TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 351, Number 7, Pages 2961–2978
S 0002-9947(99)02010-3
Article electronically published on March 1, 1999

EQUATIONS IN THE Q–COMPLETION OF A TORSION–FREE HYPERBOLIC GROUP

O. KHARLAMPOVICH, E. LIOUTIKOVA, AND A. MYASNIKOV

Abstract. In this paper we prove the algorithmic solvability of finite systems of equations over the **Q**-completion of a torsion-free hyperbolic group.

It was recently proved in [1] that finite systems of equations over the **Q**–completion of a finitely generated free group are algorithmically solvable. In this paper we generalize the results of [1] to the case of the **Q**–completion $G^{\mathbf{Q}}$ of an arbitrary torsion–free hyperbolic group G with generators $d_1, \dots, d_N$. (A detailed definition of the **Q**–completion of a group can be found in [1].)

Main Theorem. *Let G be a torsion–free hyperbolic group. Then there exists an algorithm that decides if a given finite system of equations over the* **Q**–*completion of G has a solution, and if it does, finds a solution.*

By a triangular equation we mean an equation with at most three terms. Given a system S of equations in $G^{\mathbf{Q}}$, we may assume that all equations in S are triangular. Indeed, an arbitrary equation $x_1^{\varepsilon_1} x_2^{\varepsilon_2} \dots x_n^{\varepsilon_n} = 1$ (where x_i stands either for a variable or for a constant, $\varepsilon_i = \pm 1$) is equivalent to a finite system of triangular equations: $x_1^{\varepsilon_1} x_2^{\varepsilon_2} y_1^{-1} = 1$, $y_1 x_3^{\varepsilon_3} y_2^{-1} = 1$, $\dots$, $y_{n-2} x_{n-1}^{\varepsilon_{n-1}} x_n^{\varepsilon_n} = 1$. Adding a finite number of new variables and equations, we may also assume that S contains only equations with coefficients in G. To achieve this, we replace every constant of the form $d^{\frac{m}{n}}$ by a new variable z satisfying the equation $z^m = d^n$.

From now on, we fix a finite system S of triangular equations over $G^{\mathbf{Q}}$ with coefficients in G. We will reduce the system S to a finite set of systems in a specific hyperbolic group $G * \langle t_1, \dots , t_{\psi(m)} \rangle$, where the number $\psi(m)$ can be determined effectively given the number m of equations in the original system S. The resulting systems are accompanied by the restriction that some of the variables belong to a certain subgroup $G * \langle t_1, \dots , t_i \rangle$ of $G * \langle t_1, \dots , t_{\psi(m)} \rangle$, i.e. do not contain certain t's. To each of the systems we can apply the (slightly modified) method of Rips and Sela [4] to see if they are decidable. If none of them is consistent, then the original system S has no solution; if at least one has a solution, then it is possible to find a corresponding solution to the system S.

Suppose that the system has a solution in $G^{\mathbf{Q}}$; let $\{X_1, \dots , X_L\}$ be a solution with the minimal possible number of roots. Since the solution belongs to $G^{\mathbf{Q}}$, it is contained in a certain group K, obtained from G by adding finitely many roots. It is the union of a chain of subgroups H_i defined as follows. Let $G = H_0$.

Received by the editors August 14, 1996.
1991 *Mathematics Subject Classification.* Primary 20E05, 20F10.
The first author was supported by grants from NSERC and FCAR.
The third author was supported by the NSF grant DMS-9103098.

Step 1. Consider pairwise nonconjugated cyclically minimal primitive elements $u_1, \dots, u_{k_1} \in G$, $|u_1| \le \dots \le |u_{k_1}|$ (here $|u|$ denote the length of u in G), and add roots $t_1, \dots, t_{k_1}$, such that $u_j = t_j^{s_j}$. (Notice that u_{i+1} does not become a proper power after we add roots $t_1, \dots, t_i$.) The corresponding groups are denoted by $H_1, \dots, H_{k_1}$, where $H_{j+1} = H_j *_{u_{j+1}=t_{j+1}^{s_{j+1}}} \langle t_{j+1} \rangle$.

Step 2. Consider pairwise nonconjugated primitive elements $u_{k_1+1}, \dots, u_{k_2} \in H_1$, cyclically reduced in the amalgamated product, each having the reduced form $u = t_1^{\alpha_1} c_1 \dots t_1^{\alpha_k} c_k$, where $\alpha_i \ne 0, \alpha_i \in \mathbf{Z}, c_i \in G$, $|u_{k_1+1}|_{H_1} \le \dots \le |u_{k_2}|_{H_1}$; and add roots $t_{k_1+1}, \dots, t_{k_2}$, such that $u_j = t_j^{s_j}$, to the group H_{k_1}. The corresponding groups are denoted by $H_{k_1+1}, \dots, H_{k_2}$, where $H_{j+1} = H_j *_{u_{j+1}=t_{j+1}^{s_{j+1}}} \langle t_{j+1} \rangle$.

Step i+1. Suppose that $H_1, \dots, H_{k_i}$ have been constructed.

Consider pairwise nonconjugated primitive elements $u_{k_i+1}, \dots, u_{k_{i+1}} \in H_i$, cyclically reduced in the amalgamated product, each having the reduced form $t_i^{\alpha_1} c_1 \dots t_i^{\alpha_k} c_k$, where $\alpha_i \ne 0, \alpha_i \in \mathbf{Z}, c_i \in H_{i-1}$ (c_k is not a power of u_i, because the elements are cyclically reduced), $|u_{k_i+1}|_{H_i} \le \dots \le |u_{k_{i+1}}|_{H_i}$, and add roots $t_{k_i+1}, \dots, t_{k_{i+1}}$, such that $u_j = t_j^{s_j}$, to the group H_{k_i}. The corresponding groups are denoted by $H_{k_i+1}, \dots, H_{k_{i+1}}$, where $H_{j+1} = H_j *_{u_{j+1}=t_{j+1}^{s_{j+1}}} \langle t_{j+1} \rangle$.

Finally, for some number i one has $K = H_{k_{i+1}}$.

The group $H_{k_{i+1}}$ is called the group at level i, corresponding to the sequence $u_1, \dots, u_{k_{i+1}}$. The group H_i will be called the group of rank i. We also order the set of t_j's: $t_k < t_l$ if $k < l$.

Let F_i be the free group with the same generating set as H_i, $i = 1, \dots, k_1$, and let $\beta : H_i \to F_i$ be a section, i.e. a mapping of sets such that $\pi \circ \beta = \mathrm{id}$. For our purposes, β has to be of specific form; we will define it explicitly in section 2. For $X \in H_i$ we will call $\beta(X)$ the canonical representative of X.

Definition 1. By an *equational diagram* over a group H_i we mean an equation $X_1 \dots X_n = 1$ together with a solution $A_1, \dots, A_n$ and a diagram over H_i having $\beta(A_1) \dots \beta(A_n)$ as its boundary label. An equational triangle is an equational diagram corresponding to a triangular equation. A *system of equational diagrams* is a system of equations together with the system of diagrams such that the solution associated to each equational diagram is a solution of the whole system.

A *free equational diagram* is an equational diagram with no cells; the corresponding equations are called *free equations*.

If the level i of the group $H_{k_{i+1}}$ is greater than zero, we can use the method of [1] to reduce each equation of the system S to a bounded number of free equations and at most one equation over the group in the previous rank. We can proceed in the same way until we reach level 0 (i.e. the group H_{k_1}), for which we need to develop a separate method. In what follows we will denote H_{k_1} by H.

