PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 127, Number 7, Pages 1885–1891
S 0002-9939(99)04579-7
Article electronically published on March 1, 1999

ON DEHN FUNCTIONS OF FREE PRODUCTS OF GROUPS

V. S. GUBA AND M. V. SAPIR

(Communicated by Ronald M. Solomon)

Abstract. In this article we show that the Dehn function of a nontrivial free product of groups is equivalent to its subnegative closure.

Let $\mathcal{P} = \langle \Sigma \mid \mathcal{R} \rangle$ be a finite presentation of a group G. Suppose that a word w over Σ is equal to 1 in G. By van Kampen's Lemma [6], there exists a *diagram* over $\mathcal{P}$ such that w is a *boundary label* of this diagram. By $k(w)$ we denote the smallest number of *cells* in such a diagram. In other words, $k(w)$ is the smallest number with the following property: w is equal in the free group over Σ to a product of $k(w)$ conjugates of the defining relators from $\mathcal{R}$ or their inverses.

A function $f\colon \mathbf{N} \to \mathbf{N}$ is called the *Dehn function* of the presentation $\mathcal{P}$ whenever $f(n) = \max k(w)$, where maximum is taken over all words such that w equals 1 in G and $|w| \le n$.

Let f, g be functions from $\mathbf{N}$ to $\mathbf{N}$. By definition, $f \preceq g$ means that there exists a positive integer C such that $f(n) \le Cg(Cn) + Cn$ for all n. This relation induces an equivalence relation on the set of functions from $\mathbf{N}$ into itself: $f \simeq g$ if and only if $f \preceq g$ and $g \preceq f$.

It is well known (see [7], [1], [4]) that if G is a finitely presented group and f, g are the Dehn functions of two finite presentations of G, then $f \simeq g$. In this paper, we shall not distinguish between equivalent functions. Thus we will speak about the Dehn function of a finitely presented group G.

Following Brick [3], we say that a function $f\colon \mathbf{N} \to \mathbf{N}$ is *subnegative* whenever $f(m) + f(n) \le f(m+n)$ for all $m, n \in \mathbf{N}$. For every function $f\colon \mathbf{N} \to \mathbf{N}$ one can define a function $\bar{f}(n)$ by the following formula:

$$\bar{f}(n) = \max\,(f(n_1) + f(n_2) + \cdots f(n_r)), \tag{1}$$

where maximum is taken for all $r \ge 1$, and $n_1, \ldots, n_r \in \mathbf{N}$ such that $n_1 + \cdots + n_r = n$. It is easy to see that $\bar{f}(n)$ is the smallest subnegative function which is greater than or equal to $f(n)$. The function $\bar{f}$ is said to be the *subnegative closure* of f.

The subnegative property plays an important role in [2], where it is proved that the class of Dehn functions is very large. Roughly speaking, every relatively fast computable subnegative function $\ge n^4$ is the Dehn function of a finitely presented group. More precisely, if $T(n)$ is a subnegative time function of a nondeterministic

Received by the editors June 27, 1997.
1991 *Mathematics Subject Classification.* Primary 20F32; Secondary 57M07.
The research of the first author was supported in part by grants from the ISF and the Russian Foundation for Fundamental Researches, grant no. 96–01–00420. The research of the second author was supported in part by NSF grants.

Turing machine such that $\sqrt[4]{T(n)}$ is equivalent to a time function, then $T(n)$ is the Dehn function of a finitely presented group. On the other hand, every Dehn function of a finitely presented group is equivalent to the time function of a nondeterministic Turing machine. Thus if every Dehn function were equivalent to a subnegative function and in addition the fourth root of every time function were equivalent to a time function (which is quite possible), then the main result from [2] would give a complete description of all Dehn functions of groups which are greater than n^4.

It is easy to see that if $f(n)$ is the Dehn function of a finite **presentation**, then f may not be subnegative. For example, for the presentation $\langle a \mid a^2 = 1, a^3 = 1 \rangle$ one has $f(1) = 2$, $f(2) = 2$. But it is natural to ask whether every Dehn function is **equivalent** to a subnegative function. The second author conjectured that this is always true.

Conjecture 1. The Dehn function of any finite presentation is equivalent to a subnegative function.

Recall that not every increasing function is equivalent to a subnegative one (see [3]).

Let us note that if $f \preceq g$, then $\bar{f} \preceq \bar{g}$. Therefore $f \simeq g$ implies $\bar{f} \simeq \bar{g}$. We obtain the following equivalent

Conjecture 1′. Every Dehn function of a finite presentation is equivalent to its subnegative closure.

