Filling radii of finitely presented groups

S. M. Gersten* and T. R. Riley†

Abstract

The filling radius function R of Gromov measures the minimal radii of van Kampen diagrams filling edge-circuits w in the Cayley 2-complex of a finite presentation $\mathcal{P}$. It is known that the Dehn function can be bounded above by a double exponential in R and the length of the loop, and it is an open question whether a single exponential bound suffices. We define the upper filling radius $\overline{\mathrm{R}}(w)$ of w to be the maximal radius of minimal area fillings of w and let $\overline{\mathrm{R}}$ be the corresponding filling function, so $\overline{\mathrm{R}}(n)$ is the maximum of $\overline{\mathrm{R}}(w)$ over all edge-circuits w of length at most n. We show that the Dehn function is bounded above by a single exponential in $\overline{\mathrm{R}}$ and the length of the loop. We give an example of a finite presentation $\mathcal{P}$ where R is linearly bounded but $\overline{\mathrm{R}}$ grows exponentially.

1991 Mathematics Subject Classification: 20F05, 20F32, 57M07

1 Introduction and statement of results

Gromov defined a number of filling invariants for finitely presented groups [15, Ch. 5] in terms of the geometry of van Kampen diagrams. Two of the most important are the filling radius R and the minimal isoperimetric function f_0 (also known as the Dehn function [11]), whose definitions we now recall.

Let $\mathcal{P}$ be a finite presentation for the group G. Let X be the Cayley 2-complex, so X is the universal cover of the 2-complex canonically associated to $\mathcal{P}$ (namely, one vertex, one edge for each generator, and one 2-cell for

*Partially supported by NSF grant DMS-9800158.

†Supported by EPSRC Award No. 98001683 and Corpus Christi College, Oxford.

each defining relator ρ whose attaching map spells out the cyclic word ρ in the 1-skeleton); the 1-skeleton $X^{(1)}$ is the Cayley graph Γ. Let w be an edge-circuit of Γ and let D be a van Kampen diagram[1] filling w [17], so D is the domain of a cellular map $D \to X$ where D is a 1-connected cellular planar 2-complex such that w (read anticlockwise from the base point $\star$) is the associated boundary circuit (*i.e.* w is the attaching map of the single 2-cell at infinity). The 1-skeleton $D^{(1)}$ of D is equipped with a path metric so that each edge is assigned length 1. Let $\ell(w)$ denote the length of the boundary word w. We will use four measurements one can make on D:

- $\mathrm{Area}(D)$ is the *area*, that is, the number of 2-cells in D,
- $\mathrm{R}(D)$ is the *radius*, that is, the maximum distance of a vertex to the boundary in the path metric on $D^{(1)}$,
- $\mathrm{Diam}(D)$ is the *diameter*, that is, the maximum distance (again in the path metric on $D^{(1)}$) of a vertex to the base point $\star$ of D, (note that diameter is closely related to radius: $\mathrm{R}(D) \leq \mathrm{Diam}(D) \leq \mathrm{R}(D)+\ell(w)$),
- $\mathrm{FL}(D)$ is the *filling length*, that is, the minimal bound on the length of boundary circuits in a *complete shelling* of D (*i.e.* the combinatorial notion of a null-homotopy of w to $\star$ across D). See [12] for a detailed discussion of filling length.

We define $\mathrm{Area}(w)$ (resp. $\mathrm{R}(w)$, $\mathrm{Diam}(w)$, $\mathrm{FL}(w)$) to be the minimum value of $\mathrm{Area}(D)$ (resp. $\mathrm{R}(D)$, $\mathrm{Diam}(D)$, $\mathrm{FL}(D)$), where D ranges over van Kampen diagrams filling w.

The four measurements on diagrams lead to definitions of filling functions for the presentation $\mathcal{P}$. Perhaps the most important is the *minimal isoperimetric function* (*a.k.a.* the *Dehn function*) $f_0 : \mathbb{N} \to \mathbb{N}$, defined by

$$f_0(n) = \max\{\mathrm{Area}(w) : w \text{ is an edge-circuit in } \Gamma \text{ with } \ell(w) \leq n\};$$

this takes finite values since there are only finitely many orbits of such w up to the G-action on Γ. More generally, an upper bound $f : \mathbb{N} \to \mathbb{R}$ for the Dehn function f_0 is called an *isoperimetric function* for $\mathcal{P}$. The notion in

[1] It is convenient to think of D as a tree-like arrangement of topological discs where two discs have at most one vertex in common. Some care is needed in discussing van Kampen diagrams because of the presence of 1-dimensional portions, namely, edges which are not incident with any 2-cell of D.

group theory is due to Gromov [14] in analogy with the corresponding notion in differential geometry. It is convenient to extend an isoperimetric function f to the nonnegative reals by defining $f(r) := f([r])$, where $[r]$ denotes, as usual, the integer part of the real number r.

Analogously one can define filling functions for R, Diam and FL. So in particular there is the *filling radius function* $\mathrm{R} : \mathbb{N} \to \mathbb{N}$, given by

$$\mathrm{R}(n) = \max\{\mathrm{R}(w) : w \text{ is an edge-circuit in } \Gamma \text{ with } \ell(w) \leq n\}.$$

D. E. Cohen [6] first proved the *double exponential theorem*, which states that there are constants $A, B > 1$ so that $f_0(n) \leq A^{B^{\mathrm{R}(n)+n}}$ for all n, by making use of an analysis of the complexity of the Nielsen reduction process due to Avenhaus and Madlener [1]. The first author gave a geometrical proof of this result in [10]; his argument has been generalized in several different directions, *cf.* [9], [19].

As soon as the result was proved the question arose whether in fact a single exponential bound sufficed; this is discussed by Gromov in [15, 5.C.]. The problem has remained open for almost a decade. One reason it is a difficult problem is that minimum area and minimum radius van Kampen diagrams may be completely different, even for aspherical presentations. We give an example of this phenomenon in §5 for the presentation

$$\langle x, y, s, t \mid [x, y] = 1,\ txt^{-1} = x^2,\ sys^{-1} = y^2 \rangle.$$

This suggests studying relations between area and radius in the same diagram. This is the program we initiated in [12], where we studied relations between area, diameter, and filling length on the same class of diagrams. We continue this program in this paper in relating area and filling radius in the class of minimal area diagrams; this theme is also pursued in [13] in calculating isoperimetric functions in kernels of homomorphisms to free groups.