1. Some properties of the Cayley graph of H

We will frequently use the following lemma:

Lemma 1. ([3, Lemma 1.5]) *For each geodesic triangle $[x_1, x_2, x_3]$ in a δ-space, there are points $y_i \in [x_{i-1}, x_{i+1}]$ (indices are considered modulo 3) such that*

$$d(x_i, y_{i-1}) = d(x_i, y_{i+1}) = (x_{i-1} \cdot x_{i+1})_{x_i},$$

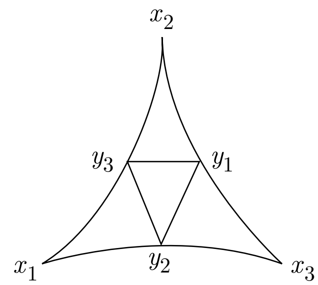


FIGURE 1

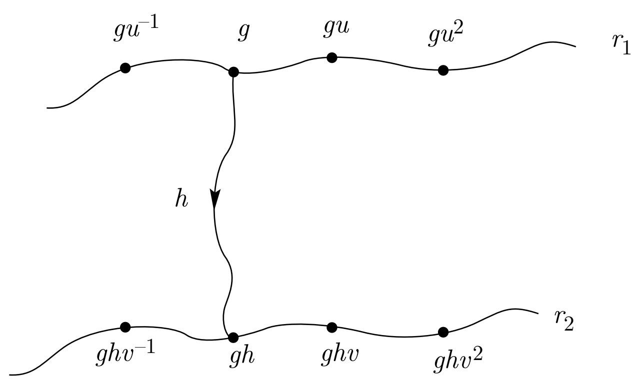


FIGURE 2

$$d(y_i, y_{i-1}) \leq 4\delta \quad \textit{and } d(u, [x_i, y_{i\pm 1}]) \leq 4\delta$$

for any point $u \in [x_i, y_{i\pm 1}]$. We will call the triangle $[y_1, y_2, y_3]$ the Gromov triangle inscribed in $[x_1, x_2, x_3]$ (see Fig. 1).

Lemma 2. [3, Lemma 1.9] *There exists a constant $\mu = \mu(\delta, \lambda, c)$ such that for any (λ, c)–quasigeodesic path p in a δ–space and any geodesic path q with the same initial and terminal points as p, the inequalities $d(Q, p) < \mu$ and $d(P, q) < \mu$ hold for any points $Q \in q$ and $P \in p$.*

Definition 2. Let $u, v \in \{u_1, \dots, u_{k_1}\}$, $|u|_G \geq |v|_G$. In $\Gamma(H)$ consider two paths: a u–path r_1 and a v–path r_2, where r_1 connects the sequence of vertices $\dots, gu^{-1}, g, gu, gu^2, \dots,$ r_2 connects the sequence of vertices $\dots, ghv^{-1}$, gh, ghv, $ghv^2, \dots,$ $h \in H$, $g \in G$. (See Fig. 2.) Then a path q that connects r_1 and r_2 is called a minimal path if

1. the image in H of the label of q is an element in reduced form in all ranks from 0 to k_1 (see [1] for terminology);
2. for each subpath q' of q connecting t–syllables, the image of the label of q' is a geodesic word in G and has minimal length among all words connecting the corresponding t–strips.

Definition 3. Let $d(P, q)$ denote the distance between a point P and a path q in the Cayley graph of G. Then a point P' is called the projection of P onto q if $P' \in q$ and $d(P, P') = d(P, q)$.

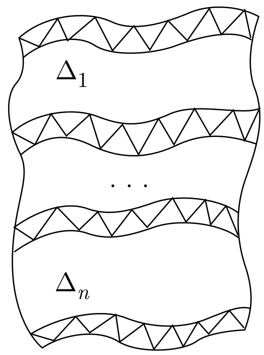


FIGURE 3

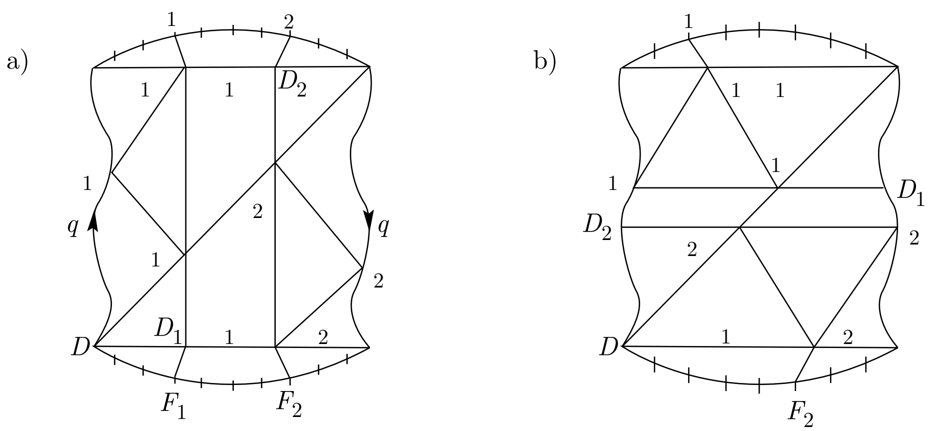


FIGURE 4

The following lemma says that the initial (resp. terminal) points of any two minimal paths are, in a sense, close to each other.

Lemma 3. *Let u, v, r_1, r_2 be as in Definition 2, with r_1 and r_2 distinct; denote by U the cyclic subgroup of G generated by u, and by V the cyclic subgroup generated by v. Let q and q' be two minimal paths connecting r_1 and r_2, X the u-periodic word corresponding to the subpath of r_1 connecting the initial points of q and q', and Y the v–periodic word corresponding to the subpath of r_2 connecting the terminal points of q and q'; i.e.,*

$$q^{-1}Xq' = Y.$$

Then there is a constant $M = M(G, U, V, |u|, |v|)$ such that

$$|X|_G, |Y|_G \leq M.$$

Proof. Whenever the context allows us to do so, we will often use the same notation for a path in $\Gamma(H)$ and the corresponding word in H.

We may assume that q and q' do not contain any t_i's (otherwise subdivide the diagram corresponding to the equality $q^{-1}Xq'$ into several subdiagrams $\Delta_1, \ldots, \Delta_n$

with side labels containing no t–syllables as shown in Fig. 3; for each of them obtain a constant M_i; take M to be $\max(M_1, \dots, M_n)$).

Observe that the subgroups U and V are quasi-isometrically embedded in G; therefore, the paths corresponding to X and Y are (λ, c)–quasigeodesic for some $\lambda > 0$ and $c \geq 0$ (see [3]). Let X_1 and Y_1 be the geodesic paths having the same initial and terminal points as the quasigeodesic paths X and Y.

In the geodesic triangle $ABCD$, draw the diagonal BD. Inside each of the resulting triangles consider the Gromov triangles from Lemma 1. The triangles can be located with respect to each other in two different ways, as shown in Fig. 4 a) and b).

