion. Namely we show that for some constant $M$ the area of any reduced diagram $\Delta$ of perimeter $n$ does not exceed $M n^2 \log n + MD(\Delta)$ for some constant $M$. (Then using the quadratic upper bound for $D(\Delta)$ we will deduce that the area is bounded by $M' n^2 \log n$ for some constant $M'$.) Roughly speaking, we are doing the following. We use induction on the perimeter of the diagram. First we remove rim $\theta$-bands (those with one side on the boundary of the diagram) with short bases. This operation decreases the perimeter and preserves the sign of $M n^2 \log n + MD(\Delta) - \text{area}(\Delta)$, so we can assume that the diagram does not have such bands. Then we use Lemma 4.9 and find a tight comb inside the diagram with a handle $C$. We also find a long enough $q$-band $C'$ that is close to $C$. We use a surgery which amounts to removing a part of the diagram between $C'$ and $C$ and then gluing the two remaining parts of $\Delta$ together. The main difficulty is to show that, as a result of this surgery, the
perimeter decreases and the area and the dispersion change in such a way that the expression
\( M n^2 \log n + MD(\Delta) - \text{area}(\Delta) \) does not change its sign. In the proof, we need to consider several
cases depending on the shape of the subdiagram between \( C' \) and \( C \). Note that neither \( M n^2 \log n \)
or \( MD(\Delta) \) nor \( \text{area}(\Delta) \) alone behave in the appropriate way as a result of the surgery, but the
expression \( M n^2 \log n + MD(\Delta) - \text{area}(\Delta) \) behaves as needed.

Let us take a big enough constant \( M \). Here “big enough” means that \( M \) satisfies the
inequalities used in the proof of Lemma 6.2 (i.e. (6.20),(6.21), (6.22), (6.31), (6.35), (6.37),
(6.38), (6.43)). Each of them has the form \( M > C \) for some constant \( C \) that does not depend on
\( M \) (but depends on the constants introduced earlier), and the number of inequalities is finite,
so the choice of \( M \) is possible.

**Lemma 6.2.** The area of a reduced diagram \( \Delta \) does not exceed \( M n^2 \log' n + MD(\Delta) \) where
\( n = |\partial \Delta|, \) and \( \log' n = \max(\log_2 n, 1) \).

**Proof.** Arguing by contradiction, we consider a counter-example \( \Delta \) with minimal perimeter \( n \).
Of course, its area is positive, and, by Lemma 3.3, we have at least 2 \( \theta \)-edges on the boundary
\( \partial \Delta \), and so \( n \geq 2 \).

**Step 1.** Assume that there are two \( Q \)-chords \( C \) and \( C' \), where \( C' \) is close to \( C \) in the BCD
\( G(\Delta) \), and suppose the number \( l' \) of the nodes on \( C' \) does not exceed a half of the number \( l \) of
nodes lying on \( C \). Let \( X' \) and \( X'' \) be the \( q \)-bands corresponding in \( \Delta \) to \( C \) and \( C' \), respectively.
Without loss of generality assume that the top subdiagram \( \Gamma \) with respect to \( X' \) contains both
\( X' \) and \( X'' \) and the top subdiagram \( \Gamma' < \Gamma \) with respect to \( X' \) does not contain \( X'' \). We suppose
in addition that \( \Gamma' \) is a comb with handle \( X'' \). Let us prove that then the basic width \( b \) of the
comb \( \Gamma' \) is not smaller than \( K \), and in particular the comb cannot be tight.

By contradiction suppose that \( b < K \).

We note that the bands \( X' \) and \( X'' \) contain \( l \) and \( l' \) cells, respectively. It follows from Lemma
4.10 and our assumptions, that the area of \( \Gamma' \) does not exceed \( C_1(l')^2 + 2al' \) for \( \alpha = |\Gamma'|a \) and
some constant \( C_1 \).

Let \( \Delta' \) be the diagram obtained by deleting the subdiagram \( \Gamma' \) from \( \Delta \). Since the boundary
of \( \Delta' \) has at least two \( q \)-edges fewer than \( \Delta \), we have \( |\partial \Delta'| \leq |\partial \Delta| - 2 \). Moreover, we have from
Lemma 4.6 and Lemma 3.3 that

\[ |\partial \Delta| - |\partial \Delta'| \geq \gamma = \max(2, \delta(\alpha - Ll')) \] (6.18)

because the top or the bottom of \( X' \) has at most \( Ll' \) \( a \)-edges.