We define $\overline{\mathrm{R}}(w)$ (resp. $\underline{\mathrm{R}}(w)$) be the maximum value (resp. minimal value) of $\mathrm{R}(D)$ over all *minimal area* van Kampen diagrams D for an edge circuit w. These exist since it is not difficult to see that there are only finitely many minimal area van Kampen diagrams for w up to the action of G. The corresponding filling functions, the *upper filling radius function* $\overline{\mathrm{R}} : \mathbb{N} \to \mathbb{N}$ and the *lower filling radius function* $\underline{\mathrm{R}} : \mathbb{N} \to \mathbb{N}$, are defined by

$$\begin{aligned}
\overline{\mathrm{R}}(n) &:= \max\{\overline{\mathrm{R}}(w) : w \text{ is an edge-circuit in } \Gamma \text{ with } \ell(w) \leq n\}, \\
\underline{\mathrm{R}}(n) &:= \max\{\underline{\mathrm{R}}(w) : w \text{ is an edge-circuit in } \Gamma \text{ with } \ell(w) \leq n\}.
\end{aligned}$$

We remark that the functions $\underline{\mathrm{R}}$ and $\overline{\mathrm{R}}$ depend on the presentation, unlike the other filling functions we discussed, which are invariants of Tietze transformations up to the appropriate notion of equivalence.[2]

One has $\mathrm{R}(w) \leq \underline{\mathrm{R}}(w) \leq \overline{\mathrm{R}}(w)$ for all edge circuits w, and hence $\mathrm{R}(n) \leq \underline{\mathrm{R}}(n) \leq \overline{\mathrm{R}}(n)$ for all n. Let M be the length of the longest relation of the finite presentation $\mathcal{P}$. A van Kampen diagram D over $\mathcal{P}$ satisfies $\mathrm{R}(D) \leq M\mathrm{Area}(D)$, since $M\mathrm{Area}(D)$ is an upper bound on the number of 1-cells in a topological disc component of D. It follows that the radius of any minimal area diagram for an edge circuit w is bounded by $M\mathrm{Area}(w)$; so $\overline{\mathrm{R}}(w) \leq M\mathrm{Area}(w)$ and thus $\overline{\mathrm{R}}(n) \leq Mf_0(n)$ for all n. On the other hand our main theorem gives an inequality in the opposite direction: in §3 we establish the following single exponential bound for the Dehn function in terms of the upper filling radius function.

Theorem 1. *Let $\mathcal{P}$ be a finite presentation. There is a constant $C > 1$ so that for all edge-circuits w in the Cayley Graph of $\mathcal{P}$ one has* $\mathrm{Area}(w) \leq \ell(w)C^{\overline{\mathrm{R}}(w)}$.

In Example 2.4 of §2 we use the Baumslag-Solitar group to show this result is best possible in this generality.

Corollary 1. *One has $f_0(n) \leq nC^{\overline{\mathrm{R}}(n)}$ for suitable constant $C > 1$ and for all n.*

We define an AR-pair (f, g) for our finite presentation $\mathcal{P}$ (in analogy to the AD-pairs introduced in [12]) to be an ordered pair of functions $f, g : \mathbb{N} \to \mathbb{N}$ such that for every edge-circuit w there exists a van Kampen diagram D_w such that $\mathrm{Area}(D_w) \leq f(\ell(w))$ and $\mathrm{R}(D_w) \leq g(\ell(w))$. Note that f is an isoperimetric function and g is an upper bound on the filling radius. As an example of the terminology, one way to state the double exponential theorem is that there is an AR-pair for $\mathcal{P}$ of the form $(A^{B^{\mathrm{R}(n)+n}}, \mathrm{R}(n))$.

An example of an AR-pair for $\mathcal{P}$ is $(f_0, \underline{\mathrm{R}})$ since for an edge circuit w we can take D_w to be minimal area diagram which has minimal radius amongst all possible minimal area fillings. Applying Corollary 1 we therefore have:

[2] For a discussion of equivalence relations for filling functions and the behaviour of filling functions on change of presentation see [12, Theorem 4] and [13, section 2]. Briefly, for f_0, Diam and FL the notion of equivalence involves affine change of variables in domain and range plus addition of a linear function, whereas for R one just performs the affine transformations and omits the addition of a linear function.

Corollary 2. *There is a constant $C > 1$ so that $(nC^{\overline{\mathrm{R}}(n)}, \underline{\mathrm{R}}(n))$ is an AR-pair and $(nC^{\overline{\mathrm{R}}(n)}, \underline{\mathrm{R}}(n) + n)$ is an AD-pair for $\mathcal{P}$.*

The next result is a bound on the filling length function $h_0 : \mathbb{N} \to \mathbb{N}$. We apply the main result (Theorem 1) of [12], which gives a bound on filling length in terms of a diameter bound multiplied by the logarithm of a simultaneously realisable area bound, to the AD-pair of Corollary 2.

Corollary 3. *There is a constant $E > 0$ so that $h_0(n) \leq E(\overline{\mathrm{R}}(n)+n)(\underline{\mathrm{R}}(n)+n))$ for all n.*

The following theorem serves to estimate the growth of the function $\overline{\mathrm{R}}$ relative to the area.

Theorem 2. *Suppose that f is an isoperimetric function for a finite presentation $\mathcal{P}$ such that $f(n)/n$ is monotone increasing for large n. Then there exist constants $A, B > 0$ so that*

$$A \log(f_0(n)/n) \leq \overline{\mathrm{R}}(n) \leq B \frac{f(n) \log n}{n} + B$$

for all $n > 0$.

The left inequality is a restatement of Theorem 1. The right inequality is addressed in §4, Corollary 4.4, and tells us in particular that

$$\lim_{n \to \infty} \overline{\mathrm{R}}(n)/f(n) = 0$$

with f as in the theorem. Its proof is itself based on a bound we prove in Proposition 4.1 for the filling radius of a minimal area van Kampen diagram in terms of area and valid for a general presentation.

Proposition 4.1 yields several corollaries. In Corollary 4.2 we show that if $\mathcal{P}$ presents a hyperbolic group then there is a logarithmic bound on $\overline{\mathrm{R}}$. Further we obtain a generalisation of a statement of Gromov for polynomial isoperimetric functions f [15, p. 100] (verified for quadratic polynomials in [18, p. 799] and for general polynomials of degree ≥ 2 in [12, Lemma 2]):

Theorem 3. *Suppose f is an isoperimetric function for the finite presentation $\mathcal{P}$ such that $g(n) := f(n)/n$ satisfies $g(2n) \geq 2g(n)$ for all $n > 0$. Then there is a constant $C > 0$ so that for all $n \geq 1$,*

$$\overline{\mathrm{R}}(n) \leq C \left(1 + \log_2 n + \frac{f(n)}{n} \right).$$

In particular to obtain such a bound it is enough for $g(n)$ to be superadditive.

We gratefully acknowledge comments and suggestions by Martin Bridson.

2 Examples

2.1 For finitely generated groups, sublinearity of the filling radius function R is a characterisation of hyperbolicity. That sublinearity of R implies hyperbolicity is Proposition 3.2.6 of [8]. (We gratefully acknowledge M. Kapovich for also providing a proof of this fact.) Both proofs rely on the characterization of hyperbolic groups as finitely generated groups all of whose asymptotic cones are $\mathbb{R}$-trees (see [8] and [15]).

