The gallery length filling function and a geometric inequality for filling length

S. M. Gersten and T. R. Riley

January 25, 2003

Abstract

We exploit duality considerations in the study of singular combinatorial 2-discs (*diagrams*) and are led to the following innovations concerning the geometry of the word problem for finite presentations of groups. We define a filling function called *gallery length* that measures the diameter of the 1-skeleton of the dual of diagrams; we show it to be a group invariant and we give upper bounds on the gallery length of combable groups. We use gallery length to give a new proof of the Double Exponential Theorem. Also we give geometric inequalities relating gallery length to the space-complexity filling function known as *filling length*.

2000 Mathematics Subject Classification: 20F05, 20F06, 57M05, 57M20
Key words and phrases: diagram, filling length, gallery length, van Kampen diagram

1 Introduction

A *diagram* is a finite planar 2-complex homeomorphic to a singular 2-disc. A *van Kampen diagram* is a labelled diagram amounting to a graphical demonstration of how a word in a finite presentation $\mathcal{P}$ of a group Γ that represents the identity, is a consequence of the defining relations. (See Definitions 2.1 and 4.1.)

The use of *diagrams* to probe the geometry of finitely presented groups and to produce invariants (*filling functions*) is well established in Geometric

Group Theory. To date, attention has focused on the area of diagrams, and the resulting *Dehn function* (also known as the *minimal isoperimetric function*), and on the diameter of the 1-skeleton of diagrams (in the combinatorial metric), which gives the *minimal isodiametric function*. And recently the *filling length* of a diagram D has proved important [9, 10] – we define $\mathrm{FL}(D)$ to be the minimal length L such that there is a *combinatorial null-homotopy* of ∂D, across D and down to a base vertex, through loops of length at most L. (See §2.1 for more details.)

The innovation in this article is to bring duality considerations to bear on the study of diagrams. We define the gallery length $\mathrm{GL}(D)$ of a diagram D to be the diameter of the 1-skeleton of its dual. We will also define $\mathrm{DGL}(D)$ which combines diameter and gallery length as realisable on *complementary pairs* of maximal trees – see §2.3. In §2 we give formal definitions of diagrams and of $\mathrm{GL}(D)$ and $\mathrm{DGL}(D)$ as well as other diagram *measurements*, and in Proposition 2.4 we list some of the inequalities that relate diagram measurements. Then in §3 we use DGL to control the filling length of a diagram:

Theorem 3.5. *Fix $\lambda > 0$. There exists $K > 0$ such that for every diagram D in which the boundary of each 2-cell has at most λ edges,*

$$\mathrm{FL}(D) \leq K\,(\mathrm{DGL}(D) + \mathrm{Perimeter}(D)\,).$$

This inequality is an amendment of a suggestion of Gromov [13, §5.C]: the term $\mathrm{DGL}(D)$ replaces $\mathrm{Diam}(D)$. Gromov's inequality was known to fail [6] in the context of 2-discs with Riemannian metrics. In §3 we give our own family of diagrams D_n that have bounded perimeter and diameter, and at most three edges in the boundary of each 2-cell, but have filling length tending to infinity. It follows that Gromov's inequality does not hold in a combinatorial context. Our proof that $\mathrm{FL}(D_n) \to \infty$ takes us into unlikely territory. We use the following result, proved in an appendix, concerning the number $\|n\|$ of terms required to express an integer n as a sum of elements of the symmetric generating set $\mathcal{A}^{\pm} := \left\{ \pm 2^k \mid k \in \mathbb{N} \right\}$ for the additive group $\mathbb{Z}$.

Theorem A.1. *If $\{a_n\,;\ n \geq 1\}$ is a sequence of rational integers which converges in the p-adic integers $\hat{\mathbb{Z}}_p$ to an element which is not a rational integer (i.e. the limit is in $\hat{\mathbb{Z}}_p \smallsetminus \mathbb{Z}$), then the sequence $\{\|a_n\|\,;\ n \geq 1\}$ is unbounded. Hence $\|\ \|$ is unbounded on $\mathbb{Z}$ (and so also on $\mathbb{N}$).*

The diagram measurement GL leads to the *gallery length* filling function $\mathrm{GL}(n)$ for finite presentations. We define $\mathrm{GL}(n)$ in §4. In Theorem 4.5 we prove that $\mathrm{GL}(n)$ is a group invariant in the sense that the gallery length functions of two finite presentations of the same group are $\simeq$-equivalent (see Definition 4.3), modulo a technical hypothesis concerning *fattening* the presentations (see Definition 4.4).

