ON DEHN FUNCTIONS OF AMALGAMATIONS AND STRONGLY UNDISTORTED SUBGROUPS

Stephen G. Brick and Jon M. Corson

May 31, 1999

Abstract. We study the Dehn functions of amalgamations, introducing the notion of strongly undistorted subgroups. Using this, we give conditions under which taking an amalgamation does not increase the Dehn function, generalizing one aspect of the combination theorem of Bestvina and Feighn.

To obtain examples of strongly undistorted subgroups, we define and study the relative Dehn function of pairs of groups. As a result we obtain a new method of constructing examples of pairs of groups that are relatively hyperbolic in the sense of Farb.

0. Introduction.

In [BF], M. Bestvina and M. Feighn prove a combination theorem for hyperbolic groups. Among other consequences, their theorem gives conditions under which an amalgamation $P = A *_C B$ of hyperbolic groups A and B is itself hyperbolic. Other results in the same vein can be found in [Gi], [KM], and [Ge2].

As is well-known, hyperbolic groups are characterized by having linear Dehn functions. Therefore the combination theorem can be viewed as giving a condition for groups with linear Dehn functions under which taking an amalgamation does not increase the Dehn function. It thus seems appropriate to try to generalize the Bestvina-Feighn result by looking for a condition on groups – which are possibly not hyperbolic – under which taking an amalgamation $P = A *_C B$ does not result in an increase of the Dehn function, i.e, the Dehn function δ_P of $P = A *_C B$ is bounded above by the maximum of the Dehn function δ_A of A and the Dehn function δ_B of B (though strictly speaking we have to take the subnegative closure of the maximum).

To that end we introduce the notion of *strongly undistorted* subgroups. This definition was inspired by the approach of [KM]. The condition on subgroups is a geometric one. Recall that a subgroup H of a group G is undistorted if there are finite generating sets for H and G so that there is a constant a for which

1991 *Mathematics Subject Classification.* Primary 20F32.

Key words and phrases. Dehn function, word problem, quasiconvex, finitely presented group, amalgamation, pairs of groups, subgroup distortion, strongly undistorted subgroups.

$|w|_H \leq a \cdot |w|_G$ for all $w \in H$ (where $|*|_H$ and $|*|_G$ denotes the length function with respect to the appropriate generating set). This can be viewed as saying something about the boundary of a bigon in the Cayley graph, where one side of the bigon represents w in the generators of H and the other side represents w in the generators of G. Our condition of being strongly undistorted is a condition not just for bigons, but more generally for $2n$-gons for any n. (See below for details.)

We prove that if C is strongly undistorted in A and undistorted in B (we only need strong undistortion in one factor), then the Dehn function δ_P of $P = A *_C B$ is bounded above by the subnegative closure of the maximum of the Dehn function δ_A of A and the Dehn function δ_B of B.

To obtain examples of strongly undistorted subgroups we define the *relative Dehn function* of pairs of groups. The basic idea is to obtain from a given pair of groups a complex of groups and then apply techniques and results from [BC1] on the Howie functions of complexes of groups. It turns out that when the relative Dehn function is linear then the subgroup is strongly undistorted. Furthermore, a group pair with linear relative Dehn function is necessarily relatively hyperbolic in the sense of Farb (see [Fa2]). And if the complex of groups is conformly hyperbolic, a result from [Co1] implies that the relative Dehn function is linear.

We also study the relative Dehn function in general, showing it does not depend on the presentation chosen. And we obtain a bound for δ_P when C is not strongly undistorted. In that case, the bound involves the relative Dehn function.

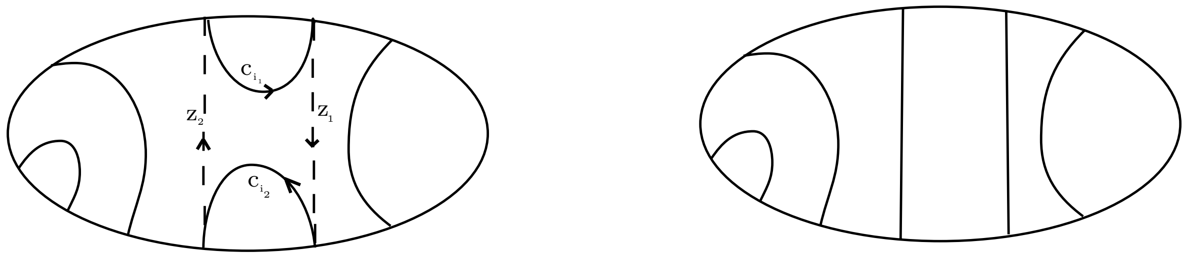
The layout of the paper is as follows. In section 1, we use a transversality approach to give an overview of our approach. In section 2 we define the notion of strongly undistorted subgroups. We prove our main theorem on the Dehn function of an amalgamation. In section 3 we turn to the construction of examples by defining the relative Dehn function of pairs of groups. We prove an invariance theorem for the relative Dehn function, showing that it is independent of the choice of relative presentation. In section 4, we turn to the hyperbolic case, showing that a linear relative Dehn function implies relative hyperbolicity in the sense of Farb as well as strong undistortion for the subgroup. Further we show that a conformal hyperbolic structure is sufficient to ensure a linear relative Dehn function. We close the paper with some remarks in section 5.

Unless otherwise specified, all two-complexes are assumed to be combinatorial and connected. Given an edge path w we let $|w|$ denotes the length of the edge path. And if $h : \mathbb{Z}_+ \to \mathbb{Z}_+$ is a function, then $\overline{h}$ denotes the subnegative (or superadditive) closure, i.e., the least function H greater than h for which $H(n)+H(m) \leq H(n+m)$ for all n, m (see [Br]).

1. Sketch using transversality.

Let $P = A *_C B$ be an amalgamation of finitely presented groups. In order to motivate the definitions and arguments that we make later on in the paper we sketch an approach using transversality.

Use presentations of A, B and C to construct complexes K_A and K_B with fundamental groups A and B respectively, such that $K = K_A \cup K_B$ has fundamental group P and $K_C = K_A \cap K_B$ has fundamental group mapping onto C. Moreover,



before the bridge move

after the bridge move

FIGURE 1

arrange things so that K_C is two-sided in K.