Case a. Let D_1 be the projection on C_1 onto B_2D and D_2 the projection of C_2 onto B_1B. Then by Lemma 1 $d(B_1, C_1) \leq 4\delta$, $d(C_1, D_1) < 4\delta$, $d(B_2, C_2) \leq 4\delta$, $d(C_2, D_2) < 4\delta$; therefore

$$\mid B_1D_1 \mid < 8\delta, \quad \mid B_2D_2 \mid < 8\delta.$$

Now take the projection of B_1 onto X and let E_1 be the phase vertex on X closest to this projection; in the same way, obtain the phase vertices $E_2 \in X$ and $F_1, F_2 \in Y$. Then $d(B_1, E_1), d(D_1, F_1), d(D_2, E_2), d(B_2, F_2)$ are all less than $\mu + \max(\mid u \mid, \mid v \mid)$. If p is the path connecting E_1 with F_1 and p' the path connecting E_2 with F_2, then $\mid p \mid, \mid p' \mid \leq 8\delta + 2(\mu + \max(\mid u \mid, \mid v \mid))$.

If $u \neq v$, the subgroups U and V satisfy the conditions of Lemma 4 from [2]: they are quasi-isometrically embedded in G, malnormal, and $U \cap g^{-1}Vg = \emptyset$ for all $g \in G$ by the choice of u and v. Therefore, there is a constant $M_0 = M_0(G, U, V)$ such that either $\max(\mid E_1E_2 \mid, \mid F_1F_2 \mid) < M_0$ or $4\max(\mid p \mid, \mid p' \mid) \geq \mid E_1E_2 \mid, \mid F_1F_2 \mid$. Since in our case p and p' have bounded length, it follows that $\mid E_1E_2 \mid, \mid F_1F_2 \mid \leq \max(M_0, 4(8\delta + 2(\mu + \max(\mid u \mid, \mid v \mid)))$.

If $u = v$, $\mid E_1E_2 \mid \geq M_0$, then by the same lemma either $\mid E_1E_2 \mid, \mid F_1F_2 \mid \leq 4\max(\mid p \mid, \mid p' \mid)$ or p, p' represent powers of u. In the latter case, the paths r_1 and r_2 coincide, which is the case that we are not interested in.

It remains to bound AE_1, BE_2, CF_2 and DF_1.

Observe that $\mid E_1B_1A_1 \mid \leq 4\delta + \mu + \mid u \mid$; if AA_1 had greater length, then AA_1D would not be a minimal path: $E_1B_1A_1D$, also connecting r_1 with r_2, would be shorter. Therefore, $\mid AA_1 \mid \leq 4\delta + \mu + \mid u \mid$, and $\mid AE_1 \mid \leq \mid AA_1 \mid + \mid E_1B_1A_1 \mid \leq 8\delta + 2\mu + 2 \mid u \mid$, as required. The same argument applies to the remaining paths.

Case b. Let D_1 be the projection of C_1 onto BA_2, D_2 the projection of C_2 onto DA_1. Observe that $\mid E_1B_1C_1D_1 \mid = \mid E_1B_1 \mid + \mid B_1C_1 \mid + \mid C_1D_1 \mid \leq \mu + \mid u \mid + 8\delta$. Therefore, $\mid BD_1 \mid \leq \mu + \mid u \mid + 8\delta$ (otherwise $BC = q'$ is not a minimal path), and $\mid E_1B \mid \leq \mid E_1B_1C_1D_1 \mid + \mid BD_1 \mid \leq 2(\mu + \mid u \mid + 8\delta)$. In a similar way, $\mid AE_1 \mid \leq 2(\mu + \mid u \mid + 4\delta)$, which implies

$$|X| \leq 4\mu + 4 \mid u \mid + 24\delta.$$

The same argument works for Y.

It follows from the above that the constant M can be taken equal to

$$\max(M_0, 4(8\delta + 2(\mu + \max \mid u \mid, \mid v \mid))) + 24\delta + 4\mu + 4 \mid u \mid .$$

□

Therefore, for $u, v \in \{u_1, \dots, u_{k_1}\}$, a u–path r_1 and a v-path r_2 from Definition 2 there are subpaths $r_1' \subset r_1$ and $r_2' \subset r_2$ such that $\mid r_1' \mid, \mid r_2' \mid \leq M$, the initial

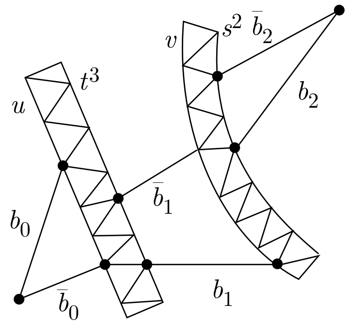


FIGURE 5

points of all minimal paths connecting r_1 and r_2 belong to r'_1, and the terminal points, to r'_2.

Definition 4. Following the terminology of [1], we call a minimal path between r_1 and r_2 a *(u,v)–pseudoconnector for h.* Let q be a (u, v)–pseudoconnector, $p_1 = \iota(q)$, $p_2 = \tau(q)$ ($p_1 \in r'_1$, $p_2 \in r'_2$). Consider the two phase vertices b_1 and b_2 closest to p_2 on each side of p_2 on the v–path r_2. The *connecting zone* for h with respect to g is the union of all phase vertices between such b_1 and b_2 for all (u, v)–pseudoconnectors for h. Denote the initial vertex of the connecting zone by $(h)_1$ and the terminal vertex by $(h)_2$.

2. Construction of canonical representatives

Let F be the free group with the same set of generators $d_1, \dots, d_N$ as G; let $F_i = F * \langle t_1, \dots, t_i\rangle$. For $0 \le i \le k_1$ we will define a section $\beta : H_i \to F_i$ (a mapping of sets such that $\pi \circ \beta = id$). For $X \in H_i$ we will call $\beta(X)$ the canonical representative of X. (See [1] for the construction of canonical representatives for elements in the groups of ranks higher than k_1.)

The canonical representative of an element $X \in G$ is simply a fixed geodesic word representing X. If X is an element of H_i for $1 \le i \le k_1$ and X contains t_i, then it can be written in the form $X = b_0 t_{i_1}^{\alpha_1} b_1 t_{i_2}^{\alpha_2} \cdots t_{i_n}^{\alpha_n} b_n$, where $i_j \in \{1, \dots, i\}$, $b_j \in G$. By $\bar{b}_j$ we denote the canonical representative of the path connecting the vertices $(b_j^{-1})_2$ and $(b_j)_2$; we will call $\bar{b}_j$ the (u, v)–pseudoconnector for the element b_j (here $u = u_{j-1},\ v = u_j$). The points $(b_{j-1})_2$ and $(b_j^{-1})_2$ can be connected by a path with label $t_{i_j}^{\gamma_j}$. Then the canonical representative of X is defined to be $\bar{b}_0 t_{i_1}^{\gamma_1} \bar{b}_1 t_{i_2}^{\gamma_2} \dots \bar{b}_n$.

In the example illustrated in Fig.5, $X = b_0 t^{11} b_1 s^{-7} b_2$, where $u, v \in \{u_1, \dots, u_{k_1}\}$, $u = t^3$, $v = s^2$. Then $\beta(X) = \beta(\bar{b_0}) t^{-5} \beta(\bar{b_1}) s^{-3} \beta(\bar{b_2})$.

Below, all the elements u_j's are always represented by the words $\beta(u_j)$ and we will write u_j instead of $\beta(u_j)$. It will be clear from the context when u_j means a word and when it means the element represented by this word.