We exploit gallery length in §5 to give a new and transparent proof of the *Double Exponential Theorem* of D.E.Cohen – a cornerstone result of Geometric Group Theory that gives an isoperimetric function for a finite presentation in terms of an isodiametric function:

Theorem 5.3. *Given a finite presentation $\mathcal{P}$ there exists $C > 0$ such that the Dehn function* $\mathrm{Area}(n)$ *and the minimal isodiametric function* $\mathrm{Diam}(n)$ *for $\mathcal{P}$ satisfy*

$$\mathrm{Area}(n) \leq n\, C^{C^{\mathrm{Diam}(n)}}$$

for all $n \in \mathbb{N}$.

The role of $\mathrm{GL}(n)$ in the proof is as a stepping-stone lying at most one exponential above the minimal isodiametric function $\mathrm{Diam}(n)$:

Theorem 5.1. *Let $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ be a finite presentation and w be an edge-circuit in its Cayley graph. Suppose D_w is a minimal diameter van Kampen diagram for w, and moreover is of minimal area amongst all minimal diameter diagrams for w. Then there exists $A > 0$ depending only on $\mathcal{P}$, such that*

$$\mathrm{GL}(D_w) \;\; \leq \;\; A^{1\, +\, 2\, \mathrm{Diam}(D_w)}.$$

It follows that $\mathrm{GL}(n) \leq A^{1\, +\, 2\, \mathrm{Diam}(n)}$ *for all n.*

The bounded valence of the graph dual to the 1-skeleton of D_w allows one to prove that a single exponential of $\mathrm{GL}(n)$ is an upper bound for the Dehn function $\mathrm{Area}(n)$ – see Proposition 5.2.

In §6 we give upper bounds on the gallery length functions of combable groups. Then in §7, building on Theorem 3.5, we relate the filling functions $\mathrm{FL}(n)$, $\mathrm{GL}(n)$ and $\mathrm{DGL}(n)$:

Theorem 7.1. *Let $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ be a finite fat presentation of a group Γ. The filling functions* GL, DGL *and* FL *for $\mathcal{P}$ satisfy* $\mathrm{GL} \leq \mathrm{DGL} \simeq \mathrm{FL}$.

Every finite presentation can be made into a *fat* finite presentation for the same groups by introducing some extra generators and relations – see Definition 4.4.

This article is the first in a series. The second is [11] in which we give conjectures concerning duality and the diameter maximal trees and we explain their implications for Geometric Group Theory, a major one being for the theory of central extensions. It would follow from one of the conjectures that the filling functions GL and DGL are essentially the same. This would lead to the simplification of the conclusion of Theorem 7.1 to $\mathrm{FL} \simeq \mathrm{GL}$. We hope that the series will culminate in a third article in which we present proofs of the conjectures. (*Added May 2005*: to date these conjectures remain unresolved.)

2 Singular disc diagrams

2.1 Definitions

Definition 2.1. A *singular 2-disc diagram* (or, more concisely, a *diagram*) is a combinatorial 2-complex D obtained from some finite combinatorial 2-complex S homeomorphic to the 2-sphere by removing the interior of a 2-cell e_∞. Equivalently D is a finite, planar, contractible, combinatorial 2-complex. It is referred to as singular because, in general, it need not be a topological 2-disc but rather may be a tree-like arrangement of topological discs connected by 1-dimensional arcs.

A finite, connected, undirected graph G embedded in the 2-sphere induces a combinatorial 2-complex structure S with 1-skeleton G. (We allow *graphs* to have multiple edges between two vertices and to have edges that meet only one vertex, thereby forming a loop. These are *multigraphs* in [4].) Associated to S is the dual 2-complex $S^\star$ that has a face dual to each vertex of S, an edge dual to each edge of S, and a vertex dual to each face of S. The 1-skeleton of $S^\star$ is $G^\star$, the dual graph of G.