The aim of this paper is to show that the conjecture is true for groups which are nontrivially decomposable into a free product.

Theorem. *Let G_1, G_2 be nontrivial finitely presented groups. Then the Dehn function of their free product $G_1 * G_2$ is equivalent to its subnegative closure.*

Corollary. *Let G_1, G_2 be nontrivial finitely presented groups and let f_1, f_2 be their Dehn functions. The Dehn function f of the group $G_1 * G_2$ satisfies the following equality: $f \simeq \max(\bar{f}_1, \bar{f}_2)$.*

This corollary improves Proposition 3.1 in Brick [3] which says that

$$\max(f_1, f_2) \preceq f \preceq \max(\bar{f}_1, \bar{f}_2).$$

The corollary follows immediately from this inequality and from our main result.

This corollary immediately implies that our conjecture is equivalent to the following statement:

Conjecture 1″. For any finitely presented group G, the Dehn functions of G and $G * \mathbf{Z}$ are equivalent.

One might say that if this conjecture is wrong, then something is wrong with the definition of Dehn functions. More seriously, it could be better to call the "true Dehn function" of a group the subnegative closure of the Dehn function. Our Corollary shows that "true Dehn functions" behave nicely with respect to free products. Since all known Dehn functions are subnegative, this change of definition will not affect any known results about Dehn functions.

To prove our theorem, we use diagrams over group presentations (see [6]). Let us recall that $\varphi(e)$ denotes the label of an edge e of a diagram. All contours of diagrams are read clockwise, all contours of cells are read counterclockwise. By $\iota(e)$, $\tau(e)$ we denote the initial and the terminal vertices of an edge e respectively.

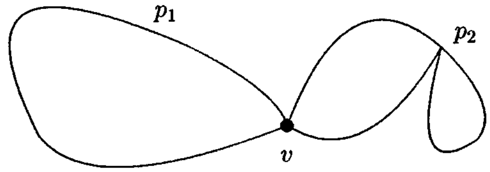

FIGURE 1

We need the following auxiliary definition.

Let Δ be a diagram over a group presentation and let v be a vertex in the boundary of Δ. We say that v is a *double vertex* of Δ if Δ has a contour of the form p_1p_2 where p_1, p_2 are nontrivial loops at v (see Figure 1).

Let $\langle \Sigma_i \mid \mathcal{R}_i \rangle$ be a finite presentation of a group G_i $(i = 1, 2)$. We assume that Σ_1 and Σ_2 do not intersect. It is obvious that $\langle \Sigma \mid \mathcal{R} \rangle$ is a finite presentation for $G_1 * G_2$, where $\Sigma = \Sigma_1 \cup \Sigma_2$, $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$. We keep these definitions throughout the paper.

Lemma 1. *Let Δ be a diagram over $\langle \Sigma \mid \mathcal{R} \rangle$. Suppose that it has a contour of the form e_1pe_2, where e_1, e_2 are edges. If $\varphi(e_i) \in \Sigma_i^{\pm 1}$ for each $i \in \{1, 2\}$, then $v = \iota(e_1) = \tau(e_2)$ is a double vertex of Δ.*

Proof. Let $h_0, h_1, \dots, h_r$ $(r \geq 1)$, where $h_0 = e_1$, $h_1 = e_2^{-1}$, be the sequence of all edges going out from v in the clockwise order. (Note that this sequence may have occurrences of the form h, h^{-1} if h forms a loop.) Each of the edges h_j $(0 \leq j \leq r)$ has a label either from $\Sigma_1^{\pm 1}$ or from $\Sigma_2^{\pm 1}$. Since h_0 and h_r have labels from different alphabets, there exist two neighbour edges h_j and h_{j+1} $(0 \leq j < r)$ such that their labels belong to different alphabets.

Since there are no defining relations from $\mathcal{R}$ that contain letters from both of the two alphabets, there are no cells "between" h_j and h_{j+1} (i.e. no contours of cells of Δ have subpaths $h_j^{-1}h_{j+1}$). It is clear that the edge h_j^{-1} cannot belong to a boundary contour of any cell since otherwise there are edges between h_j and h_{j+1}. The same argument applies for the edge h_{j+1}. This implies that h_j^{-1}, h_{j+1} belong to a boundary contour of Δ and moreover, they are consecutive edges of the cyclic contour. We obtain that the contour of Δ at v has a form p_1p_2 where $p_1 = e_1 \dots h_j^{-1}$, $p_2 = h_{j+1} \dots e_2$ are nontrivial loops at v. Thus v is a double vertex. The proof is complete.