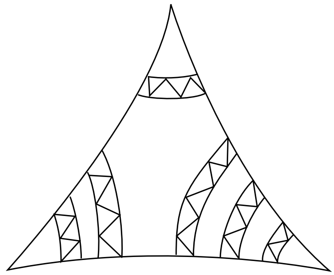


FIGURE 6

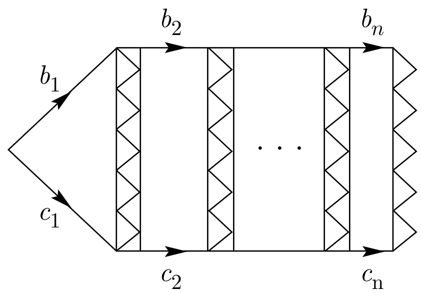


FIGURE 7

3. MIDDLES

Let $X = \{X_1, \dots, X_L\}$ be a solution of the system S in the group $G^{\mathbf{Q}}$. Suppose that this solution is minimal, i.e. contains the minimal possible number of roots. Then for some i the solution X belongs to the group $H_{k_{i+1}} = K$.

For the solution X we will construct another solution $X'_1, \dots, X'_L$ and a system of equations over the group $G * \langle t_1, \dots, t_{k_{i+1}} \rangle$ such that $\beta(X'_1), \dots, \beta(X'_L)$ will be a part of a solution of this new system, and every solution of the new system will give a solution of the system S.

Here we will deal only with equational triangles in the group $H = H_{k_1}$, since for the higher ranks we can use the reduction method developed in [1].

Consider an equational triangle in H with at least one side label containing t_i for some $i \in \{1, \dots, k_1\}$ (hence at least two side labels containing t_i). It is represented by a diagram of the form shown in Fig. 6. Note that the t–strips in Fig. 6 may correspond to distinct t's; a t_i–strip can start on a side of the equational triangle, but not on a u_j–side of a t_j–strip, since $u_j \in G$ and hence doesn't contain any t's.

The following lemma is a direct consequence of the construction of canonical representatives; it is an analog of Lemma 3 from [1].

Lemma 4. *Suppose that in H we have a diagram (see Fig. 7) with the boundary label $(b_1 t_{j_1}^{r_{j_1}} \dots t_{j_n}^{r_{j_n}} b_n t_{j_{n+1}}) t_{j_{n+1}}^{r_{j_{n+1}}} (c_1 t_{j_1}^{p_{j_1}} \dots t_{j_n}^{p_{j_n}} c_n t_{j_{n+1}})^{-1}$, where $j_1, \dots, j_{n+1} \in \{1, \dots, k_1\}$, $b_1, c_1, \dots, b_n, c_n \in H$. Then*

$$\beta(b_1 t_{j_1}^{r_{j_1}} \dots t_{j_n}^{r_{j_n}} b_n t_{j_{n+1}}) t_{j_{n+1}}^{r_{j_{n+1}}} = \beta(c_1 t_{j_1}^{p_{j_1}} \dots t_{j_n}^{p_{j_n}} c_n t_{j_{n+1}}).$$

It follows from Lemma 4 that every equational triangle either does not contain any cells or takes on the form shown in Fig. 8 and hence has a unique maximal

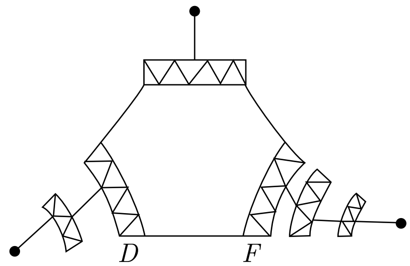


FIGURE 8

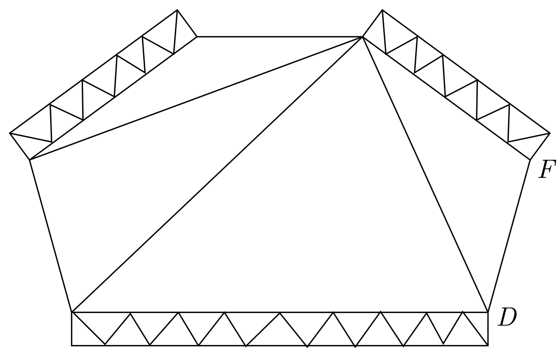


FIGURE 9

nontrivial G–subdiagram. (Some of the t–strips in Fig. 8 might, of course, be trivial.)

Definition 5. Consider an equational triangle Δ in $\Gamma(H_i)$ $(1 \leq i \leq k_1)$ with the boundary label $\beta(X_1)\beta(X_2)\beta(X_3)$. The maximal nontrivial G-subdiagram of Δ is called the middle of this triangle. For example, the subdiagram $ABEFDC$ in Fig. 8 is a middle.

Let $u \in \{u_1, \ldots, u_{k_1}\}$, $t \in \{t_1, \ldots, t_{k_1}\}$. A boundary of a middle is canonically subdivided into paths. Each of these paths is either a u–path joining two phase vertices (i.e. belongs to the u-side of a t-strip) or a connector. The u-paths are called *pseudoangles* of the middle.

Let $\nu = \max\{\nu_1, \ldots, \nu_8\}$, where ν_i are certain fixed constants depending on the group G and $u_1, \ldots, u_{k_1}$; we will see how to calculate the ν_i's in the proof of Lemma 5.

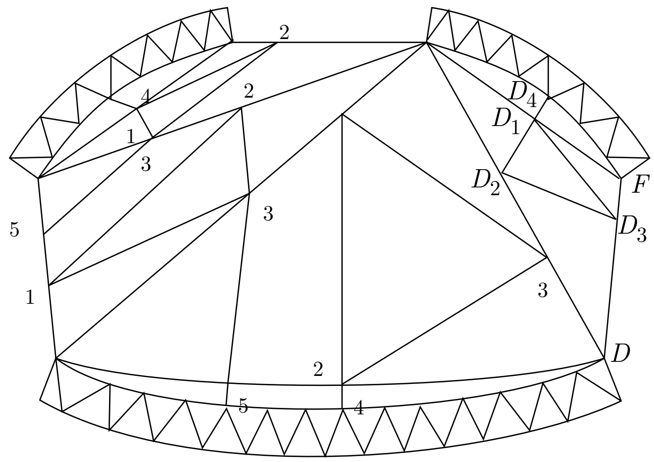
A pseudoangle p is said to be *trivial* if the corresponding u–path is trivial; *short* if $|\, p \,| \leq \nu$; *long* otherwise. A middle is called *triangular* if all the pseudoangles are trivial.

If none of X_i's contains t then the middle of the triangle coincides with the triangle itself and is a triangular middle.

As an example, consider the middle $ABEFDC$ in Fig. 8. The paths AB, CD, and EF are pseudoangles of this middle.

Each middle represents an equation over the group G; denote the corresponding system of equational diagrams by $\widetilde{S}$.

a)



b)

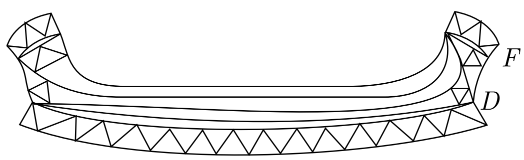


FIGURE 10

Lemma 5. *Let M be a middle of an arbitrary equational triangle over H. Then at most one pseudoangle of M is long.*

Proof. We will provide a proof for the case where none of the pseudoangles is trivial. Take an arbitrary triangulation of the geodesic hexagon obtained from the middle M by replacing the pseudoangles, which are quasogeodesic paths, by the corresponding geodesic paths. We will use the triangulation shown in Fig. 9. Here AB is a u–path, CD is a v–path and EF is a w–path for $u, v, w \in \{u_1, \dots, u_{k_1}\}$. Inside each of the resulting triangles ABE, ACE, CED and DEF consider the Gromov triangles from Lemma 1. These Gromov triangles can be arranged with respect to each other in eight different ways. Instead of providing a detailed proof for each of the eight possibilities, we will consider two typical cases; the others can be analyzed similarly. Note also that if one or two of the pseudoangles are trivial, then instead of the geodesic hexagon we obtain a pentagon or a quadrangle, which leaves us with fewer cases to consider.