Given a word w representing the identity in P, choose a disk diagram for w, i.e., a map of a disk $f : D^2 \to K$ with boundary label being w. By transversality, we may assume that $f^{-1}(K_C)$ is a properly embedded collection of arcs and simple closed curves. The arcs cut the disk D^2 up into subdisks D_i. The interior of any simple-closed curve component may be redefined so as to map entirely into either K_A or entirely into K_B. Thus we may assume that each subdisk D_i is likewise mapped either entirely into K_A or entirely into K_B.

Our goal is to bound the area of D. Clearly it suffices to bound the area of each subdisk D_i. Consider one of the subdisks D_i. Suppose it maps entirely into K_A (a similar discussion holds for the other case). The boundary of D_i is of the form

$$\alpha_1 c_1 \alpha_2 c_2 \cdots \alpha_n c_n$$

where each c_i is a path in $f^{-1}(K_C)$ and each α_i is a path mapping to K_A. Call the c_i's the *subgroup boundary paths*. We define the complexity of the disk D to be the sum of the lengths of all of the subgroup boundary paths for all of the subdisks.

The key point is that we are free to change the map of the disk D and the subdisks D_i as long as the boundary of D is mapped to the word w. Thus we may as well assume that things have been chosen so that the complexity of the disk D is minimal.

There is a situation where the complexity of the disk can be decreased. This happens when a bridge move, with respect to $f^{-1}(K_C)$, decreases the length of the subgroup boundary paths. One such bridge move is pictured in figure 1. Note that there are two paths z_1 and z_2, both of which map to K_C, where z_1 starts at the terminal point of some subgroup boundary path, call it c_{i_1}, and ends at the initial point of some c_{i_2}, while z_2 starts at the terminal point of c_{i_2}, and ends at the initial point c_{i_1} with

$$|z_1| + |z_2| < |c_{i_1}| + |c_{i_2}|.$$

We may alter the map so that the only change to the subgroup boundary paths is that c_{i_1} and c_{i_2} are replaced by z_1 and z_2. This is possible because the closed circuit $c_{i_1} z_1 c_{i_2} z_2$ is an inessential loop in K_C and may be pushed off of K_C by two-sidedness.

We must also allow for the possibility of bridge moves involving more subgroup boundary paths, i.e., ones involving more than two such subgroup boundary paths, say $c_{i_1}, \dots c_{i_m}$ and new paths $z_1, \dots z_m$, where the c_{i_j}'s will be replaced by the z_i's. It is possibilities such as these that motivate our definitions below.

2. Strongly undistorted subgroups.

Suppose G is a finitely generated group and H is a finitely generated subgroup. Fix finite generating sets X and Z for G and H respectively such that $Z \subseteq X$. Denote by Γ the Cayley graph of G with respect to X, and write Γ_Z for the full subgraph of Γ consisting of the edges with labels in Z.

Definition. An H-relative n-gon (or n-gon rel H) is a edge-circuit w in Γ of the form

$$(*) \qquad w = \alpha_1 c_1 \alpha_2 c_2 \cdots \alpha_n c_n$$

where each α_i is a nontrivial possibly non-geodesic edge path with no edges in Γ_Z, and each c_i is a geodesic edge path in Γ_Z.

Note the slight abuse of notation in that a relative n-gon consists of $2n$ subpaths. However $(*)$ is a "lift" of a possibly non-geodesic n-gon in the coset graph Γ/H, i.e., the paths $\alpha_1, \alpha_2, \dots, \alpha_n$ project to an n-gon in Γ/H.

Definition. A relative n-gon w, as described in $(*)$, is *splittable* if there exists integers $1 \le i_1 < i_2 < \cdots < i_m \le n$ and geodesic edge paths $z_1, \dots, z_m$ in Γ_Z such that

$$(**) \qquad |z_1| + \cdots + |z_m| < |c_{i_1}| + \cdots + |c_{i_m}|$$

where z_j goes from terminal vertex of c_{i_j} to the initial vertex of $c_{i_{(j+1)}}$ (double indices reduced modulo m).

The case of a splittable relative n-gon with $m = 2$ is illustrated in figure 1 of the previous section. Note that relative n-gon w may be "cut" along the circuit $\rho = c_{i_1} z_1 c_{i_2} \cdots c_{i_m} z_m$ to obtain relative n'-gons, where $n' = i_{j+1} - i_j$, of the form

$$w_j = \alpha_{i_j+1} c_{i_j+1} \cdots c_{(i_{(j+1)}-1)} \alpha_{i_{(j+1)}} (z_j)^{-1}.$$

As described in the previous section, such a cut would occur for a map of a disk into an amalgam of complexes. Moreover, there would be a joining of other relative n''-gons along u, which would map into the "other side" of the amalgam.

Observe that a relative 1-gon, i.e., an edge circuit of the form αc, with α an edge path with no Z-labels and c a geodesic in Γ_Z, is vacuously nonsplittable.

Definition. The subgroup H is *strongly undistorted* in G if there exists a constant $k \geq 1$ and a finite generating set X that contains a subset $Z \subseteq X$ which generates H so that for every $n \geq 1$ and every nonsplittable relative n-gon w, as described above, satisfies

$$|c_i| \leq k \cdot (|\alpha_1| + |\alpha_2| + \cdots + |\alpha_n|) \tag{1}$$

for each $i = 1, \ldots, n$.

A priori, this condition may depend on the generating sets chosen. So to be perfectly accurate, we should speak of being strongly undistorted with respect to a nested pair of generating sets.

There is one case where independence of generating sets can be shown. If (X_1, Z) and (X_2, Z) generate the group pair (G, H) and, if for each i, any element of H may be written as a word in $X_i \setminus Z$, then H is strongly undistorted in G with respect to (X_1, Z) if and only if H is strongly undistorted in G with respect to (X_2, Z). We leave the details of this to the interested reader.

It is a lemma of [KM] that any finitely generated malnormal quasiconvex subgroup of a hyperbolic group is strongly undistorted with respect to any nested pair of generating sets (recall that for finitely generated subgroups of hyperbolic groups being quasiconvex is equivalent to being undistorted). This is the key step in their proof that an amalgamation of hyperbolic groups along such a subgroup is itself hyperbolic.

The choice of the term "strongly undistorted" begs the question as to whether a strongly undistorted subgroup is necessarily undistorted. The condition of being undistorted can be formulated in a way analogous to formula (1) in the case of $n = 1$.