A connected graph H is equipped with the path metric in which each edge is uniformly given length 1. The diameter of H is

$$\mathrm{Diam}(H) := \max\left\{ d(a,b) \mid a, b \in H^{(0)} \right\},$$

the maximum distance between vertices, and given a vertex u of H we define $\mathrm{Diam}_u(H)$ to be the maximum distance of vertices from u:

$$\mathrm{Diam}_u(H) := \max\left\{d(u,a) \mid a \in H^{(0)}\right\}.$$

Definitions 2.2. (Diagram measurements.) We define a number of measurements that capture aspects of the geometry of a diagram D with a base vertex v_0 in $\partial D^{(0)}$.

- The *area* $\mathrm{Area}(D)$ is the number of 2-cells in D.
- The *diameter* $\mathrm{Diam}(D) := \mathrm{Diam}_{v_0}(G)$ where $G = D^{(1)}$.
- The *perimeter* $\mathrm{Perimeter}(D)$ is the length of the boundary circuit of D.
- The *gallery length* $\mathrm{GL}(D)$ is the diameter of the dual graph $G^\star$.
- The *filling length* $\mathrm{FL}(D)$ is the minimum filling length $\mathrm{FL}(\mathcal{S})$ amongst all *shellings* $\mathcal{S}$ of D.

Note that it is conventional, on account of the group theoretic applications (see §4), to define the diameter to be the maximal distance of vertices from the base vertex rather than the maximal distance between two vertices. Obvious inequalities relate these two alternatives.

The definition of *filling length* requires further explanation. A *shelling* $\mathcal{S}$ of D is a *combinatorial null-homotopy* of D down to its base vertex v_0, formally defined as follows.

Definition 2.3. A *shelling* $\mathcal{S} = (D_i)$ of D is a sequence diagrams

$$D \;=\; D_0,\, D_1,\, \ldots,\, D_m \;=\; v_0,$$

in which each D_{i+1} is obtained from D_i by one of the *shelling moves* defined below and depicted in Figure 1. (The definitions of *1-cell* and *2-cell collapse* moves follow J. H. C. Whitehead; the reason for also using *1-cell expansions* arises in the group theoretic interpretation of filling length – see [10].)

A. *1-cell collapse*: remove a pair (e^1, e^0) where e^1 is a 1-cell with $e^0 \in \partial e^1$ and $e^0 \neq v_0$, and e^1 is attached to the rest of D_i only by one of its end vertices $\neq e^0$. (We call such an e^1 a *spike*.) This reduces the length of the boundary circuit by 2.

B. *1-cell expansion*: cut along some 1-cell e^1 in the interior of D_i that has a vertex e^0 in ∂D_i, in such a way that e^0 and e^1 are doubled. This increases the length of the boundary circuit by 2.

C. *2-cell collapse*: remove a pair (e^2, e^1) where e^2 is a 2-cell which has some edge $e^1 \in (\partial e^2 \cap \partial D_i)$. The effect on the boundary circuit is to replace e^1 with $\partial e^2 \smallsetminus e^1$.

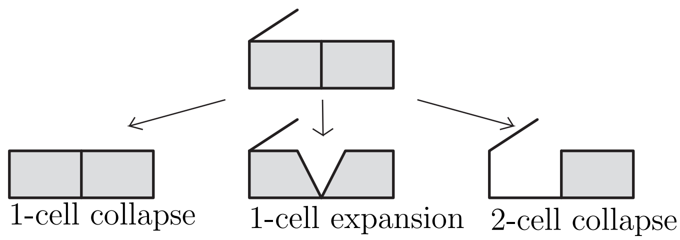

Figure 1: Shelling moves.

Note that when a *1-cell expansion* is performed with $e^0 = v_0$ in D_i a choice has to be made on account of the ambiguity over which of the two copies of e^0 in D_{i+1} should be the base vertex of D_{i+1}. An ambiguity also occurs when $D_i \smallsetminus e^0$ is not connected. Roughly speaking, we have to choose which side of the diagram to cut from, that is, where in the boundary circuit the extra two 1-cells will be inserted.

The *filling length* $\mathrm{FL}(\mathcal{S})$ *of a shelling* $\mathcal{S}$ is $\mathrm{FL}(\mathcal{S}) := \max_i \ell(\partial D_i)$, the maximum length of the boundary circuit in the course of the shelling. And the *filling length* $\mathrm{FL}(D)$ *of a diagram* D is the minimum of $\mathrm{FL}(\mathcal{S})$ as $\mathcal{S}$ ranges over all shellings of D.