For an arbitrary path q in the Cayley graph of H, we denote by $\widetilde{q}$ the geodesic path with the same initial and terminal points as q. If P_1 is the projection of a point P onto a path q_1 and P_2 is the projection of P_1 onto a path q_2, then we will call P_2 the projection of P onto q_2 through q_1. □

Case 1. (Fig. 10 a).) Let A_4 denote the projection of $A_1 \in \widetilde{AB}$ onto AB, C_4 the projection of $C_1 \in \widetilde{CD}$ onto CD and D_4 the projection of $D_1 \in \widetilde{EF}$ onto EF.

Claim: A_4B is short.

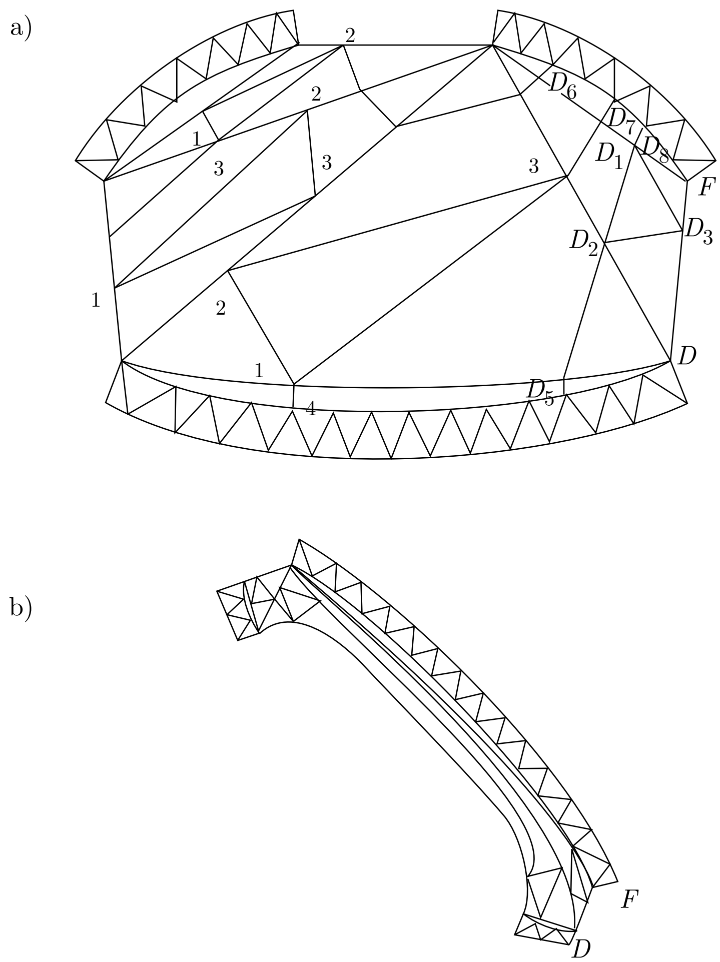


FIGURE 11

Proof. We may assume that $\mid A_4B \mid > M$ (the length of the connecting zone for BE); otherwise there is nothing to prove. Let p be the minimal path between AB and EF whose initial point is the farthest from B; $\iota(p)$ is between A_4 and B. The path BE is a connector, so its length is bounded by the following: $\mid BE \mid \leq \mid p \mid +2M$.

We now claim that $\mid BA_2 \mid \leq \mid A_4A_1A_2 \mid +2M$. Indeed, suppose for contradiction that $\mid BA_2 \mid > \mid A_4A_1A_2 \mid +2M$. Consider the path BA_2E.

$$\mid p \mid +2M \geq \mid BA_2E \mid = \mid BA_2 \mid + \mid A_2E \mid$$

$$> \mid A_4A_1A_2 \mid + \mid A_2E \mid +2M = \mid A_4A_1A_2E \mid$$

It follows that $A_4A_1A_2E$ has length less than $\mid p \mid$, and therefore is a minimal path between AB and EF, but its initial point falls outside the connecting zone for BE – a contradiction.

Therefore $| BA_2 | \leq | A_4A_1 | + | A_1A_2 | + 2M \leq \mu + 4\delta + 2M$; by Lemma 1 $| A_1B | = | BA_2 | \leq \mu + 4\delta + 2M$; finally, $| A_4B | \leq | A_4A_1 | + | A_1A_2 | \leq 2\mu + 4\delta + 2M$, as required. □

Claim: AA_4 is short.

Proof. Let A_5 denote the projection of A_3 onto AB_1; by the same argument as above $| AA_5 | \leq A_4A_1A_3A_5 | + 2M$ (otherwise AC is not a connector). Consequently, we have: $| AA_5 | \leq \mu + 8\delta + 2M$, and $| AA_4 | \leq | AA_5 | + | A_4A_1A_3A_5 | \leq 2\mu + 16\delta + 2M$. □

Combining the two claims, we obtain that AB is a short pseudoangle.

In the same way one can see that CC_5, C_4D, FD_4 and D_4E are all short (here C_5 is B_3 projected onto CD through them onto CC_1). However, it is not guaranteed that C_5C_4 is short, so it is possible for the pseudoangle CD to be long. Fig. 10 b) gives a more precise idea of what the middle looks like in this case.

As one can see from the calculations, $| AB |, | EF | \leq 4(\mu + 10\delta + M)$; denote this constant by ν_1.

Case 2. (Fig. 11 a).)

Let D_5 denote the phase vertex closest to the projection of D_2 onto CD through DC_1; C_4 the phase vertex closest to the projection of C_1 onto CD; D_6 the projection of A_2 onto EF through B_2E, B_3E, C_3E and ED_1.

In this case, AB is short by the same argument as above; it is also clear that the following are short: CC_4, D_1F, ED_6 and DD_5.

Claim: C_4D_5 is short.

Proof. Let $D_7 \in EF$ be the phase vertex closest to the projection of C_3 onto EF through ED_1; D_8 the phase vertex in EF closest to the projection of D_1. Denote the path $C_4C_1C_3D_7$ by p and $D_5D_2D_1D_8$ by q. Then by [1, Lemma 4] $| D_7, D_8 |$, $| C_4D_5 | \leq \max(M_0, | p |, | q |)$ for some constant M_0. (Note that neither v nor w participates in p or q.) Since $| p |$ and $| q |$ are bounded by $2\max(| v |, | w |) + 2\mu + 8\delta$, this gives us a boundary on the length of C_4D_5.

D_6D_7 might turn out to be long. See Fig. 11 b) for the shape of the middle in this case.

These calculations give us the constant ν_2 bounding the lengths of the short pseudoangles AB and CD:

$$\nu_2 = 32\delta + 4\mu + 8M + \max(M_0, 2\max\{ |u_1|, \dots, | u_{k_1} |\} + 2\mu + 8\delta).$$

In a similar way, one can compute the constants $\nu_3, \dots, \nu_8$ bounding the lengths of the short pseudoangles in the remaining cases. □

The following lemma can be proved by the same method as Lemma 5.