Definition. A *pseudo-bigon* is an edge circuit w in Γ of the form $w = ac$, where c is a geodesic edge path in Γ_Z and a is a possibly non-geodesic edge path in Γ (which may contain Z-edges).

Being undistorted is equivalent to there being a constant k so that given any pseudo bigon $w = ac$ we have $|c| \leq k \cdot |a|$. This seems less demanding than formula (1). But the edge path a may contain Z-edges and so u is possibly not a relative 1-gon. And the edge path a may begin or end with Z-edges, so u may not even be a relative n-gon for any n. But even if w were a relative n-gon for some n, it may be splittable, so formula (1) would not directly apply. But the following takes care of these considerations.

Proposition 2.1. *Suppose H is strongly undistorted in G with constant k, with respect to generating sets $Z \subset X$. Given $w = ac$, a pseudo bigon, we have*

$$|c| \leq k \cdot |a|$$

Proof. Decompose $a = c'a'c''$ where c' and c'' are (possibly trivial) edge paths in Γ_Z and a' is an edge path in Γ that does not start nor end with a Z-edge. Find a geodesic edge path v in Γ_Z with the same endpoints as $c''cc'$. Then $u = a'v$ is a pseudo bigon and by the triangle inequality $|c| \leq |c'| + |v| + |c''|$. Suppose we know the result for the pseudo bigon $u = a'v$. Then using $|v| \leq k \cdot |a'|$ and the fact that $k \geq 1$, we get

$$|c| \leq |c'| + |v| + |c''| \leq |c'| + (k \cdot |a'|) + |c''| \leq k \cdot (|c'| + |a'| + |c''|) = k \cdot |a|$$

as desired. So it suffices to prove the result for the pseudo bigon $u = a'v$.

We may write $u = \alpha_1 v_1 \cdots \alpha_n v_n$ where $v = v_n$, each α_i is an edge path in Γ not containing any Z-edges, and each v_i is an edge path in Γ_Z, though possibly not geodesic. We now replace each v_i for by a geodesic edge path c_i in Γ_Z, where we may assume $c_n = v_n = v$, as it was already geodesic in Γ_Z. Doing this results in a relative n-gon $w = \alpha_1 c_1 \alpha_2 c_2 \cdots \alpha_n c_n$. We claim that it is sufficient to prove

$$|c_n| \leq k \cdot (|w| - |c_n|). \tag{2}$$

By construction

$$|w| - |c_n| = |w| - |v| \leq |u| - |v| = |a'|.$$

So if we prove formula (2), we may then use the fact that $c_n = v$, giving us the desired result of $|v| \leq k \cdot |a'|$ for the pseduo bigon $u = a'v$

We will now prove formula (2) by inducting on n. If $n = 1$, then $w = \alpha_1 c_1$ must be nonsplittable. The result then follows from the strongly undistorted property using formula (1) above.

Now assume the result is true for all relative n'-gons for $n' < n$. If w is nonsplittable, then the result follows using formula (1), as in that case

$$\sum |\alpha_i| = |w| - \sum |c_i| \leq |w| - |c_n|$$

If w is splittable, then there are integers $1 \leq i_1 < i_2 < \cdots < i_m \leq n$ and geodesic edge paths $z_1, \ldots z_m$ in Γ_Z such that formula $(**)$ holds. Consider first the case of $i_m = n$. As described above, the edge circuit $\rho = c_{i_1} z_1 c_{i_2} \cdots c_{i_m} z_m$ cuts w up into relative n'-gons,

$$w_j = \alpha_{i_j+1} c_{i_j+1} \cdots c_{(i_{(j+1)}-1)} \alpha_{i_{(j+1)}} (z_j)^{-1},$$

where $n' = i_{(j+1)} - i_j < n$. By induction, formula (2) holds for each w_j. Thus we have

$$|z_j| \leq k \cdot \Big(\sum_{t=i_j+1}^{i_{(j+1)}} |\alpha_t| + \sum_{t=i_j+1}^{i_{(j+1)}-1} |c_t| \Big) \tag{3}$$

for each j. Note that no term of the form $|c_{i_j}|$ appears in of these sums. Also, using the edge circuit ρ and the triangle inequality, we have

$$|c_n| = |c_{i_m}| \leq \sum_{j=1}^{m-1} |c_{i_j}| + \sum_{j=1}^{m} |z_j| \tag{4}$$

Putting inequalities (3) and (4) together yields

$$|c_n| \leq \sum_{j=1}^{m-1} |c_{i_j}| + k \cdot (\sum_{j=1}^{m} \Big(\sum_{t=i_j+1}^{i_{(j+1)}} |\alpha_t| + \sum_{t=i_j+1}^{i_{(j+1)}-1} |c_t| \Big)$$

Collecting terms gives us the desired result:

$$|c_n| \leq k \cdot \Big(\sum_{i=1}^{n} |\alpha_i| + \sum_{i=1}^{n-1} |c_i| \Big) = k \cdot (|w| - |c_n|)$$

Finally we are left with the case of $i_m < n$, i.e., c_n is not among the c_{i_j}'s. In that case, c_n is an edge in the relative n'-gon and we have

$$w_m = \alpha_{i_m+1} c_{i_m+1} \cdots c_n \cdots c_{i_1-1} \alpha_{i_1} (z_m)^{-1}$$

Rewriting w_m and applying the inductive hypothesis results in

$$|c_n| \leq k \cdot \Big(\sum_{t=i_m+1}^{n-1} |c_t| + \sum_{t=i_m+1}^{n} |\alpha_t| + \sum_{t=1}^{i_1} |\alpha_t| + |z_m| \Big) \tag{5}$$

Using the triangle inequality applied to the cycle $\rho = c_{i_1} z_1 c_{i_2} \cdots c_{i_m} z_m$ we have

$$|z_m| < |c_{i_1}| + \cdots + |c_{i_m}| + |z_1| + \cdots + |z_{m-1}|$$

We may make use of formula (3) above to bound each of the $|z_i|$ for $i < m$. Substituting these terms and rewriting the sums, formula (5) becomes:

$$|c_n| \leq k \cdot \Big(\sum_{i=1}^{n} |\alpha_i| + \sum_{i=1}^{n-1} |c_i| \Big) = k \cdot (|w| - |c_n|)$$

as desired. And by induction we are done. □

As noted above, this implies the following:

Corollary 2.2. *If H is strongly undistorted in G then H is undistorted in G.*

We now turn to the result that is the main reason for considering strongly undistorted subgroups:

Theorem 2.3. *Let P be an amalgamation $A *_C B$, where A and B are finitely presented and C is finitely generated. If C is strongly undistorted in A and undistorted in B, then*

$$\delta_P \preccurlyeq \overline{\delta}$$

where $\delta = \max(\delta_A, \delta_B)$.