2.2 Inequalities relating diagram measurements

Below we present a number of inequalities between the various measurements on a diagram D. As before $D = S \smallsetminus e_\infty$ and $G = D^{(1)}$. Let $e_\infty^\star$ denote the vertex of $G^\star$ dual to e_∞. The *valence* λ of a vertex v in G is the number of connected components of $(G \smallsetminus v) \cap B_v(\varepsilon)$, the intersection of $G \smallsetminus v$ with a small neighbourhood of v. Equivalently, λ is the length of the boundary circuit of the face $v^\star$ dual to v. (This is different from the number of edges in the boundary of the face in the event that there is an edge in G that forms a loop based at v.) In particular, $\mathrm{Perimeter}(D)$ is the valence in $G^\star$ of the vertex $e_\infty^\star$ dual to the 2-cell e_∞.

Proposition 2.4. *Suppose B is an upper bound on the valences of the vertexes in $G^\star \smallsetminus e^\star_\infty$, that is, on the lengths of the boundary circuits of the 2-cell of D. Define $n :=$ Perimeter(D). Then*

$$\begin{aligned} \mathrm{Diam}(D) &\leq B\,\mathrm{GL}(D) + n/2, & (1)\\ \mathrm{Diam}(D) &\leq \mathrm{FL}(D)/2, & (2)\\ \mathrm{GL}(D) &\leq \mathrm{Area}(D), & (3)\\ \mathrm{FL}(D) &\leq 2\,B\,\mathrm{Area}(D) + n, & (4)\\ \mathrm{Area}(D) &\leq n(B-1)^{\mathrm{GL}(D)}. & (5) \end{aligned}$$

Proofs.

(1). Suppose u is a vertex of G in the interior of $D^{(1)}$. Then there is a vertex $a^\star$ of $G^\star$ such that u is in the boundary of the dual 2-cell a. Let γ be a geodesic edge-path in $G^\star$ from $a^\star$ to $e^\star_\infty$ and let $\hat{\gamma}$ be the edge-path consisting of all but the final edge of γ. The number of vertices of $G^\star$ along the length of $\hat{\gamma}$ is at most $\mathrm{Diam}_{e^\star_\infty}(G^\star)$. Let $C(\hat{\gamma})$ be the subcomplex of D made up of all the 2-cells dual to vertices along $\hat{\gamma}$. There is a path from u to ∂D in the boundary of $C(\hat{\gamma})$. This has length at most $B\,\mathrm{Diam}_{e^\star_\infty}(G^\star)$, an upper bound on the total number of 1-cells in $C(\hat{\gamma})$. It follows that the distance in G of any vertex to ∂D is at most $B\,\mathrm{GL}(D)$. The fact that it is possible to reach the base vertex v_0 from any point on ∂D by traversing at most half the length of ∂D accounts for the remaining $n/2$ term in (1).

(2). Suppose v is a vertex of G. Let $\mathcal{S} = (D_i)$ be a shelling of D with $\mathrm{FL}(D) = \max_i \ell(\partial D_i)$. For each i there is a natural combinatorial map $(D_i, v_0) \twoheadrightarrow (D, v_0)$, and for some i the image of ∂D_i is an edge circuit in $D^{(1)}$ that passes through v. This edge circuit has length at most $\mathrm{FL}(D)$ and provides two paths in $D^{(1)}$ from v to v_0.

(3). This is straight-forward.

(4). It suffices to note that for any shelling $\mathcal{S} = (D_i)$ of D the length of each boundary circuit ∂D_i is at most twice the number of edges in D.

(5). Let $T^\star$ be a maximal geodesic tree in $G^\star$ based at $e^\star_\infty$. The valence of $e^\star_\infty$ is at most n and every other vertex has valence at most B. So the total number of vertices in $T^\star$ is at most $n(B-1)^{\mathrm{Diam}_{e^\star_\infty}(G^\star)}$ and the result follows.

■

2.3 Complementary pairs of trees and DGL

Suppose G is a finite, connected, undirected graph embedded in the 2-sphere (whence G is planar). Given a maximal tree (that is, a spanning tree) T in G, we can define the *complementary* tree $T^\star$ in $G^\star$ to be the subgraph whose edges are duals of edges in $G \smallsetminus T$. We refer to $(T, T^\star)$ as a *complementary pair* of maximal trees because one easily checks that $T^\star$ is a maximal tree in $G^\star$ and that $T^{\star\star} = T$.