Lemma 2. *Let Δ be a diagram over $\langle \Sigma \mid \mathcal{R} \rangle$. Suppose that G_1 and G_2 are nontrivial groups. Let a_i denote an element in Σ_i such that $a_i \neq 1$ in G_i $(i = 1, 2)$. Then there exists a vertex v in the boundary of Δ and there exists an index $j \in \{1, 2\}$ such that for any decomposition pq of the contour of Δ as a loop at v, we have $\varphi(p) \neq a_j$ in G.*

Proof. Suppose that the conclusion is not true. Let v_0 be an arbitrary vertex in the boundary of Δ. Let us define an infinite sequence v_m of vertices in the boundary of Δ and an infinite sequence p_m of subpaths of the cyclic contour of Δ by induction on m. Let $m \geq 1$ and let v_j, p_j be defined for all $0 \leq j < m$. By our assumption, for any $i = 1, 2$ there exists a decomposition $p(i)q(i)$ of the contour of Δ at v_{m-1} such that $\varphi(p(i)) = a_i$ in G. Let p_m be $p(1)$ for odd m and $p(2)$ for even m. By definition, v_m is the terminal vertex of p_m. Note that all p_m are nontrivial since a_1, a_2 are not equal to 1 in G.

Since the set of the boundary vertices of Δ is finite, we have $v_m = v_{m+s}$ for some $m \geq 0$, $s \geq 1$. Thus the path $p = p_{m+1} \dots p_{m+s}$ is a nontrivial power of a boundary label of Δ. Since the boundary label of any diagram is equal to 1 in the corresponding group, this implies that $\varphi(p) = 1$ in G. But $\varphi(p)$ is a nontrivial product of letters a_1, a_2 such that all neighbour letters are different. This shows that $\varphi(p)$ represents a normal form of the free product $G_1 * G_2$. But a nonempty normal form represents a nontrivial element of a free product (see [6]). We have a contradiction. This completes the proof.

Now we are able to prove our Theorem.

Let f be the Dehn function of $\langle \Sigma \mid \mathcal{R} \rangle$ where G_1, G_2 are nontrivial. We shall prove the inequality

$$f(n_1) + f(n_2) \leq f(n_1 + n_2 + 6) \tag{2}$$

for all n_1, n_2. Let us show that inequality (2) implies the statement of the Theorem. Notice that the Dehn function of any presentation is increasing. It follows easily from this fact and inequality (2) that $f(n_1)+\dots+f(n_r) \leq f(n_1+\dots+n_r+6(r-1))$. Now if any integer n is presented as a sum of positive integers $n_1, \dots, n_r$, then we have $r \leq n$ and $f(n_1) + \dots + f(n_r) \leq f(n + 6(r-1)) \leq f(7n)$. Using (1) we obtain $\bar{f}(n) \leq f(7n)$ for all n. This means that $\bar{f} \preceq f$. Since $f(n) \leq \bar{f}(n)$ for all n, we always have $f \preceq \bar{f}$. Now it is clear that $f \simeq \bar{f}$. So it remains to prove inequality (2).

By definition of f, there exist diagrams Δ_1, Δ_2 over $\langle \Sigma \mid \mathcal{R} \rangle$ with boundary labels W_i that have $f(n_i)$ cells and the length of W_i does not exceed n_i $(i = 1, 2)$. Moreover, no diagram with boundary label W_i can have fewer than $f(n_i)$ cells $(i = 1, 2)$.

Starting with Δ_1, Δ_2, we shall construct a word W and a diagram Δ with boundary label W and the following properties:

- The perimeter of Δ does not exceed $n_1 + n_2 + 6$.
- Δ has exactly $f(n_1) + f(n_2)$ cells.
- No diagram with the same boundary label W can have fewer cells.

Inequality (2) follows from these properties by definition of Dehn functions. So it remains to construct such a diagram Δ.

First of all, we apply Lemma 2 to Δ_1. We find a boundary vertex v_1 of Δ_1 and an integer $i_1 = 1, 2$ such that no initial subpath of the boundary contour of Δ_1 at v_1 has a label equal to a_{i_1} in G. After that, applying Lemma 2 to the mirror copy of Δ_2, we find a boundary vertex v_2 of Δ_2 and an integer $i_2 = 1, 2$ such that no terminal subpath of the boundary contour of Δ_2 at v_2 has a label equal to $a_{i_2}^{-1}$.

Denote by t_i the boundary contour of Δ_i at v_i $(i = 1, 2)$. We distinguish two cases: $i_1 \neq i_2$ and $i_1 = i_2$.