Lemma 6. *There are constants*

$$M_1 = M_1(G, u_1, \dots, u_{k_1}) \text{ and } M_2 = M_2(G, u_1, \dots, u_{k_1})$$

such that for every middle $ABEFDC$ (Fig. 12) with a long pseudoangle CD there exist points $P, Q \in BE$ and phase vertices $P_1, Q_1 \in DC$ with the following properties:

1. *$d(P, P_1) < M_1$, $d(Q, Q_1) < M_1$ and for all $P' \in PQ$ we have $d(P', P_1Q_1) < M_1$;*

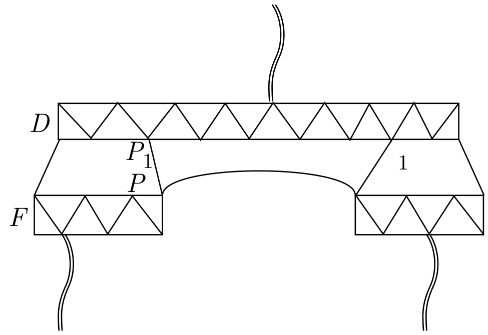


FIGURE 12

2. $| DP_1 |, | Q_1C | < M_2 \max(| u_1 |, \dots, | u_{k_1} |)$.

Proof. Let us take the same triangulation of the middle $ABEFDC$ as in the proof of Lemma 5. We will provide a proof for the case where the Gromov triangles inscribed in the triangles ABE, AEC, CED and DEF are located with respect to each other as shown in Fig. 10 a); denote by M_1^1 and M_2^1 the constants satisfying the conditions of the lemma in this case. Similarly, one can consider the remaining seven cases to obtain the constants M_i^j, $i = 1, 2$, $j = 2, \dots, 8$. Then put $M_1 = \max(M_1^1, \dots, M_1^8)$, $M_2 = \max(M_2^1, \dots, M_2^8)$.

In the notation of Lemma 5, let P be the projection of C_2 onto EA_2 through EB_2, Q the projection of B_2 onto PA_2, and P_1, Q_1 the phase vertices on CD closest to C_4 and C_5, respectively. Then $d(P, P_1)$, $d(Q, Q_1) \leq 12\delta + \mu + \max(|u_1|, \dots, |u_{k_1}|) = M_1^1$.

In the same way as in the proof of Lemma 5, one can see that $|CC_5|$, $|C_4D| \leq 16\delta + 2\mu + 2M$. Dividing this constant by $\max(|u_1|, \dots, |u_{k_1}|)$, taking the integer part of the resulting number and adding 1, we obtain the desired constant M_2^1. □

4. SHRINKING OF THE MIDDLE t-STRIPS

Our next objective is to bound the powers of u_j for $j = 1, \dots, k_1$ participating in the long pseudoangles. This will allow us to bound the number of triangular equations over the group G obtained from the middles by triangulation. (For instance, the middle with the boundary label $\beta(s_1)\beta(u_j^l)\beta(s_2)\beta(s_3)$ shown in Fig. 13 produces $l + 1$ triangular equations.)

Taking the middles of all equational triangles over the group H corresponding to the system S, we obtain a system of equational diagrams; let us denote this

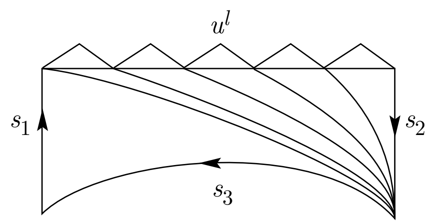


FIGURE 13

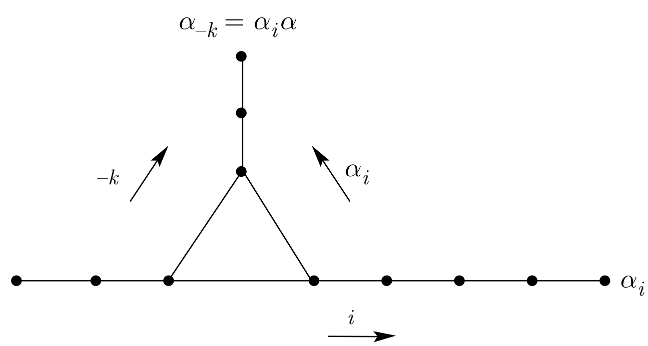


FIGURE 14

system by $\widetilde{S}$. Each diagram of the system $\widetilde{S}$ has the boundary label of the form $\beta(X_1)\dots\beta(X_l)$. In what follows, it will be convenient to temporarily replace $\beta(X_i)$ for some of the X_i by the canonical representative of X_i in the sense of E. Rips and Z. Sela ([4]). (We will explain precisely for what X_i we do this replacement after we define shrinkable words later in this section.) Let us recall the definition of Rips-Sela representatives, following the interpretation given by T. Delzant in [5].

Let D be an integer, $\widetilde{S} : \langle a_1, \dots, a_n; R_1, \dots, R_T\rangle$ a triangular system of equations in a torsion–free hyperbolic group G, and $\alpha_1, \dots, \alpha_n$ a solution. Denote $a_i^{-1} = a_{-i}$, $\alpha_i^{-1} = \alpha_{-i}$; the equations are written in the form $\tau_{i,j,k} : a_ia_ja_k = 1$, where $i, j, k \in \{\pm 1, \dots, \pm n\}$.

Definition 6. A family of D–canonical representatives in the sense of E. Rips and Z. Sela of $\alpha_1, \dots, \alpha_n$ is a family of quasigeodesic polygonal paths $A_i(t)$ of length L_i in the Caley graph $\Gamma(G)$ of G ($A_i(t) : [0, L_i] \longrightarrow \Gamma(G)$, $L_i \in \mathbf{N}$, $\alpha_i = A_i(L_i)$) parametrized by arc length and satisfying the condition below.

Set $A_{-i}(t) = \alpha_{-i}A_i(L_i - t)$; let $\tau_{ijk} : a_ia_ja_k$ be one of the equations. Then the triangle $A_i, \alpha_iA_j, \alpha_i\alpha_jA_k$ is D–flat (see Figure 14) in the following sense:

$$A_i(t) = A_{-k}(t) \ \text{ if } \ 0 \le t \le \frac{L_i + L_k - L_j}{2} - \frac{D}{2};$$

$$A_{-i}(t) = A_j(t) \ \text{ if } \ 0 \le t \le \frac{L_i + L_j - L_k}{2} - \frac{D}{2};$$

$$A_{-j}(t) = A_k(t) \ \text{ if } \ 0 \le t \le \frac{L_j + L_k - L_i}{2} - \frac{D}{2}.$$

The paths A_i are quasigeodesic, in other words,

$$\frac{|t-u|}{v_{10\delta}} \le |A_i(t) - A_i(u)| \le |t-u|(20\delta + 1),$$

where $v_r \le (2N)^r$ denotes the number of elements in a ball of radius r, N is the number of generators of G.

Let λ' denote the constant $1/v_{10\delta}$. It follows from Lemma 2 that there exists a constant $\mu' = \mu'(\delta, \lambda')$ such that for every geodesic word α and its Rips-Sela representative A the inequalities $d(Q, A) < \mu'$ and $d(P, \alpha) < \mu'$ hold for any points $Q \in \alpha$ and $P \in A$.

The main technical result of E. Rips and Z. Sela ([4]) consists in the following.

Theorem 1. [4] *Let T be a positive integer. Then there exists a number D depending only on T and the generators of G such that for every system $\widetilde{S}$ of T triangular equations over G, each solution $\alpha_1, \dots, \alpha_n$ of $\widetilde{S}$ admits a family of D–canonical representatives.*

Now let us return to the system $\widetilde{S}$ of equational diagrams over G obtained by reduction from the original system S.