Thus we can define a new diagram measurement $\mathrm{DGL}(D)$ that gives both a diameter and a gallery length bound realisable *simultaneously* by a complementary pair of maximal trees.

Definition 2.5. Define $\mathrm{DGL}(D)$ for a diagram D by

$$\mathrm{DGL}(D) \;:=\; \min\left\{\mathrm{Diam}(T) + \mathrm{Diam}(T^\star) \mid T \text{ is a maximal tree in } D^{(1)}\right\}.$$

3 The filling length of a diagram

In §5.C. of [13] Gromov asks the following geometric question about discs D with Riemannian metrics.

Question 3.1. Is it possible to contract the boundary loop ∂D in D by a homotop of length L bounded by

$$L \leq K\ \max\{\text{length}\ \partial D,\ \mathrm{Diam}\, D\}, \tag{6}$$

for some universal constant K?

In other words, Gromov is asking for a homotopy $H_t : [0,1] \to D$ with $H_t(0) = H_t(1) = v_0$ for all $t \in [0,1]$, and such that H_0 is the boundary loop, H_1 is the constant loop at a base point $v_0 \in \partial D$, and the length of the intermediate loops H_t satisfy the bound (6) for all $t \in [0,1]$. This question was quickly answered negatively by Frankel and Katz [6]. We pose a combinatorial analogue of Gromov's question replacing Riemannian discs by diagrams. In this context, controlling the length of a homotop means bounding the filling length FL as follows.

Question 3.2. Suppose D is a diagram and every 2-cell has boundary circuit length at most λ. Does there exist $K > 0$ depending only on λ such that

$$\mathrm{FL}(D) \;\leq\; K\,(\mathrm{Diam}(D) + \mathrm{Perimeter}(D)\,)\,? \tag{7}$$

This question also has a negative answer, as we shall now show, before giving a correct estimate on filling length. We present a family of diagrams D_n whose 2-cells each have boundary length at most three and for which $\mathrm{FL}(D_n) \to \infty$ as $n \to \infty$, and

$$\begin{array}{rclrcl} \mathrm{Diam}(D_n) & = & 0, & \qquad \mathrm{Perimeter}(D_n) & = & 1, \\ \mathrm{FL}(D_n) & \leq & n, & \mathrm{Area}(D_n) & = & 2^n - 1. \end{array}$$

It will follow that these diagrams do not admit some $K > 0$ such that $\mathrm{FL}(D_n) \leq K\,(\mathrm{Diam}(D_n) + \mathrm{Perimeter}(D_n)\,)$ for all n. The diagrams are defined inductively starting with a monogon D_1: one vertex and one edge forming a loop enclosing a 2-cell. The diagram D_{n+1} has only one vertex e^0, and its boundary is one edge e^1. When e^1 and the 2-cell e^2 of D_{n+1} with $e^1 \subset \partial e^2$ are removed the remaining diagram is two copies of D_n wedged together at e^0. We depict the diagram D_4 in Figure 2.

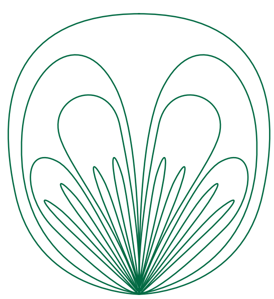

Figure 2: The diagram D_4.

The measurements of the diameter, perimeter and area of D_n given above are easily checked. The bound $\mathrm{FL}(D_n) \leq n$ on filling length follows from the result that $\mathrm{FL}(D_n) \leq 1 + \mathrm{FL}(D_{n-1})$ for all $n \geq 2$, which is obtained by shelling D_n by first removing the outermost 1-cell and 2-cell using a *2-cell collapse* move, revealing two copies of D_{n-1} which are then shelled one after the other. The result that $\mathrm{FL}(D_n) \to \infty$ is harder to prove.

Proposition 3.3. *The filling lengths of the diagrams D_n satisfy $\mathrm{FL}(D_n) \to \infty$ as $n \to \infty$.*

Proof. Let $\mathcal{S} = (C_i)$ be a shelling of D_n. There are natural combinatorial maps $C_i \to D_n$, which fail to be embeddings only when there are *1-cell expansion* moves in the shelling. Let $\overline{C}_i$ be the image of C_i in D_n. One can obtain $\overline{C}_i$ from C_i by identifying some of the boundary edges in C_i, so $\ell(\partial C_i) \geq \ell(\partial \overline{C}_i)$. (Note that edges in ∂C_i that are not part of the boundary of a 2-cell in $\overline{C}_i$ contribute 2 to $\ell(\partial C_i)$.)