Case 1. For simplicity, let us assume that $i_1 = 1$, $i_2 = 2$. Let us construct a diagram Δ taking Δ_1, Δ_2 and joining them by a "bridge" of two edges e_1, e_2 (see Figure 2).

Here $\varphi(e_i) = a_i$ for $i = 1, 2$ and Δ has a contour $e_1^{-1} t_1 e_1 e_2^{-1} t_2 e_2$ at $w = \tau(e_1) = \tau(e_2)$. It is clear from our construction that Δ has exactly $f(n_1) + f(n_2)$ cells, and its perimeter does not exceed $n_1 + n_2 + 4$. We shall prove that no diagram with the same boundary label has fewer cells.

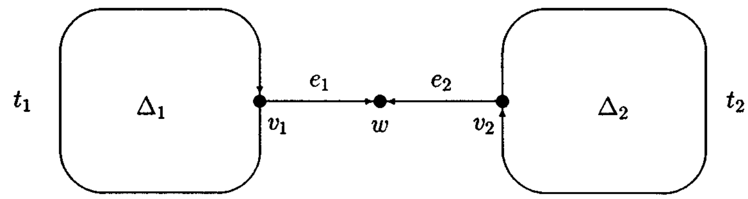

FIGURE 2

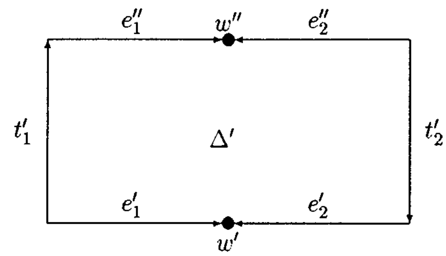

FIGURE 3

Suppose that a diagram Δ' has the same boundary label as Δ. Thus its contour can be represented as $(e_1')^{-1}t_1'e_1''(e_2'')^{-1}t_2'e_2'$ where $W_i = \varphi(t_i) = \varphi(t_i')$, $\varphi(e_i) = \varphi(e_i') = \varphi(e_i'') = a_i$ $(i = 1, 2)$. (See Figure 3.)

Let $w' = \tau(e_1') = \tau(e_2')$, $w'' = \tau(e_1'') = \tau(e_2'')$. Applying Lemma 1 we conclude that w' is a double vertex of Δ'. Thus w' coincides with some other boundary vertex.

Suppose that $w' = w''$. In this case we can consider diagrams Δ_1' and Δ_2' with contours $(e_1')^{-1}t_1'e_1''$ and $(e_2'')^{-1}t_2'e_2'$ respectively. In each of these two diagrams we can glue edges e_i' and e_i'' (they have the same label) obtaining two diagrams Δ_i'' $(i = 1, 2)$. Notice that when gluing these edges we may need to cut off some spherical subdiagrams. As a result, we get two diagrams that have boundary labels W_1 and W_2. By the remark above, Δ_i'' has at least $f(n_i)$ cells $(i = 1, 2)$. Therefore, Δ' has at least $f(n_1) + f(n_2)$ cells as desired.

Now suppose that $w' \neq w''$. Thus w' coincides with one of the vertices that belongs to either t_1' or t_2'. In the first case we can find a subdiagram in Δ' with contour $e_1't_1''$ where t_1'' is an initial subpath of t_1'. Thus $\varphi(t_1'') = \varphi(e_1') = a_1$ in G. This is a contradiction since t_1 has an initial subpath whose label is equal to a_1 in G. In the second case we have a subdiagram in Δ' with boundary contour $t_2''e_2'$ where t_2'' is a terminal subpath of t_2'. This implies that $\varphi(t_2'') = \varphi(e_2')^{-1} = a_2^{-1}$ in G. We have also a contradiction since t_2 cannot have a terminal subpath whose label is equal to a_2^{-1} in G.

Case 2. Now without loss of generality assume that $i_1 = i_2 = 1$. We also construct a diagram Δ taking Δ_1, Δ_2 and joining them by a "bridge" of three edges e_1, e, e_2 (see Figure 4).

Here $\varphi(e_1) = \varphi(e_2) = a_1$, $\varphi(e) = a_2$ and Δ has a contour $e_1^{-1}t_1e_1ee_2^{-1}t_2e_2e^{-1}$ at $w = \tau(e_1) = \iota(e)$. It is clear from our construction that Δ has exactly $f(n_1)+f(n_2)$ cells, and its perimeter does not exceed n_1+n_2+6. We shall prove that no diagram with the same boundary label can have fewer cells.