Proof. Let X and Z be finite generating sets of A and C with $Z \subseteq X$ that satisfy the definition for C to be strongly undistorted in A, say with constant a. Choose a finite generating set Y for B such that $Y \supseteq Z$. Then, since C is undistorted in B, there exists a constant b such that for all $g \in C$, we have $|g|_Z \leq b \cdot |g|_Y$. Choose finite presentations $A = \langle X \mid R \rangle$ and $B = \langle Y \mid S \rangle$, using the generating sets X and Y. Then we have the finite presentation $P = \langle X \cup Y \mid R \cup S \rangle$. We take the Dehn functions of these presentations for our representatives of δ_A, δ_B, and δ_P.

Let w be a word in $(X \cup Y)^{\pm 1}$ such that $w = 1$ in P. We show that $\text{Area}(w) \leq \overline{\delta}(c \cdot |w|)$, where $c = 1 + 2(a + b + ab)$. First note that if w is a word in $X^{\pm 1}$ or in $Y^{\pm 1}$, then $\text{Area}(w) \leq \delta(|w|) \leq \overline{\delta}(c \cdot |w|)$. Having dispensed with this trivial case, we may now assume, after replacing w by a cyclic conjugate if necessary, that $w = u_1 v_1 u_2 v_2 \cdots u_m v_m$, where each u_i is a nontrivial word in $(X \setminus Z)^{\pm 1}$ and each v_j is a nontrivial word in $Y^{\pm 1}$.

We consider certain connected planar graphs for w, one of which can be constructed as follows. Subdivide the boundary of the unit disk D^2 into $2m$ subintervals and label them consecutively by the words $u_1, v_1, \ldots, u_m, v_m$. Since the word w represents 1 in P, some u_i or some v_j represents an element $g \in C$ (the amalgamated subgroup). Let s_1 be a properly embedded arc in D^2 with the same endpoints as the boundary interval labelled by this subword, and label s_1 by a word in $Z^{\pm 1}$ representing g. Continuing in this fashion, we obtain a disjoint collection $s_1, \ldots, s_m$ of properly embedded arcs in D^2, each joining a pair of distinct vertices in our subdivision of ∂D^2 and labelled by a word in $Z^{\pm 1}$. The arcs cut D^2 into $m+1$ subdisks $D_1, \ldots, D_{m+1}$. Reading around the boundary of D_i we either obtain a word in $X^{\pm 1}$ of the form $u_{i_1} w_1 u_{i_2} w_2 \cdots u_{i_k} w_k$ representing 1 in A, or a word in $Y^{\pm 1}$ of the form $v_{i_1} w_1 v_{i_2} w_2 \cdots v_{i_k} w_k$ representing 1 in B, where the w_j are words in $Z^{\pm 1}$ and $1 \leq i_1 < i_2 < \cdots < i_k \leq m$. In the first case we say that D_i is an A-region, whereas in the second case we say that D_i is a B-region.

Among all such labelled planar graphs for w, choose one for which $\sum_{i=1}^{m} |s_i|$ is minimal. (By $|s|$ we mean the length of the label on the arc s.) Then the boundary label of an A-region of the above form $u_{i_1} w_1 u_{i_2} w_2 \cdots u_{i_k} w_k$, viewed as a circuit in the Cayley graph $\Gamma(A, X)$, is a nonsplittable k-gon rel C. Otherwise, we could perform a bridge move as described above that would reduce the sum of the lengths of words on the properly embedded arcs.

Next observe that there is a one-to-one function

$$D : \{s_1, \ldots, s_m\} \to \{D_1, \ldots, D_{m+1}\}$$

such that $s_i \subset \partial D(s_i)$, for all i. To see this, choose an inner-most arc; say it is s_1. Then the closure of one component of $D^2 \setminus s_1$ is one of the subdisks, D_i say. Define $D(s_1) = D_i$, and continue by induction.

Renumber the arcs $s_1, \ldots, s_m$ so that $D(s_i)$ is an A-region for $1 \leq i \leq l$, and $D(s_i)$ is a B-region for $l+1 \leq i \leq m$. Then for $1 \leq i \leq l$, the boundary of $D(s_i)$

can be viewed as a nonsplittable k-gon in $\Gamma(A, X)$ rel C, as noted above. Thus $|s_i| \leq a \cdot |\partial D(s_i) \cap \partial D^2|$ and it follows that

$$\sum_{i=1}^{l} |s_i| \leq a \sum_{i=1}^{l} |\partial D(s_i) \cap \partial D^2| \leq a \cdot |w|$$

For $l+1 \leq i \leq m$, the boundary of $D(s_i)$ is labelled by a word in $Y^{\pm 1}$ of the form uv, where u is the label on the arc s_i. Thus, u and v^{-1} represent the same element of B and, by the minimal nature of our labelled graph, u is a shortest word in $Z^{\pm 1}$ representing this element. So $|u| \leq b \cdot |v| = b \cdot |\partial D(s_i) \setminus s_i|$, where b is the linear distortion constant for C in B that we chose above. Furthermore, $\partial D(s_i) \setminus s_i$ consists only of arcs on ∂D^2 and arcs s_j for $j \leq l$, as the neighboring regions of a B-region are all A-regions. It follows that

$$\begin{aligned}\sum_{i=l+1}^{m} |s_i| &\leq b \sum_{i=l+1}^{m} |\partial D(s_i) \setminus s_i| \\ &\leq b\Big(\sum_{i=l+1}^{m} |\partial D(s_i) \cap \partial D^2| + \sum_{i=1}^{l} |s_i|\Big) \\ &\leq b(|w| + a \cdot |w|)\end{aligned}$$

Putting these two bounds together, we have that $\sum_{i=1}^{m} |s_i| \leq (a+b+ab)|w|$.