The difference of the dispersions \( D(\Delta) - D(\Delta') \) is at least \( \frac{1}{K^2}l'(l - l') \) by Lemmas 5.7 and
5.5. Hence \( D(\Delta) - D(\Delta') \geq \frac{1}{K^2}l'(l - l')^2 \) as \( l' \leq l - l' \). This inequality and the inductive assumption
related to the area of \( \Delta' \), imply that the area of \( \Delta' \) is not greater than

\[ M(n - \gamma)^2 \log(n - \gamma) + MD(\Delta) - \frac{M}{K^2}(l')^2. \]

Adding the area of \( \Gamma' \), we see that the area of \( \Delta \) does not exceed

\[ Mn^2 \log' n + MD(\Delta) - M\gamma n \log' n - \frac{M}{K^2}(l')^2 + C_1(l')^2 + 2al'. \]

(Keep in mind that \( \gamma \leq n \).) This will contradict the choice of the counter-example \( \Delta \) when we prove that

\[ -M\gamma n \log' n - \frac{M}{K^2}(l')^2 + C_1(l')^2 + 2al' < 0 \] (6.19)
Consider two cases.

(a) Let \( \alpha \leq 2Ll' \). Then inequality (6.19) follows from the inequality

\[
M \geq 2K^2(C_1 + 4L).
\]

(b) Assume that \( \alpha > 2Ll' \). Then by (6.18) we have \( \gamma \geq \frac{1}{2} \delta \alpha \) and \( M \gamma n \log' n > 2\alpha l' \) since \( n \geq 2l' \) by Lemma 3.3, and

\[
M > 2\delta^{-1}.
\]

Since \( \frac{M}{K^2} l'^2 > C_1 l'^2 \) by (6.20), the inequality (6.19) follows.

**Step 2.** Assume that \( \Delta \) has a rim \( \theta \)-band \( T \) whose base has \( s \leq K \) letters and \( \text{top}(T) \) is in \( \partial(\Delta) \). By deleting \( T \), we obtain, by Lemma 4.5, a diagram \( \Delta' \) with \( |\partial\Delta'| \leq n - \delta \). Since \( \text{top}(T) \) lies on \( \partial \Delta \), we have from the definition of the length (Section 4.1), that the number of \( a \)-edges in \( \text{top}(T) \) is less than \( \delta^{-1}(n - Ls) \). By Lemma 4.3, the length of \( T \) is at most \( (L + 1)s + \delta^{-1}(n - Ls) < \delta^{-1}n \) since \( \delta^{-1} > \frac{L + 1}{L} \) by (4.14). Thus, by applying the inductive hypothesis to \( \Delta' \), we have that area of \( \Delta \) is not greater than \( M(n - \delta)^2 \log(n - \delta) + MD(\Delta) + \delta^{-1} n \) because \( D(\Delta') \leq D(\Delta) \) by Lemma 5.5 (a). But this sum does not exceed \( Mn^2 \log' n + MD(\Delta) \) provided

\[
M \geq \delta^{-2}.
\]

This contradicts the choice of \( \Delta \). Hence the base of every rim \( \theta \)-band of \( \Delta \) has more than \( K \) letters.

**Step 3.** Now we can apply Lemma 4.9. By that lemma, there exists a tight comb \( \Gamma < \Delta \). Let \( \mathcal{T} \) be a \( \theta \)-band of \( \Gamma \) with \( \mathcal{B} \)-tight base. In particular, the basic width of \( \Gamma \) is smaller than \( K \). Since the base of \( \mathcal{T} \) is \( \mathcal{B} \)-tight, it is equal to \( u xv \) for some \( x = Q_i \) (we read the base starting at the boundary of \( \Delta \)) where \( \text{base}(xvx) \) is \( \mathcal{B} \)-covered.

The second occurrence of \( x \) in \( u xv \) corresponds to the last cell of \( \mathcal{T} \) belonging to the \( x \)-band \( Q \). Let \( Q' \) be the maximal \( x \)-band of \( \Gamma \) crossing \( \mathcal{T} \) at the cell corresponding to the first occurrence of \( x \) in \( u xv \).