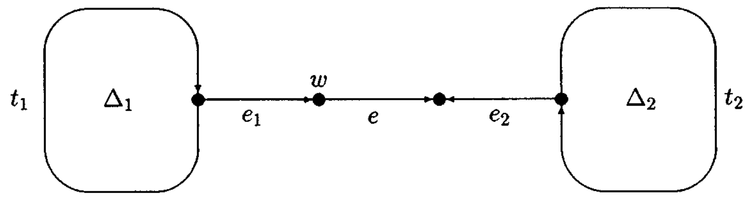

FIGURE 4

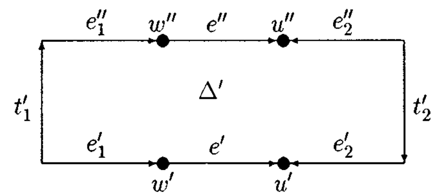

FIGURE 5

Suppose that a diagram Δ' has the same boundary label. Thus its contour can be represented as $(e'_1)^{-1}t'_1e''_1e''(e''_2)^{-1}t'_2e'_2(e')^{-1}$ where $W_i = \varphi(t_i) = \varphi(t'_i)$, $\varphi(e_i) = \varphi(e'_i) = \varphi(e''_i) = a_1$ $(i = 1, 2)$, $\varphi(e) = \varphi(e') = \varphi(e'') = a_2$. (See Figure 5.)

Let $w' = \tau(e'_1) = \iota(e')$, $w'' = \tau(e''_1) = \iota(e'')$, $u' = \tau(e') = \tau(e'_2)$, $u'' = \tau(e'') = \tau(e''_2)$. Applying Lemma 1 we conclude that w' is a double vertex of Δ'. Thus w' coincides with some other boundary vertex.

Suppose that $w' = w''$. In this case we have two subdiagrams in Δ' with contours $(e'_1)^{-1}t'_1e''_1$ and $e''(e''_2)^{-1}t'_2e'_2(e')^{-1}$. Gluing e'_1 and e''_1, e' and e'', and e'_2 and e''_2 in these subdiagrams, and cutting off spherical subdiagrams if necessary, we obtain diagrams Δ''_i with boundary labels W_i $(i = 1, 2)$. These diagrams have at least $f(n_1)$, $f(n_2)$ cells respectively, so Δ' has at least $f(n_1) + f(n_2)$ cells, as desired.

If w' coincides with a vertex in t'_1, we finish the proof in a similar way as in Case 1. Note that w' does not coincide with u' since otherwise $a_2 = \varphi(e') = 1$ in G. Therefore w' coincides with a vertex in $(e''_2)^{-1}t'_2$. By Lemma 1, u' is also a double vertex. If w' coincides with a vertex in t'_2, then u' also coincides with a vertex in t'_2. If $w' = u''$, then u' cannot coincide with u'' so it coincides with a vertex in t'_2 anyway. Now we finish the proof similarly to Case 1.

The Theorem is proved.

REFERENCES

1. G. Baumslag, C. F. Miller III, and H. Short. Isoperimetric inequalities and the homology of groups. *Invent. Math.*, 113:531–560, 1993. MR **95b:**57004
2. M.V. Sapir, J.-C. Birget, E.Rips. Dehn functions of groups, to appear.
3. S. G. Brick, Dehn functions and products of groups. *Trans. Amer. Math. Soc.*, 335 (1993),369–384. MR **93c:**57003
4. S. Gersten. Isoperimetric and isodiametric functions of finite presentations. In *Geometric group theory*, Vol. 1 (Sussex, 1991), London Math. Soc. Lecture Note Ser., 181 (1993), 79–96. MR **94f:**20066
5. M. Gromov. Hyperbolic groups. In *Essays in Group Theory* (S. Gersten, ed.), MSRI Publ. 8. Springer-Verlag, 1987, 75–263. MR **89e:**20070

6. R. C. Lyndon and P. E. Schupp. *Combinatorial group theory*. Springer-Verlag, 1977. MR **58:**28182
7. K. Madlener, F. Otto. Pseudo-natural algorithms for the word problem for finitely presented monoids and groups. *J. Symbolic Computation*, 1:383–418, 1985. MR **87h:**03067

Department of Mathematics, Vologda State Pedagogical University, S. Orlov St., 6, Vologda 160600, Russia
E-mail address: guba@vgpi.vologda.su

Department of Mathematics, Vanderbilt University, Nashville, Tennessee 37240
E-mail address: msapir@math.vanderbilt.edu