We can now bound the area of w as follows. The boundary of each subdisk D_i is labelled by a word in either $X^{\pm 1}$ or $Y^{\pm 1}$, and thus bounds a van Kampen diagram with area $\leq \delta(|\partial D_i|)$. Gluing these van Kampen diagrams together in the obvious fashion produces a van Kampen diagram for w, and hence by subnegativity,

$$\begin{aligned}\text{Area}(w) &\leq \sum_{i=1}^{m+1} \delta(|\partial D_i|) \\ &\leq \overline{\delta}\Big(\sum_{i=1}^{m+1} |\partial D_i|\Big) \\ &\leq \overline{\delta}\Big(|\partial D^2| + 2 \sum_{i=1}^{m} |s_i|\Big) \\ &\leq \overline{\delta}(c \cdot |w|)\end{aligned}$$

where $c = 1 + 2(a+b+ab)$.

We can therefore conclude that $\delta_P \preccurlyeq \overline{\delta}$, as required. □

3. Relative Dehn functions.

Let (L, K) be a pair of (connected, combinatorial) two-complexes with finite one-skeletons such that there are only finitely many two-cells in $L \setminus K$, none of which have their entire boundary in K. Then we can form the combinatorial quotient $Q = L/K$; see [Ge1]. Choose a vertex in K as base point. Let $p : \widetilde{L} \to L$ be the universal covering projection. Construct a two-complex $E = E(L, K)$ by identifying each component of $p^{-1}(K)$ to a point. (We shall always assume that the inclusion map induces a monomorphism of fundamental groups $\pi_1(K) \to \pi_1(L)$, in which case every component of $p^{-1}(K)$ is a copy of the universal cover of K.) Then E inherits an action of $\pi_1(L)$ from the covering action on $\widetilde{L}$ and there is a natural projection map $E \to Q$ forming a commutative diagram

$$\begin{array}{ccc} \widetilde{L} & \longrightarrow & E \\ \downarrow & & \downarrow \\ L & \longrightarrow & Q \end{array}$$

where the vertical arrows are the orbit maps of the $\pi_1(L)$ actions. Notice that the map $\widetilde{L} \to E$ induces an epimorphism of fundamental groups, and thus E is simply connected. (The projection $L \to Q$ is a complex of spaces and E is the universal cover of Q relative to the associated complex of groups; see [Co2].)

Definition. The *relative Dehn function of the pair* (L, K), denoted $\delta_{L,K}$, is the Dehn function of $E(L, K)$. (Note that since $E(L, K)$ may not be finite, the Dehn function of the pair may take on the value ∞.)

The relative Dehn function is a special case of the Howie function for a complex of groups as studied in [BC1].

Let (G, H) be a pair of finitely generated groups such that G is *finitely presented relative to* H. By this we mean that there exist finitely generated presentations $G = \langle X \mid R\rangle$ and $H = \langle Z \mid S\rangle$ such that $Z \subseteq X$, $S \subseteq R$, and $R \setminus S$ is finite.

Definition. A *geometric realization* of (G, H) is a pair of two-complexes (L, K) with finite one-skeletons and with only a finite number of cells in $L\setminus K$ such that, for any choice of base point in K, there exist isomorphisms $\pi_1(L) \to G$ and $\pi_1(K) \to H$ for which the following diagram is commutative:

$$\begin{array}{ccc} \pi_1(L) & \longrightarrow & G \\ \uparrow & & \uparrow \\ \pi_1(K) & \longrightarrow & H \end{array}$$

(The vertical maps are inclusion or induced by inclusion.)

It should be noted that every pair of relatively finitely presented groups (G, H) has a geometric realization. Simply choose finitely generated presentations $\langle X \mid R\rangle$ and $\langle Z \mid S\rangle$ for G and H, respectively, such that $Z \subseteq X$, $S \subseteq R$, and $R \setminus S$ is finite. Let L be the two-complex canonically associated to the presentation of G, and let K be the subcomplex corresponding to the presentation of H. Then (L, K) is a geometric realization of the pair (G, H).

For any geometric realization (L, K) of the pair (G, H), we may assume (as we do tacitly throughout) that no two-cell in $L \setminus K$ has its boundary entirely in K. If such a two-cell were to exist, we could simply remove its interior, having no effect on the fundamental groups of L and K. In this way, we guarantee that the combinatorial quotient, and thus relative Dehn function, of a geometric realization always exists. And it is natural to wonder if the relative Dehn functions of different geometric realizations of the same pair of groups are related in some way. We next give an answer to this question.

As above, let (G, H) denote a pair of finitely generated groups such that G is finitely presented relative to H.

Theorem 3.1. *If (L_1, K_1) and (L_2, K_2) are geometric realizations of (G, H), then*

$$\delta_{L_1,K_1} \equiv \delta_{L_2,K_2}$$

Proof. For $i = 1, 2$, select a vertex in K_i as base point and an isomorphism $h_i : \pi_1(L_i) \to G$ such that $h_i|_{\pi_1(K_i)}$ is an isomorphism onto H. Also choose a maximal tree T_i in L_i^1 such that $T_i \cap K_i$ is a maximal tree in K_i^1. Then for each directed edge e of L_i, put $\gamma(e) = set^{-1}$, where s and t are the unique reduced edge paths in T_i from the base point to the initial and terminal vertices of e, respectively. Thus $\gamma(e)$ represents an element of $\pi_1(L_i)$.

Define a map $f : L_1 \to L_2$ inducing an isomorphism of fundamental groups such that $h_2 \circ f_* = h_1$, as follows. Initially, we let f map all of T_1 to the base point in L_2. Then for each edge e of $L_1^1 \setminus T_1$, define f so that $f(e)$ is a closed edge path (at the base point) such that $\gamma(e)$ and $f(e)$ represent the same element of G, under the identifications $h_i : \pi_1(L_i) \to G$. Furthermore, if e is in K_1, we require that $f(e)$ be in K_2. This defines f on the one-skeleton of L_1. Then take any extension to a map of L_1 to L_2 that sends the subcomplex K_1 to K_2; there is no obstruction to doing so.

Form the combinatorial quotient spaces $Q_i = L_i/K_i$. Since $f(K_1) \subseteq K_2$, f determines a map $\overline{f} : Q_1 \to Q_2$. Set $E_i = E(L_i, K_i)$ and let $q_i : E_i \to Q_i$ denote the projection map.