The proof that hyperbolicity implies sublinear (indeed logarithmic) radius R is straightforward (see Proposition 3.2.6 of [8] for example). However in Corollary 4.2 we will actually prove more: if $\mathcal{P}$ presents a hyperbolic group G, then there is a constant $C > 0$ so that one has $\overline{\mathrm{R}}(n) \le C(1 + \log_2 n)$. Hence the logarithmic bound on filling radius is achieved on every minimal area van Kampen diagram. This result is stated in Gromov [15, 5.C., page 100].

2.2 Another type of example is given by a presentation $\mathcal{P}$ which satisfies a polynomial isoperimetric inequality of degree $d \ge 2$. In this case, it was proved in [12, §5] (generalising the $d = 2$ case in [18]) that there is a positive constant A so that $\overline{\mathrm{R}}(n) \le A\ n^{d-1}$. This result is a special case of Theorem 3.

2.3 The problem of calculating $\overline{\mathrm{R}}(w)$ for a loop w is quite formidable, since it requires some knowledge of all minimal area fillings of w. There is, however, one case where this is easy. Suppose that $\mathcal{P}$ is *aspherical* and suppose that there exists a van Kampen diagram D_w for w that imbeds in the Cayley 2-complex X. Then it is the case that $\overline{\mathrm{R}}(w) = \mathrm{R}(D_w)$. This is a consequence of what Gromov calls "Gersten's lemma" in [3] 4.C$_2$: under the hypotheses that $\mathcal{P}$ is aspherical and D_w imbeds in X, if $c(D_w)$ denotes the integral cellular 2-chain of D_w in $C_2(X, \mathbb{Z})$, then D_w is determined by $c(D_w)$.

Here is a proof of this lemma. It follows from the asphericity of $\mathcal{P}$ that the 2-chain of D_w is determined by w. The 2-cells of D_w are determined by its 2-chain whereas the 1-dimensional portions of the van Kampen diagram D_w are determined by w. Hence the image of D_w in X is determined. Since the map $D_w \to X$ is an imbedding, it follows that this map is uniquely determined by a choice of base point in X.

We deduce that D_w is the unique minimal area filling of w (and its imbedding in X is unique up to left translation by group elements), whence $\overline{\mathrm{R}}(w) = \mathrm{R}(D_w)$.

2.4 Let G be the mapping torus of an injective endomorphism ϕ of the finitely generated free group $F = F(x_1, x_2, \ldots, x_r)$ and let $\mathcal{P} = \langle x_1, x_2, \ldots, x_r, t \mid tx_it^{-1} = \phi(x_i);\ 1 \leq i \leq r\rangle$. Then a minimal area van Kampen diagram D for an edge circuit w contains no annular t-corridors. Therefore the 2-dimensional portion of D must consist entirely of t-corridors, each connecting an instance of t in w to an instance of t^{-1}. There can be at most $\ell(w)/2$ such corridors. A vertex on the path along one side of one of these corridors is a distance at most $M/2$ from the path along the other side, where M is the length of the longest relator in $\mathcal{P}$. We deduce that $\overline{\mathrm{R}}(n) \leq Mn/4$.

A special case is the Baumslag-Solitar group, which has the aspherical presentation $\langle x, t \mid txt^{-1} = x^2\rangle$. We find $\overline{\mathrm{R}}$ is linearly bounded but the area is exponential. In this example, $\overline{\mathrm{R}}$ is actually bounded below by a linear function. There are obvious imbedded van Kampen diagrams for the null-homotopic words $[t^nxt^{-n}, x]$. By 2.3 these are the unique minimal area fillings. Using some hyperbolic geometry one shows that these fillings have depth at least n, so $\overline{\mathrm{R}}$ has a linear upper bound and a (nonzero) linear lower bound.[3]

2.5 In [4] M. R. Bridson gives a family $J(a, b)$ (where a, b are positive integers with $a \geq b$) and proves (in Proposition 7.2) that $\underline{\mathrm{R}}(n) \simeq n^{a/b}$. (See [13] for the definition of $\simeq$-equivalence.)

3 A single exponential bound on Dehn functions

We need a couple of definitions before we proceed to a proof of Theorem 1.

Star neighbourhoods. For a subcomplex K of $\mathcal{D}$ define $\mathrm{Star}(K)$ to be the

[3] There are hyperbolic groups G among those occurring in example 2.4. For example, when $\phi : F \to F$ is an automorphism, $G = F \rtimes_\phi \mathbb{Z}$ is a hyperbolic group iff ϕ is hyperbolic in the sense of Gromov, by a theorem of P. Brinkmann [5]; in this case $\overline{\mathrm{R}}$ is sublinear, by 2.1.

union of closed 2-cells meeting K. Define $\mathrm{Star}_i(K)$ to be the i-th iterate of the star operation for $i \geq 1$; by convention $\mathrm{Star}_0(K) = K$.

Diamond moves. These can also be referred to as *Dehn surgeries* and are introduced and discussed in [7]. A diamond move can be performed in a van Kampen diagram D of a finitely presented group $G = \langle \mathcal{A} \mid \mathcal{R} \rangle$, when there is an $a \in \mathcal{A}^{\pm 1}$ such that the string aa^{-1} can be found in the 1-skeleton of D. In other words there are two distinct oriented edges e_1 and e_2 with the same initial vertex v, with the same edge labels a; let γ be the path of length 2 along e_1 in the direction towards v and then along e_2 in the direction away from v; then along γ one reads a^{-1} followed by a.

The diamond move consists of the following. Cut the diagram along γ: this introduces a hole with boundary label $a^{-1}aa^{-1}a$; we can remove this hole by identifying pairs of adjacent edges in two possible ways - one (necessarily) returns us the original diagram D; performing the other is a diamond move.

The diamond moves we will use in the proof of Theorem 1 are illustrated in Figure 2. As with the examples pictured there, applying the diamond move when the terminal vertices of the edges e_1 and e_2 are distinct produces a van Kampen diagram D_1. Further $\mathrm{Area}(D) = \mathrm{Area}(D_1)$. In the case where e_1 and e_2 have the same terminal vertex v_t, the two edges enclose a subdiagram D'. Let S be the spherical 2-complex obtained from D' by identifying e_1 with e_2. The result of the diamond move is to produce a 2-complex which consists of S attached at v_t to the van Kampen diagram obtained by removing D' from D and identifying e_1 with e_2.

Remark 3.1. In a minimal area van Kampen diagram for a word w, if e_1 and e_2 are edges in D with the same initial vertices and the same edge labels, then their terminal vertices are different - otherwise we could remove the subdiagram enclosed by e_1 and e_2 to produce a diagram of lower area.

This remark together with the comments on diamond moves above lead us to:

Lemma 3.2. *Let D be a minimal area van Kampen diagram for the word w. Suppose D_1 is a 2-complex resulting from applying a diamond move to D. Then*

1. *D_1 is itself a van Kampen diagram, and*
2. $\mathrm{Area}(D) = \mathrm{Area}(D_1)$, *whence D_1 is also a minimal area van Kampen diagram for w.*

We come now to our main result.