Definition 7. A middle t_j–strip is a t_j–strip whose u_j–side belongs to the long pseudoangle of a middle of an equational triangle.

Definition 8. A subword w of the label of the boundary of a middle is said to be *shrinkable* if one of the following conditions is satisfied:

1. w is the label of the u–side of a middle t–strip (e.g., CD in Fig. 10 b);
2. w is a connector such that a subword v of w is "close" to the u–side q of a middle t–strip, meaning that for every point $P \in v$ $d(P, q) \leq M_1 + \mu'$ (e.g., BE in Fig. 10 b);
3. w is a t–periodic word on the t–side of a middle t–strip;
4. if w is the label of a common subpath of two paths p and p' shrinkable as a subword of the label of p, then it is shrinkable as a subword of the label of p'.

For each $j = 1, \dots, k_1$ and each middle t_j–strip, we will replace the power of u_j on its boundary by a bounded power of u_j; during this procedure, the solution might be altered, but it will remain a solution of the original system of equations.

When we replace a u_j–word w by another word, we should make appropriate changes to the diagram whose boundary contains the word w, as well as to the adjacent diagrams. Consider, for example, the diagram $ABCED$ shown in Fig.15 a); here CD is "close" to AB in the sense of Definition 8. If we remove a subword w from AB, we will also have to remove a subword of CD and a subword of DE (or EC) together with some cells. Similarly, in the diagram $ABCFED$ in Fig. 15b), removing a subword of AB forces us to remove some subwords of CD and DE (or CF). In this diagram we assume that both CD and EF are close to AB, and EF is a short pseudoangle of the diagram $CDEF$ (if it were long, it could not be close to AB by Lemma 3).

For the sake of convenience, we will now denote by $\beta(X_i)$ the Rips–Sela representatives for the words X_i that are on the boundary of triangular diagrams from the system $\widetilde{S}$ (e.g., CED from Fig. 15 a)) and the diagrams obtained by triangulation from polygonal diagrams containing no long pseudoangles ($CDEF$ in Fig. 15 b)). The diagrams from Fig 15 a),b) now take shape shown in Fig 15 c),d) respectively.

Definition 9. Let σ be the least integer greater than

$$3M_2 + 2D/\max(|u_1|, \dots, |u_{k_1}|).$$

The length $L(w)$ of a shrinkable word w is defined as follows.

1. $L(w) = k - \sigma$ if $w = u_j^k$, $j \in \{1, \dots, k_1\}$.
2. $L(w) = k - \sigma$ if $w = t_j^{s_j k + r}$, $0 < r < s_j$, $j \in \{1, \dots, k_1\}$.
3. Let w (or a subword of w) be close to the u_j–side q of a middle t_j–strip in the sense of Definition 8. Denote by P and Q the phase vertices closest to the projections of $\iota(w)$ and $\tau(w)$ onto q, respectively; the path PQ corresponds to a certain subword u_j^k of q. Then we put $L(w) = k - \sigma$.

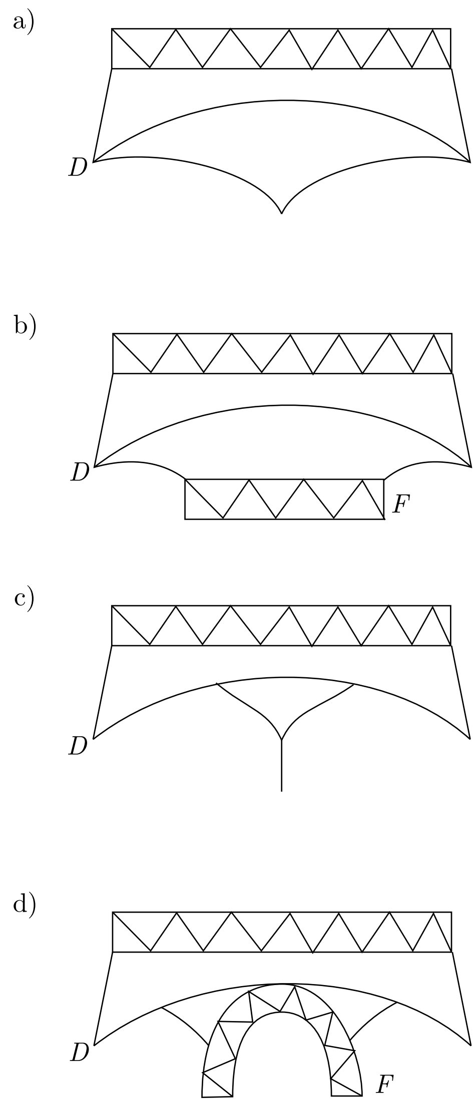


Figure 15

Definition 10. A maximal shrinkable subdiagram Δ of an equational triangle is the minimal subdiagram having three maximal shrinkable words on the boundary.

In Fig. 16 we illustrate all possible shapes of maximal shrinkable subdiagrams. (The diagram ABC in Fig. 16 d) is shrinkable since the path AB is "close" to a u–side of a middle t–strip.)

Every maximal shrinkable subdiagram Δ_k gives us an equation over the integers as described below. Let a_{k1}, a_{k2} and a_{k3} be the maximal shrinkable words on the boundary of Δ_k and let u be the element from the set $\{u_1, \dots, u_{k_1}\}$ used in the definition of length for the words a_{kj}; denote $L(a_{kj})$ by l_{kj}. (We will be only interested in the case where at least one of l_{kj} is positive, since we do not need

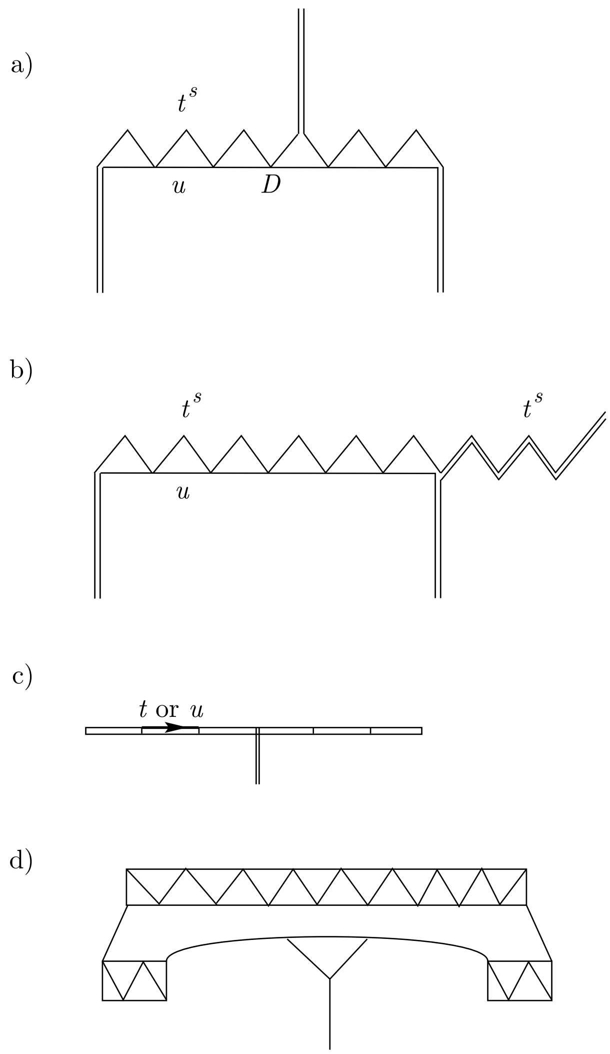


FIGURE 16

to "shrink" any words of nonpositive length.) Then to the subdiagram Δ_k we associate the equation $l_{k1} + l_{k2} + \sigma + s = l_{k_3}$, where $s \in \{0, \ldots, D'\}$ and D' is the least integer greater than $D/|u|$. (For example, if Δ_k is the diagram shown in Fig. 16 a), then the t_j–word AD is of length l_{k1}, the t_j–word EC of length l_{k2} and the u_j–word AC of length l_{k_3}; we obtain the equation $l_{k1} + l_{k2} + \sigma + 1 = l_{k_3}$.)