We consider the smallest subdiagram \( \Gamma' \) of \( \Delta \) containing all the \( \theta \)-bands of \( \Gamma \) crossing the \( x \)-band \( Q' \). It is a comb with handle \( Q_2 \subset Q \). The comb \( \Gamma' \) is covered by a trapezium \( \Gamma_2 \) placed between \( Q' \) and \( Q \), and a comb \( \Gamma_1 \) with handle \( Q' \). The \( Q_i \)-band \( Q' \) belongs to both \( \Gamma_1 \) and \( \Gamma_2 \). The remaining part of \( \Gamma \) is a disjoint union of two combs \( \Gamma_3 \) and \( \Gamma_4 \) whose handles \( Q_3 \) and \( Q_4 \) contain the cells of \( Q \) that do not belong to the trapezium \( \Gamma_2 \). The handle of \( \Gamma \) is the composition of handles \( Q_3, Q_2, Q_4 \) of \( \Gamma_3, \Gamma_1, \Gamma_4 \) in that order.

Let the lengths of \( Q_3 \) and \( Q_4 \) be \( l_3 \) and \( l_4 \), respectively. Let \( l' \) be the length of the handle of \( \Gamma' \).

Then

\[
l = l' + l_3 + l_4,
\]

and, as we proved in Step 1, \( l' > l/2 \).

For \( i \in \{3, 4\} \) and \( \alpha_i = |\partial \Gamma_i|_a \), Lemma 4.10 gives inequalities

\[
A_i \leq C_{11}^2 + 2\alpha_i l_i
\]

for some constant \( C_1 \) where \( A_i \) is the area of \( \Gamma_i \). Let \( p_3, p_4 \) be the top and the bottom of the trapezium \( \Gamma_2 \). Here \( p_3 \) (resp. \( p_4 \)) share some initial edges with \( \partial \Gamma_3 \) (with \( \partial \Gamma_4 \)), the rest of these
Figure 6: Step 3 in Lemma 6.2.

paths belong to the boundary of $\Delta$. We denote by $d_3$ the number of $a$-edges of $p_3$ and by $d'_3$ the number of its edges which do not belong to $\Gamma_3$. Similarly, we introduce $d_4$ and $d'_4$. Let $A_2$ be the area of $\Gamma_2$. Then, by Lemma 4.4, $A_2 \leq C_2 l' (d_3 + d_4 + \log l' + 1)$ (6.24) for some constant $C_2$.

Now we note that the handle $Q_2$ of $\Gamma'$ is a copy of $Q$ because both maximal $q$-bands of the trapezium $\Gamma_2$ correspond to the same basic letter $x$. This makes the following surgery possible.

The diagram $\Delta$ is covered by two subdiagrams: $\Gamma$ and another subdiagram $\Delta_1$, having only the band $Q$ in common. We construct a new auxiliary diagram by attaching $\Gamma_1$ to $\Delta_1$ with identification of the band $Q'$ of $\Gamma_1$ and the band $Q_2$. We denote the constructed diagram by $\Delta_0$. It is a reduced diagram because every pair of its cells having a common edge, has a copy either in $\Gamma_1$ or in $\Delta_1$. It follows from our constructions that the area of $\Delta$ does not exceed $A_2 + A_3 + A_4 + A_0$, where $A_0$ is the area of $\Delta_0$.

Let $p_3$ be the segment of the boundary $\partial \Gamma_3$ that joins $Q$ and $\Gamma_2$ along the boundary of $\Delta$. It follows from the definition of $d_3$, $d'_3$, $l_3$ and $\alpha_3$, that the number of $a$-edges lying on $p_3$ is at least $\alpha_3 - (d_3 - d'_3) - 2Ll_3$.

Let $u_3$ be the part of $\partial \Delta$ that contains $p_3$ and connects $Q$ with $Q'$. It has $l_3$ $\theta$-edges. Hence we have, by Lemma 4.6, that the length $|u_3|$ of $u_3$ is at least

$$\max(l_3, l_3 + \delta(|p_3| - Ll_3)) \geq \max(l_3, l_3 + \delta(\alpha_3 - (d_3 - d'_3) - 2Ll_3)).$$

Since $u_3$ includes a subpath of length $d'_3$ having no $\theta$-edges, we also have by inequality (4.15) that $|u_3| \geq l_3 + \delta (d'_3 - L)$.