Theorem 1. *Let $\mathcal{P}$ be a finite presentation. There is a constant $C > 1$ so that for all edge-circuits w in the Cayley Graph of $\mathcal{P}$ one has* $\text{Area}(w) \leq \ell(w) C^{\overline{\text{R}}(w)}$.

Proof. Suppose D is a minimal area van Kampen diagram for an edge circuit w. We use the Star operation to decompose D into *annuli*. Let $N_i := \text{Star}_i(\partial D) \subseteq D$. For $i \geq 1$ let c_i be the inner boundary of N_i, and let c_0 be the boundary of the closure of the interior of D (*i.e.* ∂D without the 1-dimensional portions). For $i \geq 1$ let $\Delta_i = N_i \setminus N_{i-1}$, so the inner boundary of Δ_i is c_i and the outer boundary is c_{i-1}. The Δ_i constitute an *annular* decomposition of D as depicted in Figure 1.

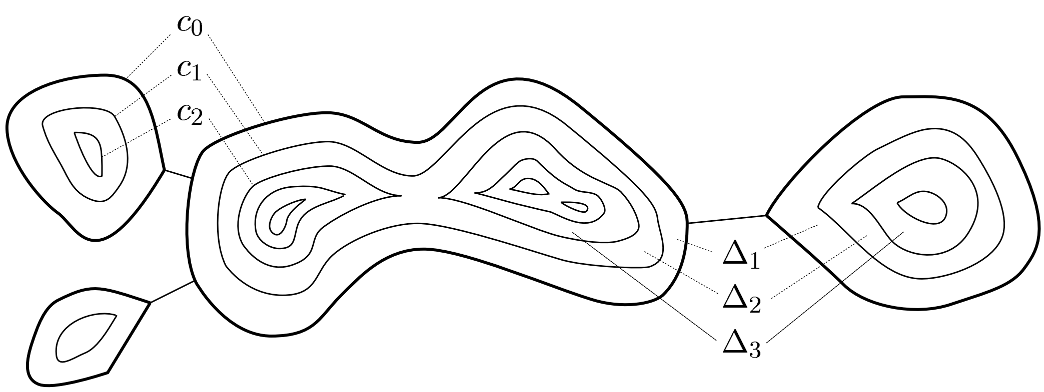

Figure 1: The *annular* decomposition of D.

Observe that $D = \partial D \cup \bigcup_{i=1}^{\overline{\text{R}}(w)} \Delta_i$. Let M be the length of the longest relator in $\mathcal{P}$ and L be four times the number of generators.

Our object is to produce D satisfying the inequalities

$$\text{Area}(\Delta_i) \leq L\, \ell(c_{i-1}), \tag{1}$$

$$\ell(c_i)/M \leq \text{Area}(\Delta_i), \tag{2}$$

relating the area of Δ_i (for $i \geq 1$) to the lengths of its outer and inner boundaries; for $i \geq 2$, these combine to give $\text{Area}(\Delta_i) \leq LM\text{Area}(\Delta_{i-1})$ and hence for $i \geq 1$ we have $\text{Area}(\Delta_i) \leq L^i M^{i-1} \ell(c_0) \leq L^i M^{i-1} \ell(w)$.

It will then follow that

$$\mathrm{Area}(D) \leq \sum_{i=1}^{\overline{\mathrm{R}}(w)} \ell(w) L (LM)^{i-1} = \ell(w) L \frac{(LM)^{\overline{\mathrm{R}}(w)} - 1}{LM - 1}$$

completing the proof of the theorem.

Now inequality (2) follows from the observation that each 1-cell of c_i is an edge of a 2-cell in Δ_i and the total number of edges of 2-cells in Δ_i is at most $M\mathrm{Area}(\Delta_i)$.

Obtaining inequality (1) is less straightforward. There is no *a priori* bound on the valence of a vertex v of c_{i-1} in the closure of the interior of Δ_i, and hence on the number of 2-cells in the interior of Δ_i that are incident with v. However we use diamond moves to prove the following lemma, which says that there is *some* minimal area diagram D for w for which such a bound exists. It follows that (1) holds for D, and hence the theorem.

Lemma 3.3. *There exists a minimal area van Kampen diagram D filling w such that there is a uniform upper bound L on the valences of vertices of the curves c_i in the closure of the interior of c_i.*

Proof. Start with any minimal area van Kampen diagram diagram D' for w. We use diamond moves to transform D' to a van Kampen diagram D for w with the required properties. In the process we may increase the radius of the diagram, but by Lemma 3.2 the area remains minimal. Thus $\mathrm{R}(D) \leq \overline{\mathrm{R}}(w)$.

Consider a topological disc component D_1 of D' and a boundary vertex v of D_1. Let e_1, e_2 be distinct oriented edges in D_1 with initial vertex v, and with the same label. Then a diamond move is possible, but shall we do it? The answer depends on the end points x, y respectively of the edges. Observe that $x \neq y$ by minimality of the area of D' (Remark 3.1).

The rule is do the diamond move only if either

1. x and y are both in ∂D_1, or

2. x and y are both in the interior of D_1.

But we still have to decide in what order to do the surgeries.

These two types of diamond moves are illustrated in Figure 2. In case 1 the effect of the surgery is to increase the number of disc components, or, better, to decrease the number of distinct vertices on the boundary (when

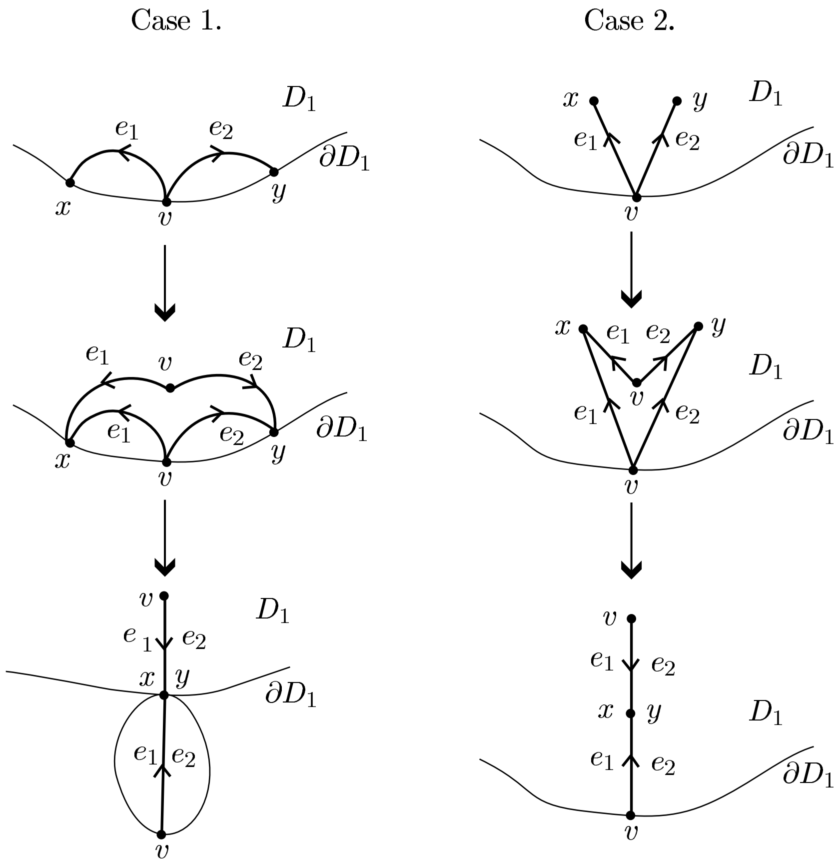

Figure 2: Diamond moves.

e_1, e_2 have terminal vertices on ∂D_1). In case 2 there is no change on the boundary, but after surgery there is at least one fewer edge incident with the boundary of a disc component. One defines the induction parameter to be the ordered pair (a, b) ordered lexicographically, where a is the number of vertices on the boundary of D' and b is the number of edges in which are in topological disc components of D' which are incident with the boundary. Note that each of the two types of allowed diamond moves decreases the induction parameter.