Choose a base point in the universal cover $\widetilde{L}_i$ sitting over the base point of L_i, and let $f' : \widetilde{L}_1 \to \widetilde{L}_2$ be the base point preserving map covering f. By the way f is constructed, f' maps each lift of K_1 to a lift of K_2, and hence determines a map $\overline{f'} : E_1 \to E_2$.

In the same fashion, define a map $g : L_2 \to L_1$ and obtain maps $\overline{g} : Q_2 \to Q_1$, $g' : \widetilde{L}_2 \to \widetilde{L}_1$ and $\overline{g'} : E_2 \to E_1$.

Now we define three constants. Note first that for every directed edge e of L_1, the edge-loops $g(f(e))$ and $\gamma(e)$ are homotopic (rel base point).

$$a = \max\{\mathrm{Area}_{Q_1}(\overline{g}(\delta)) \mid \delta \text{ is the boundary cycle of a two-cell in } Q_2\}$$

$$b = \max\{|\overline{f}(e)| \mid e \text{ is a directed edge in } Q_1\}$$

$$c = \max\{\mathrm{Area}_{L_1}(g(f(e))\gamma(e)^{-1}) \mid e \text{ is a directed edge in } L_1\}$$

We claim that for every edge-loop w in E_1 with $|w| \le n$,

$$\mathrm{Area}_{E_1}(w) \le a \cdot \delta_{E_2}(b \cdot n) + c \cdot n$$

To this end, we make a few observations. Since $T_i \cap K_i$ is a subtree of T_i, it follows that T_i projects to a tree $\overline{T}_i$ in Q_i. Furthermore, the preimage of T_i in $\widetilde{L}_i$ is a disjoint union of copies of T_i, each of which projects to a tree in E_i. Thus $q_i^{-1}(\overline{T}_i)$ is a disjoint union of copies of $\overline{T}_i$ in E_i, and for every vertex v in E_1, $\overline{g'}(\overline{f'}(v))$ lies in the same component of $q_1^{-1}(\overline{T}_1)$ as v.

(1) For every directed edge e in E_1, $|\overline{f'}(e)| \le b \cdot n$.
(2) For every vertex v in E_1, there is a unique reduced edge path $t(v)$ in $q^{-1}(\overline{T}_1)$ from v to $\overline{g'}(\overline{f'}(v))$.
(3) If e is a directed edge in E_1 with initial vertex v_0 and terminal vertex v_1, then $\mathrm{Area}_{E_1}(t(v_0)\overline{g'}(\overline{f'}(e))t(v_1)^{-1}e^{-1}) \le c$
(4) If δ is a boundary cycle of a two-cell of E_2, then $\mathrm{Area}_{E_1}(\overline{g'}(\delta)) \le a$

Suppose $w = e_1 \cdots e_m$, where each e_i is a directed edge of E_1 $(m \le n)$. Then $\overline{f'}(w) = \overline{f'}(e_1) \cdots \overline{f'}(e_m)$ is an edge-loop in E_2 with $|\overline{f'}(w)| \le b \cdot n$. Hence, $\mathrm{Area}_{E_1}(\overline{g'}(\overline{f'}(w))) \le a \cdot \mathrm{Area}_{E_2}(\overline{f'}(w)) \le a \cdot \delta_{E_2}(b \cdot n)$. Denoting the terminal vertex of e_i by v_i, we can construct a van Kampen diagram for w in E_1 from one for $\overline{g'}(\overline{f'}(w))$. The claim follows.

As a consequence of the claim above, we have that $\delta_{E_1} \preccurlyeq \delta_{E_2}$. However, by symmetric reasoning, $\delta_{E_2} \preccurlyeq \delta_{E_1}$ and our result follows. □

By virtue of the previous theorem, we can speak of the relative Dehn function of a pair of groups (G, H) which we write as $\delta_{G,H}$. It is worth noting that theorem 3.1 in [BC1] implies that $\delta_G \preccurlyeq \overline{\delta}_H \circ \overline{\delta}_{G,H}$.

4. Relative Hyperbolic pairs.

We turn to pairs of groups with linear relative Dehn functions. Recall the definition (Farb) of relatively hyperbolic groups: Let G be a finitely generated group and let H be a finitely generated subgroup. For any Cayley graph Γ of G form the quotient graph $\widehat{\Gamma}$ of Γ by identifying, for each $g \in G$, all vertices of Γ corresponding to elements lying in the left coset gH. We say that the pair (G, H) is *relatively hyperbolic* or that the group G is *hyperbolic relative to* H if $\widehat{\Gamma}$ is a hyperbolic metric space. Here a graph is made into a metric space in the usual way, so that each edge has length one.

Theorem 4.1. *Let (G, H) be a pair of finitely generated groups such that G is finitely presented relative to H. If $\delta_{G,H}$ is linear, then (G, H) is relatively hyperbolic.*

Proof. Let (L, K) be a geometric realization of (G, H). Choose a maximal tree T of L^1 that contains a maximal tree of K^1. Choose an orientation of each edge of $L^1 \setminus T$, and let X denote the corresponding set of generators of $G = \pi_1(L)$. Note that the Cayley graph Γ of G relative to X is obtained from the one-skeleton of

$\widetilde{L}$ by collapsing each lift of T (a finite tree) to a point, and thus Γ and $\widetilde{L}^1$ are quasi-isometric graphs. Furthermore, $H = \pi_1(K)$ is generated by the elements of X that lie in K. And, by the way $\widehat{\Gamma}$ and $E = E(L, K)$ are constructed, it follows that $\widehat{\Gamma}$ and E^1 are also quasi-isometric.

Now E is a one-connected two-complex that satisfies a linear isoperimetric inequality, and there is a bound on the number of edges in the attaching maps of two-cells (since we have a combinatorial map $E \to Q$ to a finite two-complex). It follows that the one-skeleton E^1 is a hyperbolic graph (see [Sh]). Thus $\widehat{\Gamma}$ is hyperbolic and G is hyperbolic relative to H. □

Proposition 4.2. *If G is hyperbolic relative to H, then G is finitely presented relative to H.*

Proof. Let X be a finite generating set of G such that some subset $Z \subseteq X$ generates H. Let Γ be the corresponding Cayley graph of G, and form the graph $\widehat{\Gamma}$ described above. Then $\widehat{\Gamma}$ is δ-hyperbolic for some $\delta \geq 0$, so it satisfies the $2(\delta+1)$-fellow traveler property; see [E]. Hence, by attaching a two-cell along every circuit in $\widehat{\Gamma}$ which is labelled by a reduced word in $(X \setminus Z)^{\pm 1}$ of length $\leq 4\delta + 6$, we obtain a simply connected two-complex E with $E^1 = \widehat{\Gamma}$.