We obtain an equation of this form from every maximal shrinkable subdiagram; denote the system of these equations by L.

It is possible to obtain only a finite number of distinct linear systems of this form. Every such system is algorithmically solvable, since the elementary theory of the natural numbers with addition is decidable. Choose arbitrarily a solution for each system. Let l be the maximum of l_{ij} in these solutions. Then the system L also has a solution bounded by l. Let $\{\bar{l}_{ij} \mid 1 \le i \le m', 1 \le j \le 3\}$ be this solution.

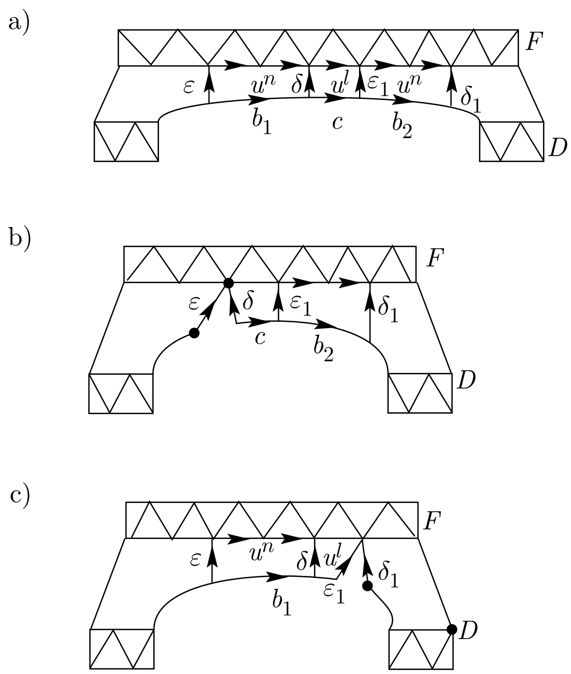


Figure 17

We now replace each word a_{kj}, which has length l_{kj}, by a word $\bar{a}_{kj}$ of length $\bar{l}_{kj}$ as follows. If $a_{kj} = u^{l_{kj}+\sigma}$ for $u \in \{u_1, \dots , u_{k_1}\}$, then $\bar{a}_{kj} = u^{\bar{l}_{kj}+\sigma}$; the same procedure applies if a_{kj} is a power of $t \in \{t_1, \dots , t_{k_1}\}$. Now let a_{kj} be a maximal shrinkable word of the third type, meaning that a subword of a_{kj} is close to the u–side q of a middle t–strip. Let P be the initial point of the path corresponding to the word a_{kj} and Q its endpoint. Denote by P' and Q' the phase vertices closest to the projections of P and Q onto q, respectively; the path $P'Q'$ represents the subword $u^{l_{ij}+\sigma}$ of q. Replace an arbitrary entry of $u^{l_{ij}}$ in $P'Q'$ by $u^{\bar{l}_{ij}}$, making the appropriate changes to the diagram $PP'Q'Q$. Then $\bar{a}_{kj}$ is the word corresponding to the path PQ after these changes.

Replacing all the words a_{jk} by $\bar{a}_{jk}$, we again get a solution of the original system of equations (see [1, Lemma 11]).

When we cut a power of u out of a long pseudoangle of a middle, we also have to cut the corresponding cells out of the middle. Here we should make sure that no matter where on the path EF (Fig. 17) we cut out u^n, the effect on the word BC will be the same . More precisely,

Lemma 7. *In the diagram shown in Fig. 17 a), $\varepsilon\delta^{-1}cb_2 = b_1c\varepsilon_1\delta_1^{-1}$ (in other words, after removing the first copy of u^n, we get the same result as after removing the second copy).*

Proof. Indeed,

$$b_1c\varepsilon_1 = \varepsilon u^n u^l = \varepsilon u^{n+l}; \tag{1}$$

$$\delta^{-1}cb_2 = u^{l+n}\delta_1^{-1}. \tag{2}$$

Multiplying (2) by ε and replacing εu^{n+l} in the right-hand side by the right-hand side of (1), we obtain the required equality. □

The proof of Main Theorem is now essentialy the same as that of [1, Theorem 1]. It has to be mentioned that the original system S is finally reduced to a finite set F of systems in the hyperbolic group $\widetilde{G} = G * \langle t_1, \ldots, t_{\psi(m)} \rangle$, with the restriction that some of the variables belong to subgroups generated by only a part of the generating system of $\widetilde{G}$ (i.e., $u_{k_i+1}, \ldots, u_{k_{i+1}} \in H_i$). This restriction, however, does not prevent us from applying the Rips-Sela method to test each system from the set F for solvability. Indeed, if a system has a solution $X_1, \ldots, X_L$ with some X_j belonging to the group $G * \langle t_1, \ldots, t_i \rangle$, then the Rips-Sela representative of X_j is forced to be in the group $F * \langle t_1, \ldots, t_i \rangle$. (Here F stands for the free group with the same generating system as G.) Therefore, in the Rips-Sela method, the system over the hyperbolic group with the restrictions described above gets reduced to a system over a free group with the same restrictions, which is algorithmically decidable by [6].

References

[1] O. Kharlampovich, A. Myasnikov. *Equations in a free* $\boldsymbol{Q}$*–group.* Trans. Amer. Math. Soc. **350** (1998), 947–974. CMP 98:05

[2] O. Kharlampovich, A. Myasnikov. *Hyperbolic groups and free constructions.* Trans. Amer. Math. Soc. 350 (1998), 571–613.

[3] A. Yu. Ol'shanskii. *On residualing homomorphisms and G-subgroups of hyperbolic groups.* International Journal of Algebra and Computation, 3(4):365–409, 1993. MR **94i:**20069

[4] E. Rips and Z. Sela. *Canonical representatives and equations in hyperbolic groups.* Invent. Math., 120: 489–512, 1995. MR **96c:**20053

[5] Thomas Delzant. *Finitely presented subgroups of hyperbolic groups.* (Preprint)

[6] G. S. Makanin. *Equations in a free group.* Math. USSR Izv., V. 21, 1983. MR **84m:**20040

Department of Mathematics and Statistics, McGill University, 805 Sherbrooke St. West, Montreal, Quebec, H3A 2K6 Canada

E-mail address: `olga@triples.math.mcgill.ca`

Department of Mathematics and Statistics, McGill University, 805 Sherbrooke St. West, Montreal, Quebec, H3A 2K6, Canada

E-mail address: `kate@triples.math.mcgill.ca`

Department of Mathematics, City College (CUNY), Convent Ave. at 138th Street, New York, New York 10031-9100

E-mail address: `alexei@jolly2.sci.ccny.cuny.edu`