Let L be twice the number of vertices in J. H. C. Whitehead's star graph ([17, p. 61]) of the finite presentation[4] $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$, that is, $L := 4|\mathcal{A}|$.

[4]This is where the group theory enters, or, more precisely, where the finite presentation

Suppose that the induction parameter cannot be reduced by the two types of diamond moves allowed, and let v be a vertex of ∂D_1, a disc component of D'. If there were three oriented edges e_1, e_2, e_3 in D_1, each incident at v, and having the same label a, then two of them would either both end in the interior or both end in the boundary of D_1. In either case one of the allowed diamond moves is possible, and the induction parameter can be reduced, contrary to assumption. It follows that each label a can occur at most twice among the oriented edges in D_1 and incident at v. The number of labels is the same as the number of edges of the star graph, namely $2|\mathcal{A}|$. It follows that the number of edges in D_1 and incident at v is at most $4|\mathcal{A}| = L$.

After having reduced the induction parameter to a minimum by the allowed types of diamond moves, we achieve a minimal area van Kampen diagram D_1' for w, so that the number of corners of 2-cells incident with $\partial D_1'$ is $\leq L \cdot \ell(\partial D_1')$. Then one takes the star neighbourhood N' of $\partial D_1'$ in D_1' and considers the inner boundary of N' and repeats the process with w replaced by the inner boundary label(s) and D' replaced by their interiors in D_1'. Successively repeating this procedure (at most $\overline{\mathrm{R}}(w)$ times) we eventually arrive at a diagram D with the properties we require.

This completes the proof of the lemma, and the proof of the theorem is complete. □

4 Bounding upper filling radius in terms of isoperimetric functions.

In §1 we gave the crude estimate on the radius of a van Kampen diagram D over a finite presentation $\mathcal{P}$:

$$\mathrm{R}(D) \leq M \mathrm{Area}(D),$$

where M is the length of the longest relation of $\mathcal{P}$. From this it follows that

$$\mathrm{R}(n) \leq \overline{\mathrm{R}}(n) \leq M f_0(n).$$

The following proposition provides an improved bound on the filling radius in terms of the Dehn function; examples 2.1 and 2.2 of §2 are special cases. We use $\lceil x \rceil$ to denote the smallest integer bounding the real number

is used.

x from above. Throughout this section isoperimetric functions are assumed to have domain $[0, \infty)$.

Proposition 4.1. *Let $\mathcal{P}$ be a finite presentation and M the length of its longest relation. Then for all minimal area van Kampen diagrams D for edge-circuits w of length at most n (with $n \geq 1$) one has*

$$\mathrm{R}(D) \leq M^2 \left(\left\lceil \frac{f_0(n)}{n/2} \right\rceil + \left\lceil \frac{f_0(n/2)}{n/4} \right\rceil + \cdots + \left\lceil \frac{f_0(n/2^{s-1})}{n/2^s} \right\rceil + f_0(1) \right),$$

where $s = \lceil \log_2 n \rceil$.

It follows that $\mathrm{R}(D) \leq M^2 \left(f_0(1) + s + \sum_{i=0}^{s-1} \frac{f_0(n/2^i)}{n/2^{i+1}} \right)$.

Proof. We prove the proposition by induction on n.

Let w be an edge circuit of length at most n and let D be a minimal area van Kampen diagram filling w. If $n = 1$ then $\mathrm{R}(D) \leq M f_0(1)$ and the proposition holds.

For the induction step we employ the *annular* decomposition of D used in the proof of Theorem 1. Recall that for all $i \geq 1$ we established the inequality:

$$\mathrm{Area}(\Delta_i) \geq \ell(c_i)/M.$$

It follows that there exists $m \leq M \left\lceil \frac{f_0(n)}{n/2} \right\rceil$ such that $\ell(c_m) \leq n/2$, for otherwise

$$\mathrm{Area}(D) = \sum_{i=1}^{\mathrm{R}(D)} \mathrm{Area}(\Delta_i) \geq \sum_{i=1}^{\mathrm{R}(D)} \ell(c_i)/M > \left\lceil \frac{f_0(n)}{n/2} \right\rceil n/2 \geq f_0(n).$$

Now for such m we find that for any vertex $v \in N_m$ (recall from the proof of Theorem 1 that $N_i = \mathrm{Star}_i(\partial D)$, so v is not in the interiors of the subdiagrams of D enclosed by c_m) we have

$$d(v, \partial D) \leq M^2 \left\lceil \frac{f_0(n)}{n/2} \right\rceil.$$

Now c_m is a union $c_m = \bigcup_{i \in I_m} c_m^i$ of simple closed curves c_m^i any two of which meet at no more than one vertex. These c_m^i each have length at most $n/2 \geq 1$

and enclose a minimal area van Kampen diagrams D_m^i. By induction, for all $i \in I_m$ the diagram D_m^i satisfies

$$\mathrm{R}(D_m^i) \leq M^2 \left(\left\lceil \frac{f_0(n/2)}{n/4} \right\rceil + \left\lceil \frac{f_0(n/4)}{n/8} \right\rceil + \cdots + \left\lceil \frac{f_0((n/2)/2^{s-1})}{(n/2)/2^s} \right\rceil + f_0(1) \right),$$

where $s = \lceil \log_2(n/2) \rceil = \lceil \log_2 n \rceil - 1$. Now

$$\mathrm{R}(D) \leq M^2 \left\lceil \frac{f_0(n)}{n/2} \right\rceil + \max_{i \in I_m} \mathrm{R}(D_m^i).$$

The result therefore follows. □

A first corollary is that hyperbolic groups have a logarithmic bound on their upper filling length $\overline{\mathrm{R}}$.