Let $w = x_1 \cdots x_k$ be a word in $(X \setminus Z)^{\pm 1}$ such that paths in $\widehat{\Gamma}$ that are labelled by w are closed. Then we can choose reduced words $u_1, u_2, \ldots, u_k$ in $Z^{\pm 1}$ such that $\widetilde{w} = u_1 x_1 u_2 x_1 \cdots u_k x_k$ is the label on a closed path in Γ. Let

$$R_0 = \{\widetilde{w} \mid w \text{ is a word in } (X \setminus Z)^{\pm 1} \text{ and } |w| \leq 4\delta + 6\}$$

and set $R = S \cup R_0$; it is easy to see that $\langle X \mid R \rangle$ is a presentation of G. Since $R \setminus S = R_0$ is finite, it follows that G is finitely presented relative to H. □

It is a simple exercise to show that $(\mathbb{Z}, 2\mathbb{Z})$ is a relatively hyperbolic pair (in the sense of Farb), but the relative Dehn function of this pair is not linear. Thus not every pair of relatively hyperbolic groups has a linear relative Dehn function.

The next result provides a sufficient condition for having a linear relative Dehn function. At the same time it gives a purely combinatorial means of constructing relatively hyperbolic groups.

First we need to recall a definition, due to Gersten [Ge1, Appendix]. We say that a two-complex Q is *conformally hyperbolic* if it is possible to assign "angles" (i.e., nonnegative real numbers) to the corners of Q such that two axioms hold:

(1) For every simple closed curve C in the link of a vertex of P, the sum the angles at the corners corresponding to the edges of C is greater than or equal to 2π.
(2) For each two-cell σ of Q, the sum of angles of the corners of σ is (strictly) less than $(d(\sigma) - 2)\pi$, where $d(\sigma)$ denotes the number of corners of σ.

We call such an assignment of angles a *hyperbolic conformal structure* for Q.

Theorem 4.3. *Suppose a pair of relatively finitely presented groups* (G, H) *has a geometric realization* (L, K) *such that the combinatorial quotient* $Q = L/K$ *is conformally hyperbolic. Then the relative Dehn function of the pair* (G, H) *is linear.*

Proof. Notice that the orbit map $E \to Q$ is a branched covering of two-complexes, in the sense that the restriction to $E \setminus E^0$ is a covering map onto $Q \setminus Q^0$. Thus the link of every vertex of E is immersed into the link of a vertex of Q. It follows that the pullback of a hyperbolic conformal structure for Q is a hyperbolic conformal structure for E with only finitely many "types" of two-cells (since Q is finite). Thus the argument given in ([Co1, Section 5]) applies showing that E satisfies a linear isoperimetric inequality, and the result follows. □

A linear relative Dehn function provides us with examples of strongly undistorted subgroups as illustrated in the following theorem:

Theorem 4.4. *If the relative Dehn function of the pair* (G, H) *is linear, then* H *is strongly undistorted in* G.

Proof. Choose presentations $\langle X \mid R\rangle$ and $\langle Z \mid S\rangle$ for G and H, respectively, where X and Z are finite and such that $Z \subseteq X$, $S \subseteq R$, and $R\backslash S$ is finite. Then (L, K) is a geometric realization of (G, H), where L is the two-complex canonically associated to the presentation of G and K is the subcomplex associated to the subpresentation defining H.

Let $Q = L/K$ and denote the quotient mapping by $p : L \to Q$. Associated to this quotient ("complex of spaces") is a complex of groups: The single vertex of Q is assigned the group H, and the oriented corners of Q are labelled as follows. Each two-cell σ of Q is the image (under p) of a two-cell in L which is attached by an edge circuit $\beta_1 e_1 \beta_2 e_2 \cdots \beta_k e_k$, where each β_i is an edge path in K and each e_i is a directed edge in $L \setminus K$. Thus, $p(e_1)p(e_2)\cdots p(e_k)$ is the attaching circuit of σ in Q and β_i corresponds to a directed corner of σ, to which we assign the element of H represented by the path β_i (for $i = 1, 2, \dots, k$). See [BC1] for more details.

We define two constants. Let $m_1 = \max\{|h|_Z \mid h \text{ is the label on a corner of } Q\}$, and let m_2 be the largest degree (i.e., length of the attaching circuit) of a two-cell in Q.

Note that the one-skeleton of $\widetilde{L}$ is the Cayley graph of G relative to X. So we can view an n-gon rel H as a null-homotopic edge circuit $w = \alpha_1\gamma_1 \cdots \alpha_n\gamma_n$ in L, where each γ_i is an edge path with no edges in K, and each α_i is an edge path in K representing an element $h_i \in H$ and is a shortest such representative. Thus, there exists a Howie diagram (D, j, λ) such that reading the labels on exterior corners and boundary edges (in order around a boundary cycle) gives the "word" $h_1\gamma_1 h_2\gamma_2 \cdots h_n\gamma_n$. Furthermore, we can find such a Howie diagram with area $\leq \delta_{L,K}(|\gamma_1| + \cdots + |\gamma_n|)$, as $\delta_{L,K}$ is the Howie function of this complex of groups, by [BC1, Theorem 4.2]. (We should point out that there is a typesetting misprint in the statement of this theorem in the published version of [BC1]; the undefined symbol $\delta_{\mathcal{G}}$ should be the Howie function, denoted $h_{\mathcal{G}}$ elsewhere in the paper.)