One can similarly define $p_4$ and $u_4$ for $\Gamma_4$. When passing from $\partial \Delta$ to $\partial \Delta_0$ we replace the end edges of $Q'$, $u_3$ and $u_4$ by two subpaths of $\partial Q$ having lengths $l_3$ and $l_4$. Let $n_0 = |\partial \Delta|$. Then it follows from the previous paragraph that

$$n - n_0 \geq 2 + \delta(\max(0, d'_3 - L, \alpha_3 - (d_3 - d'_3) - 2Ll_3) + \max(0, d'_4 - L, \alpha_4 - (d_4 - d'_4) - 2Ll_4))$$ (6.25)

In particular, $n_0 \leq n - 2$. By the inductive hypothesis,

$$A_0 \leq M n_0^2 \log' n_0 + MD(\Delta_0)$$ (6.26)
We note that the dispersion $\mathcal{D}(\Delta_0)$ of $\Delta_0$ is not greater than $\mathcal{D}(\Delta) - \frac{1}{K^2}l'(l - l')$ by Lemmas 5.7 and 5.5(b).

Therefore, by inequality (6.26), the area of $\Delta$ is not greater than

$$Mn^2 \log' n + M\mathcal{D}(\Delta) - Mn(n-n_0) \log' n - \frac{M}{K^2}l'(l - l') + A_2 + A_3 + A_4$$  
(6.27)

In view of inequalities (6.24) and (6.23), to obtain the desired contradiction, we should prove that

$$Mn(n-n_0) \log' n + \frac{M}{K^2}l'(l - l') \geq C_3(l_3 + l_4 + \log' l' + 1) + C_3(l_3^2 + l_4^2) + 2\alpha_3 l_3 + 2\alpha_4 l_4$$  
(6.28)

where $C_3 = \max(C_1, C_2)$ is a constant that does not depend on $M$. Note that we can assume that

$$C_3 \gg L.$$  
(6.29)

First we can choose $M$ big enough so that $\frac{M}{3K^2}l'(l - l') \geq C_3(l_3 + l_4)^2 \geq C_3(l_3^2 + l_4^2)$. Indeed

$$l - l' = l_3 + l_4 < l'$$  
(6.30)

since $l' > l/2$, and $\frac{M}{3K^2}l'(l - l') \geq \frac{M}{3K^2}(l_3 + l_4)(l_3 + l_4)$, so it is enough to assume that

$$M > 3K^2C_3.$$  
(6.31)

We also have that

$$\frac{M}{2}n(n-n_0) \log' n \geq C_3l'(\log' l' + 1)$$  
(6.32)

because $n - n_0 \geq 2$, $n \geq 2l'$ and $M \geq 2C_3$ by (6.31).

It remains to prove that

$$\frac{M}{2}n(n-n_0) \log' n + \frac{2M}{3K^2}l'(l - l') > C_3l'(d_3 + d_4) + 2\alpha_3 l_3 + 2\alpha_4 l_4.$$  
(6.33)

We assume without loss of generality that $\alpha_3 \geq \alpha_4$, and consider two cases.

(a) Suppose $\alpha_3 \leq 2C_3(l - l')$.

Since $d_i \leq \alpha_i + d_i'$ for $i = 3, 4$, we also, by inequality (6.25), have

$$d_3 + d_4 \leq \alpha_3 + \alpha_4 + d_3' + d_4' < 4C_3(l - l') + \delta^{-1}(n - n_0) + 2L - 2\delta^{-1} < 4C_3(l - l') + \delta^{-1}(n - n_0).$$

since $\delta^{-1} > L$ by (4.14).

Therefore

$$\frac{M}{3K^2}l'(l - l') + \frac{M}{2}n(n-n_0) \log' n \geq 4C_3^2l'(l - l') + C_3\delta^{-1}(n - n_0)l' \geq C_3l'(d_3 + d_4)$$  
(6.34)

since we can assume that

$$M > 12K^2C_3^2, \quad M/2 > C_3\delta^{-1}.$$  
(6.35)

We have also by (6.30):

$$\frac{M}{3K^2}l'(l - l') \geq \frac{M}{3K^2}(l_3 + l_4)(l_3 + l_4) \geq \frac{M}{3} \frac{\alpha_3 + \alpha_4}{4C_3}(l_3 + l_4) > 2\alpha_3 l_3 + 2\alpha_4 l_4$$  
(6.36)
since we can assume that
\[ M > 24K^2C_3. \]  
(6.37)

The sum of inequalities (6.34) and (6.36) gives us the desired inequality (6.33).