Corollary 4.2. *Let $\mathcal{P}$ be a finite presentation for a hyperbolic group G. Then there exists $C > 0$ such that for $n \geq 1$,*

$$\mathrm{R}(n) \leq \overline{\mathrm{R}}(n) \leq C\left(1 + \log_2 n\right).$$

Proof. In [14] Gromov characterises hyperbolic groups as the finitely presented groups that satisfy a linear isoperimetric function. Take $K > 0$ such that $n \mapsto Kn$ is an isoperimetric function for $\mathcal{P}$. Applying this to the bound of Proposition 4.1 we get $\overline{\mathrm{R}}(n) \leq M^2 \left(f_0(1) + (1 + 2K)\left(1 + \log_2 n\right)\right)$. □

Remark 4.3. The significance of this result is that all minimal area van Kampen fillings in a presentation of a hyperbolic group satisfy a uniform logarithmic bound on their radius. Special constructions of diagrams with logarithmic bounds on their radius are given in [20].

The next corollary tells us that any reasonably well-behaved upper bound on the area is a gross overestimate of $\overline{\mathrm{R}}$, so the theorem and proposition give better estimates on $\overline{\mathrm{R}}$ from below and from above, respectively, than the naive area estimate.

Corollary 4.4. *Suppose that f is an isoperimetric function for our finite presentation $\mathcal{P}$ and that $f(n)/n$ is monotone increasing for all $n \geq N$ (for some fixed $N > 0$). Then there exists $C > 0$ such that*

$$\overline{\mathrm{R}}(n) \leq C\left(1 + \frac{f(n)}{n}\right)\left(1 + \log_2 n\right)$$

for all $n \geq 1$. So in particular $\lim_{n \to \infty} \overline{\mathrm{R}}(n)/f(n) = 0$.

Proof. We use $1 + 2\frac{f(n)}{n}$ as a bound on all but finitely many of the $\lceil \log_2 n \rceil$ terms in the summation in Proposition 4.1. This gives

$$\overline{\mathrm{R}}(n) \leq M^2 \left(f_0(1) + \left(1 + 2\frac{f(n)}{n}\right) \lceil \log_2 n \rceil + C' \right)$$

for some constant $C' > 0$. $\square$

Many familiar functions, like x^r, $r \geq 2$, and $\exp x$, satisfy the hypothesis of Corollary 4.4, and indeed satisfy the hypothesis of our final corollary:

Corollary 4.5. *Suppose f is an isoperimetric function for $\mathcal{P}$ and that $f(n/2^i) \leq f(n)/2^{2i}$ for all $n > 0$ and $i = 0, 1, \ldots, \lceil \log_2 n \rceil - 1$. Then there exists $C > 0$ such that for all $n \geq 1$,*

$$\overline{\mathrm{R}}(n) \leq C \left(1 + \log_2 n + \frac{f(n)}{n} \right).$$

Proof. This bound is obtained by applying $f(n/2^i) \leq f(n)/2^{2i}$ to each term in the summation in Proposition 4.1, which are then seen to be bounded by terms in a geometric series. $\square$

Corollaries 4.2 and 4.5 generalize results of Gromov [15] p. 100, Papasoglu [18] p. 799, and [12] Proposition 3.

The hypothesis of Corollary 4.5 appears somewhat mystifying. In Theorem 3 we will give a couple of stronger but more transparent hypotheses to clarify the result and thereby to enhance its utility (we hope). First we recall a definition:

Definition. If g is a nonnegative real valued function on $\mathbb{N}$, then g is called *superadditive* if $g(m+n) \geq g(m) + g(n)$ for all $m, n > 0$. Note that superadditive functions are increasing. Superadditive functions have been considered previously in relation to isoperimetric functions in [16] and [2] (in the latter they are called "subnegative").

Theorem 3. *Suppose f is an isoperimetric function for the finite presentation $\mathcal{P}$ such that $g(n) := f(n)/n$ satisfies $g(2n) \geq 2g(n)$ for all $n > 0$. Then there is a constant $C > 0$ so that for all $n \geq 1$,*

$$\overline{\mathrm{R}}(n) \leq C \left(1 + \log_2 n + \frac{f(n)}{n} \right).$$

In particular to obtain such a bound it is enough for $g(n)$ to be superadditive.

Proof. By Corollary 4.5 it suffices to show that $f(n/2^i) \leq f(n)/2^{2i}$ for all positive numbers n, i with $2^i \leq n$. The condition $g(2n) \geq 2g(n)$ for all $n > 0$ is the same as $f(2n) \geq 4f(n)$ for all $n > 0$, that is, $f(n/2) \leq f(n)/4$ for all $n > 0$. An induction on i then shows that $f(n/2^i) \leq f(n)/2^{2i}$ as required. $\square$

5 Comparing R with $\underline{\mathrm{R}}$ and $\overline{\mathrm{R}}$

Open problem. Given a finite presentation $\mathcal{P}$, bound $\overline{\mathrm{R}}$ in terms of R. The double exponential theorem states that $f_0(n) \leq A^{B^{\mathrm{R}(n)+n}}$ for suitable constants $A, B > 1$. Since $\overline{\mathrm{R}}(n) \leq M f_0(n)$, it follows that $\overline{\mathrm{R}}$ is bounded by a double exponential function of R. This seems wildly extravagant, and one might hope for a single exponential bound.

Example.[5] Here is an example (due to M. R. Bridson [3]) of a finite presentation for which the radius R differs exponentially (at least on an infinite subset of $\mathbb{N}$) from the both the upper radius $\overline{\mathrm{R}}$ and the lower radius $\underline{\mathrm{R}}$.

Let $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ be the aspherical presentation

$$\langle x, y, s, t \mid [x, y] = 1,\ txt^{-1} = x^2,\ sys^{-1} = y^2 \rangle.$$

Define a family of edge circuits $w_n := [t^n x t^{-n}, s^n y s^{-n}]$. Let D_n be the obvious van Kampen diagram D_n filling w_n: consisting of two *triangular* diagrams for the edge circuits $t^n x t^{-n} x^{-2^n}$ and two *triangular* diagrams for $s^n y s^{-n} y^{-2^n}$, surrounding a square diagram for the circuit $[x^{2^n}, y^{2^n}]$ (see Figure 3). The diagram D_n imbeds in the Cayley 2-complex, thus by Gersten's Lemma (see Example 2.3) is the unique minimal area diagram for w_n. It follows that $\underline{\mathrm{R}}(w_n) = \overline{\mathrm{R}}(w_n) = \mathrm{R}(D_n) \geq 2^{n-1}$, that is, the upper radius $\overline{\mathrm{R}}(n)$ and the lower radius $\underline{\mathrm{R}}(n)$ are at least exponential for $n \in \{8m + 4 : m \in \mathbb{N}\}$.

On the other hand one can make use a method of *shortcuts* (independently due to both the first author and M. R. Bridson) to provide a linear bound on the filling radius R. In the instance of the diagram D_n, shortcuts are inserted by cutting along each of the $2^n - 1$ lines labelled x^{2^n} within the square portion of D_n and then inserting two copies of the minimal area diagram T_n for $t^n x t^{-n} = x^{2^n}$ in the subpresentation $\langle x, t \mid txt^{-1} = x^2 \rangle$, where the two copies have $t^n x t^{-n}$ in common on their boundaries; the two segments of the boundaries labelled x^{2^n} are identified with the two cuts. The blown-up

[5] We are grateful to the referee for suggesting we include consideration of $\underline{\mathrm{R}}$ in this example.