For each $i = 1, \dots, n$, there is an exterior corner of D (say at the vertex v) which is labelled by h_i. Furthermore, if the relative n-gon is nonsplittable, then the sum of the lengths of the exterior corner labels at v (with respect to the length function on H determined by Z) is less than or equal to the sum of the lengths on the interior corner labels at v. But each interior corner label is represented by a word in $Z^{\pm 1}$ of length $\leq m_1$, and D has $\leq m_2 \cdot \mathrm{Area}(D)$ interior corners. Thus, in the nonsplittable case, it follows that $|\alpha_i| = |h_i|_Z \leq m_1 \cdot m_2 \cdot \delta_{L,K}(|\gamma_1| + \cdots + |\gamma_n|)$. However, by hypothesis, $\delta_{L,K}$ is a linear function and hence H is strongly undistorted in G, as required. □

As noted above, the property of being strongly undistorted depends on an appropriate choice of generating sets. However in the case of a linear relative Dehn function the choice does not matter.

Using Theorem 2.3, we immediately get the following result:

Corollary 4.5. *Let P be an amalgamation $A *_C B$, where A and B are finitely presented groups and C is finitely generated. Suppose the relative Dehn function of the pair (A, C) is linear and C is undistorted in B. Set $\delta = \max(\delta_A, \delta_B)$. Then*

$$\delta_P \preccurlyeq \overline{\delta}$$

And of course, restricting ourselves to the hyperbolic case, yields the following version of the Bestvina-Feighn combination theorem:

Corollary 4.6. *Let P be an amalgamation $A *_C B$, where A and B are hyperbolic groups and C is finitely generated. If the relative Dehn function of the pair (A, C) is linear, and if C is undistorted in B, then P is hyperbolic.*

5. Remarks.

To simplify the exposition we have only discussed the case of amalgamated free products. There are, of course, analogous results for HNN extensions. We leave the formulation and proofs to the reader.

Although our aim in this paper was to better understand what types of amalgamations do not increase the Dehn functions of the factors, our techniques apply more generally. The proof of the following is a simple generalization of that of Theorem 2.3 above.

Theorem 5.1. *Let P be an amalgamation $A *_C B$, where A and B are finitely presented groups and C is finitely generated. Then*

$$\delta_P \preccurlyeq \overline{\delta} \circ \overline{h}_{B,C} \circ \overline{\delta}_{A,C}$$

where $h_{B,C}$ is the distortion function for C in B.

These results should also be compared with a result of Bernasconi (see [Be]), where a bound is obtained which uses the distortion function alone, however it doesn't use the distortion functions of the amalgamating subgroup into each of the factor, but rather the distortion function of the amalgamating subgroup into the resulting amalgamation (see [BC2]):

Theorem (Bernasconi). *Let P be an amalgamation $A *_C B$, where A and B are finitely presented groups and C is finitely generated. Then*

$$\delta_P \preccurlyeq L \cdot (\overline{\delta} \circ h_{P,C})$$

where L is a linear function and $h_{P,C}$ is the distortion function for C in $P = A *_C B$.

The results of [KM] were the inspiration for this paper. As was noted above, one of their results can be viewed as showing that any malnormal undistorted subgroup C of a hyperbolic group G is strongly undistorted. There remains the question of whether similar results hold for such subgroups in the non-hyperbolic case.

Another question worth considering is how the property of being strongly undistorted is affected by change of presentation. Is there an invariance result involving quasi-isometries ?

We could also introduce the notion of a strong distortion function by replacing the right-hand side of formula (1) in section 2 by a function of $\sum |\alpha_i|$. Then, the proof of Proposition 2.1 can be adapted to show that the subnegative closure of the strong distortion function is an upper bound for the standard distortion function. And, of course, the proof of Theorem 4.4 can be generalized to show that the relative Dehn function is an upper bound for the strong distortion function.

It should be noted that there is an immediate connection between the relative Dehn function and the generalized word problem. Since a recursive distortion function is equivalent to a solvable generalized word problem [Fa1], it follows that if $\delta_{G,H}$ is a recursive function, then H has solvable generalized word problem in G. The converse however is false; $(\mathbb{Z}, 2\mathbb{Z})$ is a pair of groups with solvable generalized word problem, but its relative Dehn function is not recursive.

References

[Be] A. A. Bernasconi, *On HNN extensions and the complexity of the word problem for one-relator groups*, Ph.D Thesis (unpublished), University of Utah (1994).

[BF] M. Bestvina and M. Feighn, *A combination theorem for negatively curved groups*, J. Diff. Geom. **35** (1992), 85–101.

[Br] S. G. Brick, *On Dehn functions and products of groups*, Trans. Amer. Math. Soc. **335** (1993), 369–384.

[BC1] S. G. Brick and J. M. Corson, *Dehn functions and complexes of groups*, Glasgow Math. J. **40** (1998), 33–46.

[BC2] ———, *A note on the Dehn functions of certain amalgamations*, preprint (1998).

[Co1] J. M. Corson, *Conformally non-spherical two-complexes*, Math. Z. **214** (1993), 511–519.

[Co2] ———, *Howie diagrams and complexes of groups*, Communications in Algebra **23**, no. 14 (1995), 5221–5242.

[E] D. B. A. Epstein et al, *Word processing in groups*, Jones and Bartlett, 1992.

[Fa1] B. Farb, *The extrinsic geometry of subgroups and the generalized word problem*, Proc. London Math. Soc. **68** (1994), 577–593.

[Fa2] ———, *Relatively hyperbolic groups*, Geom. Funct. Anal. **8** (1998), 1-31.

[Ge1] S. Gersten, *Reducible diagrams and equations over groups*, Essays in Group Theory (S. Gersten, eds.), MSRI Publications, Springer-Verlag, 1987, pp. 15–74.

[Ge2] ———, *Distortion and ℓ_1-homology*, preprint (1998).

[Gi] R. Gitik, *On the combination theorem for negatively curved groups*, International J. of Algebra and Computation **7**, no. 2 (1997), 267–276.

[KM] O. Kharlampovich and A. Myasnikov, *Hyperbolic groups and free constructions*, Trans. Amer. Math. Soc. **350**, no. 2 (1998), 571–613.

[Sh] H. Short, (ed.), *Notes on negatively curved groups*, Group Theory from a Geometrical Viewpoint (E. Ghys, A. Haefliger, and A. Verjovsky, eds.), Proc. ICTP Trieste, World Scientific, 1992, pp. 3–64.

Department of Mathematics and Statistics, University of South Alabama, Mobile, AL 36688

E-mail address: `brick@math.usouthal.edu`

Department of Mathematics, University of Alabama, Tuscaloosa, AL 35487-0350

E-mail address: `jcorson@mathdept.as.ua.edu`