(b) Assume now that \( \alpha_3 > 2C_3(l - l') \). Then, applying Lemma 4.10 to the comb \( \Gamma_3 \), we obtain
\[ d_3 - d_3' < \frac{1}{7} \alpha_3 + 2CKl_3 \leq \frac{5}{6} \alpha_3 \]  
since \( l_3 < \frac{\alpha_3}{2C_3} \) and
\[ 2C_3 > 12CK. \]  
(6.38)

We also have \( d_4 - d_4' < \frac{1}{7} \alpha_4 + 2CKl_4 \leq \frac{5}{6} \alpha_3 \). These two inequalities and inequality 6.25 lead to
\[ d_3 + d_4 < \frac{5}{3} \alpha_3 + \delta^{-1}(n - n_0) \]  
(6.39)

In addition,
\[ \alpha_3 - (d_3 - d_3') - 2Ll_3 \geq \frac{1}{6} \alpha_3 - \frac{2L}{2C_3} \alpha_3 \geq \frac{1}{7} \alpha_3, \]
since \( l_3 < \frac{\alpha_3}{2C_3} \) and \( C_3 > 42L \) by (6.29). Therefore, by 6.25,
\[ n - n_0 \geq \frac{1}{7} \delta \alpha_3. \]  
(6.40)

Thus by (6.39)
\[ d_3 + d_4 < 13 \delta^{-1}(n - n_0). \]  
(6.41)

Since \( 2l' < n \) and \( n - n_0 \geq 2 \), inequality (6.41) implies
\[ \frac{M}{3} n(n - n_0) \log' n > C_3'l'(d_3 + d_4) \]  
(6.42)
because we can assume that
\[ M >> C_3 \delta^{-1} \]  
(6.43)
\((M > 21C_3 \delta^{-1} \) is enough).

Inequalities (6.40), (6.43), \( \alpha_3 \geq \alpha_4 \), and \( 4(l_3 + l_4) \leq n \) give us
\[ \frac{M}{6} n(n - n_0) \log n' \geq \frac{7}{2} C_3 \delta^{-1}(n - n_0) n \geq 2\alpha_3(l_3 + l_4) \geq 2\alpha_3l_3 + 2\alpha_4l_4 \]  
(6.44)

The inequality (6.33) follows now from inequalities (6.42), and (6.44).

**Lemma 6.3.** Let \( n \) be the combinatorial length of a reduced diagram \( \Delta \). Then the area of \( \Delta \) is \( O(n^2 \log n) \).

**Proof.** By Lemmas 3.3 and 5.3, we have
\[ D(\Delta) \leq (n/2)^2. \]  
(6.45)

Let \( n' = |\Delta| \). It follows from the definition of the length that \( n' \leq n \). By this inequality, inequality (6.45) and Lemma 6.2, we have the inequality \( \text{area}(\Delta) \leq Mn^2 \log' n + \frac{M}{4}n^2 \). The lemma is proved.
7 The end of the proof

Lemma 7.1. The Dehn function of \( S \circ Z \) is, up to equivalence, at least \( n^2 \log n \).

Proof. Let us use the fact that for some fixed word \( W \), \( S \) had the computation (3.6) for every \( n \). The length of that computation is \( 2n \) and the width is \( n + N \).

Consider the corresponding computation of \( S \circ Z \) (see the proof of Lemma 4.2). Its width is at most \( n + 2N \). Let \( l(n) \) be the length of that computation and let \( a(n) \) be its area. Then by Remark 3.19 and the description of the computation of \( S \circ Z \) (see Section 3.7),

\[
2 \sum_{m=1}^{n-1} (2^m + 2N - 3) + (2^n + 2N - 3) + 2N - 1 < l(n) < 2 \sum_{m=1}^{n-1} (6 \cdot 2^m + 2N - 3) + (6 \cdot 2^n + 2N - 3) + 2N - 1.
\]

So \( 3 \cdot 2^n + C_1 \leq l(n) < 18 \cdot 2^n + C_2 \) for some constants \( C_1, C_2 \). In addition \( a(n) > C' l(n) \log_2 l(n) \) for some constant \( C' \).

The sides of the corresponding trapezium \( \Delta \) have the same labels. Thus we can take \( l(n) \) copies of \( \Delta \) and glue them side by side to obtain a new trapezium \( \bar{\Delta} \). The perimeter \( d \) of \( \bar{\Delta} \) is \( O(l(n)) \) and the area is \( O(l(n)^2 \log l(n)) \). By Lemmas 3.3 and 3.11, there is only one reduced van Kampen diagram with the same boundary label as \( \bar{\Delta} \).

Since \( l(n) \) is between \( 3 \cdot 2^n + C_1 \) and \( 18 \cdot 2^n + C_2 \), for every sufficiently large number \( d \) there exists a number of the form \( l(n) \) between \( d \) and \( 15d \). Indeed it is enough to find a natural number \( n \) satisfying \( 2^n > (d - C_1)/3 \) and \( 2^n < (15d - C_2)/18 \). Such a number \( n \) exists for all sufficiently large \( d \) since \( 15/18 > 2/3 \).