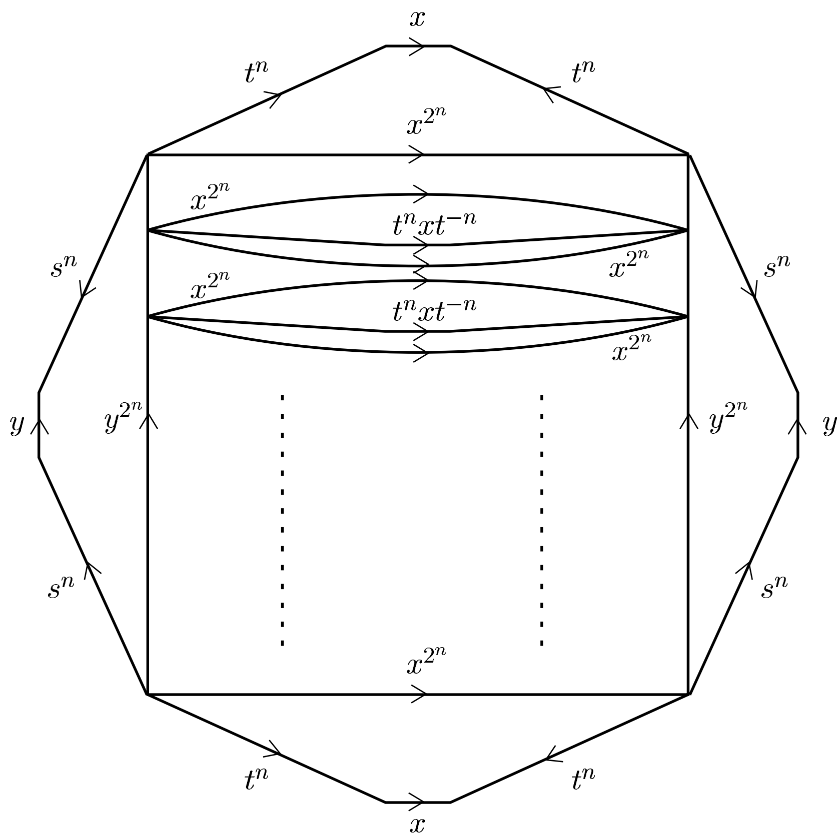

Figure 3: The diagram D'_n for $w_n = [t^n x t^{-n}, s^n y s^{-n}]$ using shortcuts.

diagram D'_n (illustrated in Figure 3) has radius linearly bounded in n; one sees this because it is a linear distance from any vertex to one of the lines labelled $t^n x t^{-n}$, and a linear distance from there to the boundary.

The diagram D_n has area $4(2^n-1)+2^{2n}$; inserting the shortcuts to produce D'_n increases the area by $2(2^n-1)^2$. So D'_n is far from having minimal area.

More generally we can prove

Proposition 5.1. *There is a linear bound on the filling radius* R *of*

$$\mathcal{P} = \langle x, y, s, t \mid [x,y] = 1,\ txt^{-1} = x^2,\ sys^{-1} = y^2 \rangle.$$

Proof. Given any edge-circuit w in the Cayley graph of $\mathcal{P}$ we use shortcuts to produce a van Kampen diagram D' for w with radius bounded linearly in $n = \ell(w)$. We need a couple of preliminary notions.

Call an edge-circuit w_0 *standard* if it is reduced (*i.e.* is without backtracking) and every segment w_0' of w_0 consisting entirely of x, y, x^{-1}, y^{-1}'s is of the form $x^a y^b$ for some $a, b \in \mathbb{Z}$.

Consideration of the s and t-corridors in any van Kampen diagram for an edge-circuit w_0 leads us to:

Lemma 5.2. *Suppose that w_0 is a nonempty standard edge-circuit. Then w_0 contains a segment u of one of the following forms*

$$tx^k t^{-1},\ t^{-1}x^k t,\ sy^k s^{-1},\ s^{-1}y^k s,$$

for some $k \in \mathbb{Z} - \{0\}$.

Given such a standard edge-circuit w_0, we can *eliminate an s or t-pair* – that is, say $u = tx^k t^{-1}$, then $u = x^{2k}$ in G; we replace u in w_0 by x^{2k} (and similarly for the other 3 possibilities) to produce an edge-circuit w_0'. This serves as a step in the process of constructing a van Kampen diagram for w_0: producing an s or t-corridor with k 2-cells.

So let us now use these ideas to construct a van Kampen diagram D for our arbitrary edge-circuit w in the Cayley graph of $\mathcal{P}$. First we repeatedly apply the relation $[x, y] = 1$ together with free reduction to express w in standard form w_0, a process which does not increase the length of the word. This produces an annular diagram with outer boundary circuit w and inner boundary circuit w_0. The distance from a vertex on the inner boundary circuit to the outer is at most n.

Now eliminate an s or t-pair from w_0 to produce w_0'. Next convert w_0' into an edge-circuit w_1 in standard form. This involves re-expressing a segment of w_0' of the form $x^{a_1}y^{b_1}x^k x^{a_2}y^{b_2}$ or $x^{a_1}y^{b_1}y^k x^{a_2}y^{b_2}$ in the form $x^a y^b$. The minimal area diagram (over the subpresentation $\langle x, y \mid [x, y]\rangle$) for this transformation consists of a $|b_1| \times |k + a_2|$ or $|b_1 + k| \times |a_2|$ *rectangle* (respectively) together with 1-dimensional portions corresponding to free reduction.

We can now repeat the process: eliminate an s or t-pair from w_1 to produce w_1' and then convert to standard form giving w_2, and so on. There are less than $n/2$ pairs s, s^{-1} and t, t^{-1} in w, so for some $m < n/2$ we find w_m is the empty word and the construction of D is complete.

Lemma 5.3. *The length of the sides of the* rectangles *in the construction of D is bounded by 2^{2n}.*

Proof. To obtain this (crude) bound it is enough to control $\max\{L_i : i = 0, 1, \ldots\}$ where

$$L_i := \max\{k : x^{\pm k} \text{ or } y^{\pm k} \text{ is a segment of } w_i\}.$$

Now $L_{i+1} \leq 4L_i$ because when a segment $x^{a_1}y^{b_1}x^{2k}x^{a_2}y^{b_2}$ or $x^{a_1}y^{b_1}y^{2k}x^{a_2}y^{b_2}$ in w_i' is re-expressed as $x^a y^b$ to produce the edge-circuit w_{i+1} we find

$$|a|, |b| \leq 2|k| + \max\{|a_1| + |a_2|, |b_1| + |b_2|\}$$

with $|a_1|, |a_2|, |b_1|, |b_2|, |k| \leq L_i$. So each L_i is bounded by $4^i L_0 \leq 4^{n/2} L_0 \leq 2^n \cdot n < 2^{2n}$. Hence the bound stated in the lemma.

In the construction of D, formation of a *rectangle* follows the construction of an s or t-corridor. So the boundary of each *rectangle* meets the side of an s or t-corridor. To produce D' we will insert *shortcuts* into all of the *rectangles* and also along the sides of all s and t-corridors. Then it will be possible to proceed from any vertex in D' to the boundary within a linearly bounded distance.

So let us now describe how to create shortcuts. Recall that the shortcuts in the diagrams D_n' of Figure 3 were provided by two copies of the diagram T_n, the minimal area diagram for the edge-circuit $t^n x t^{-n} x^{-2^n}$ over the subpresentation $\langle x, t \mid txt^{-1} = x^2 \rangle$. In the more general setting we will use subdiagrams $T_{m,k}$ of T_m (defined for $k = 1, 2, 3, \ldots, 2^m$): obtain $T_{m,k}$ by reading along the x^{2^m} boundary segment of T_m until x^k is reached and then cut along a shortest path to the $t^m x t^{-m}$ boundary segment; the resulting diagram has boundary circuit labelled $x^k u$ with $\ell(u) \leq 3m + 1$, and further the distance from any vertex of the x^k segment to the u segment is at most $2m$. Let $T^2_{m,k}$ denote the diagram with boundary circuit $x^k x^{-k}$ obtained by gluing together two copies of $T_{m,k}$ along the segment u. Then the distance in $T^2_{m,k}$ between any two vertices is at most $2 \cdot 2m + (3m + 1) = 7m + 1$.

Define $S^2_{m,k}$ to be the diagram constructed analogously over the subpresentation $\langle y, s \mid sys^{-1} = y^2 \rangle$.

We insert shortcuts into all the *rectangles* of D: if $[x^k, y^{k'}]$ is the boundary circuit of such a *rectangle* then $|k|, |k'| \leq 2^{2n}$ by Lemma 5.3; cut along each of the $|k'| + 1$ paths labelled x^k, and glue in copies of $T^2_{2n,|k|}$; further cut along the two sides of the rectangle labelled $y^{k'}$ and glue in copies of $S^2_{2n,|k'|}$. The distance between any two vertices in a *rectangle* after shortcuts have been added is at most $3(7 \cdot 2n + 1) = 42n + 3$.

A path along the side of an s or t-corridor in D is labelled by x^k or y^k for some k with $|k| \leq 2^{2n}$. Cut along all of these paths and insert copies of $T^2_{2n,|k|}$ or $S^2_{2n,|k|}$ respectively.

Let D' be the diagram for w obtained from D by inserting these shortcuts.

We now claim the bound

$$\mathrm{R}(D') \leq (42n+3) + (14n+1) + n = 57n + 4$$

on the radius of D' in terms of $n = \ell(w)$. From any vertex of one of the rectangles one can reach an s or t-corridor within a distance $42n+3$ (as calculated above). Then one can reach the end of the corridor (and hence the edge-circuit w_0) in a further distance $14n+1$. Finally one arrives at the boundary circuit w from w_0 within a distance of at most n. □

References

[1] J. Avenhaus and K. Madlener. The Nielsen reduction and P-complete problems in free groups. *Theoret. Comput. Sci.*, 32(1-2):61–76, 1984.

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

[3] M. R. Bridson. Area versus diameter for van Kampen diagrams. Unpublished Notes, Oxford University, 1997.

[4] M. R. Bridson. Fractional isoperimetric inequalities and subgroup distortion. *J. of the A. Math. Soc.*, 12(4):1103–1118, 1999.

[5] P. Brinkmann. Hyperbolic automorphisms of free groups. *Geom. Funct. Anal.*, 10(5):1071–1089, 2000.

[6] D. E. Cohen. Isoperimetric and isodiametric inequalities for group presentations. *Int. J. of Alg. and Comp.*, 1(3):315–320, 1991.

[7] D. J. Collins and J. Huebschmann. Spherical diagrams and identities among relations. *Math. Ann.*, 261(2):155–183, 1982.

[8] C. Druţu. Asymptotics cones and quasi-isometry invariants for hyperbolic metric spaces. *Ann. Inst. Fourier Grenoble*, 51:81–97, 2001.

[9] B. Farb. The extrinsic geometry of subgroups and the generalized word problem. *Proc. London Math. Soc. (3)*, 68(3):577–593, 1994.

[10] S. M. Gersten. The double exponential theorem for isoperimetric and isodiametric functions. *Int. J. of Alg. and Comp.*, 1(3):321–327, 1991.

[11] S. M. Gersten. Dehn functions and l_1-norms of finite presentations. In G. Baumslag and C. Miller, editors, *Algorithms and classification in combinatorial group theory (Berkeley, CA, 1989)*, volume 23 of *MSRI Publications*, pages 195–224. Springer-Verlag, 1992.

[12] S. M. Gersten and T. R. Riley. Filling length in finitely presentable groups. Preprint, to appear in *Geometriae Dedicata*, http://www.math.utah.edu/∼gersten/grouptheory.htm/, 2000.

[13] S. M. Gersten and H. Short. Some isoperimetric inequalities for free extensions. Preprint, http://www.math.utah.edu/∼gersten/grouptheory.htm/, 2000.

[14] M. Gromov. Hyperbolic groups. In S. M. Gersten, editor, *Essays in group theory*, volume 8 of *MSRI publications*, pages 75–263. Springer-Verlag, 1987.

[15] M. Gromov. Asymptotic invariants of infinite groups. In G. Niblo and M. Roller, editors, *Geometric group theory II*, number 182 in LMS lecture notes. Camb. Univ. Press, 1993.

[16] V. S. Guba and M. V. Sapir. On Dehn functions of free products of groups. *Proc. Amer. Math. Soc.*, 127:1885–1891, 1999.

[17] R. C. Lyndon and P. E. Schupp. *Combinatorial Group Theory*. Springer-Verlag, 1977.

[18] P. Papasoglu. On the asymptotic invariants of groups satisfying a quadratic isoperimetric inequality. *J. Differential Geom.*, 44:789–806, 1996.

[19] P. Papasolglu. Isodiametric and isoperimetric inequalities for complexes and groups. *J. London Math. Soc. (2)*, 63(1):97–106, 2000.

[20] H. Short *et al.*. Notes on word hyperbolic groups. In E. Ghys, A. Haefliger, and A. Verjovsky, editors, *Group Theory from a Geometrical Viewpoint (Trieste, 1990)*, volume 8 of *MSRI publications*, pages 3–63. World Scientific Publishing, River Edge, N.J., 1991.

STEVE M. GERSTEN
Mathematics Department, Univ. of Utah
155S. 1400E., Rm. 233
Salt Lake City UT 84112, USA
E-mail: gersten@math.utah.edu
URL: http://www.math.utah.edu/~gersten

TIM R. RILEY
Mathematical Institute
24-29 St Giles
Oxford OX1 3LB, UK
E-mail: rileyt@maths.ox.ac.uk