Hence \( n^2 \log n \) is equivalent to a lower bound for the Dehn function of \( S \circ Z \).

Lemmas 7.1 and 6.2 show that the Dehn function of the group \( S \circ Z \) is equivalent to \( n^2 \log n \).

Finally the undecidability part of Theorem 1.1 follows from Lemmas 3.15 and 3.16. Indeed, by these lemmas if \( L \) is a non-recursive language then \( S \circ Z \) has undecidable conjugacy problem.

Remark 7.2. An easy example of a group with Dehn function \( n^2 \log n \) constructed using the method described in this paper is the group \( S_0 \circ Z \) where \( S_0 \) has only one \( q \)-letter \( \{k\} \), one \( a \)-letter \( a \), and two rules \( \tau_1, \tau_2 \), both equalling \([k \to ak] \) (and their inverses). One can easily write down an explicit presentation of this group. It is an HNN extension of a free group of rank 10 with 14 free letters.

Remark 7.3. The method used in this paper allows one to construct groups with other small Dehn functions. In fact, a similar argument to the one used in the proof of Lemma 6.2 proves the following statement.

Let \( f(n) > 1 \) be any function. Let \( S \) be an \( S \)-machine satisfying the condition of Lemma 3.31 with \( \log_2 h \) replaced by \( f(h) \). Then the Dehn function of the group \( S \) does not exceed \( n^2 f(n) \).

Using the technique from [11] of converging Turing machines into \( S \)-machines, and Remark 7.3, one can construct \( S \)-machines with many other Dehn functions between \( n^2 \) and \( n^3 \) (the fact that the Dehn function of any \( S \)-machine does not exceed \( n^3 \) has been proved in [11]).
References

[1] J. C. Birget, A.Yu. Ol’shanskii, E.Rips, M. V. Sapir. Isoperimetric functions of groups and computational complexity of the word problem. Annals of Mathematics, 156, 2 (2002), 467-518

[2] M. R. Bridson. Fractional isoperimetric inequalities and subgroup distortion. J. Amer. Math. Soc. 12 (1999), no. 4, 1103–1118.

[3] N. Brady, M.R. Bridson. There is only one gap in the isoperimetric spectrum. Geom. Funct. Anal. 10 (2000), no. 5, 1053–1070.

[4] M. Gromov. Hyperbolic groups, in Essays in Group Theory (S. Gersten, ed.), MSRI Publ. 8, Springer-Verlag (1987), 75-263.

[5] Charles F. Miller III, Paul E. Schupp. The geometry of Higman-Neumann-Neumann extensions. Collection of articles dedicated to Wilhelm Magnus. Comm. Pure Appl. Math. 26 (1973), 787–802.

[6] A. Yu. Olshanskii. Hyperbolicity of groups with subquadratic isoperimetric inequality, Internat. J. Algebra Comput. 1 (1991), no. 3, 281–289.

[7] A.Yu. Olshanskii and M.V. Sapir. Non-amenable finitely presented torsion-by-cyclic groups, Publ. Math. Inst. Hautes Études Sci. No. 96 (2002), 43–169 (2003).

[8] A.Yu. Olshanskii, M.V. Sapir. The conjugacy problem and Higman embeddings. Mem. Amer. Math. Soc. 170 (2004), no. 804, viii+133 pp

[9] Michael Polyak. Invariants of curves and fronts via Gauss diagrams. Topology 37 (1998), no. 5, 989–1009.

[10] Michael Polyak, Oleg Viro. Gauss diagram formulas for Vassiliev invariants. Internat. Math. Res. Notices 1994, no. 11, 445ff., approx. 8 pp.

[11] M. V. Sapir, J. C. Birget, E. Rips. Isoperimetric and isodiametric functions of groups, Annals of Mathematics, 157, 2 (2002), 345-466.

Alexander Yu. Ol’shanskii
Department of Mathematics
Vanderbilt University
alexander.olshanskiy@vanderbilt.edu
http://www.math.vanderbilt.edu/~olsh
and
Department of Higher Algebra, MEHMAT
Moscow State University
olshan@shabol.math.msu.su

Mark V. Sapir
Department of Mathematics
Vanderbilt University
m.sapir@vanderbilt.edu
http://www.math.vanderbilt.edu/~msapir

40