Let n_{C_i} be the number of monogons in $\overline{C}_i$, that is the number of monogons in D_n that have not yet been collapsed when one reaches C_i in the course of the shelling $\mathcal{S}$. Then n_{C_i} is a monotone decreasing sequence starting at 2^{n-1} and terminating at 0.

Let $\mathcal{A}^{\pm} := \left\{ \pm 2^k \mid k \in \mathbb{N} \right\}$, a symmetric set of generators for the additive group $\mathbb{Z}$. For $m \in \mathbb{Z}$ define $\|m\|$ to be the distance of m from 0 in the word metric on $\mathbb{Z}$ with respect to $\mathcal{A}^{\pm}$. We will explain a means of reading off an expression for n_{C_i} as a sum of $\ell(\partial \overline{C}_i)$ elements of $\mathcal{A}^{\pm}$.

We refer to boundary edges of a 2-cell in D_n as either *upper* or *lower* according to the sense in which they are depicted in Figure 2. So each triangular 2-cell has one upper and two lower boundary edges, and each monogon just has an upper edge.

Each edge e in ∂D_n forms a loop and we label it by the number k_e of monogons contained within that loop; this is always a power of 2. We express n_{C_i} as a sum of elements of $\mathcal{A}^{\pm}$ as follows. Every edge in $\partial \overline{C}_i$ contributes a term k_e if e is not one of the *lower* boundary edges of a 2-cell in $\overline{C}_i$, and contributes $-k_e$ if it is not one of the *upper* boundary edges of a 2-cell in $\overline{C}_i$. (So edges e in 1-dimensional portions of $\overline{C}_i$ contribute terms k_e and $-k_e$.)

As an example, take D_n and remove the last monogon on the right, to give a subcomplex $\overline{C}$ that includes $n_{\overline{C}} = 2^{n-1} - 1$ monogons. There are two boundary edges in $\overline{C}$; the outer one contributes 2^{n-1} and the inner one -1 to express $n_{\overline{C}}$ as the sum $2^{n-1} - 1$.

In Appendix A we prove that $\| \ \|$ is unbounded. As

$$\mathrm{FL}(D_n) \geq \max \left\{ \|m\| \mid m \leq 2^{n-1} \right\}$$

we deduce that $\mathrm{FL}(D_n)$ is unbounded. ∎

Remark. If the *1-cell expansion* move is removed from the definition of a shelling then it becomes easy to prove that $\mathrm{FL}(D_n) = n$. However, as the definition stands it is not true that $\mathrm{FL}(D_n) = n$. For example one can prove that $\mathrm{FL}(D_5) \leq 4$.

We now amend the bound from Gromov's question by replacing the $\mathrm{Diam}(D)$ term in (7) by the diameter of a maximal tree T in $D^{(1)}$ and adding the diameter of the complementary tree $T^\star$ (defined in §2.3).

Proposition 3.4. *Suppose $D = S \smallsetminus e_\infty$ is a diagram in which the boundary of each 2-cell has at most λ edges. Then given any maximal tree T in $G = D^{(1)}$,*

$$\mathrm{FL}(D) \ \le \ \mathrm{Diam}(T) \, + \, 2\,\lambda\, \mathrm{Diam}(T^\star) \, + \, \mathrm{Perimeter}(D)\,. \tag{8}$$

Recalling the diagram measurement DGL of Definition 2.5, we see that Proposition 3.4 gives:

Theorem 3.5. *Fix $\lambda > 0$. There exists $K > 0$ such that for every diagram D in which the boundary of each 2-cell has at most λ edges,*

$$\mathrm{FL}(D) \le K\,(\mathrm{DGL}(D) + \mathrm{Perimeter}(D)\,).$$

Proof of Proposition 3.4. Take the vertex $e_\infty^\star$ dual to the 2-cell e_∞ *at infinity* to be the base vertex of the maximal tree $T^\star$ in the dual graph $G^\star$. As usual D is equipped with a base vertex v_0 on its boundary.

Let m be the total number of edges in T and $\gamma : [0, 2m] \to T$ be the edge-circuit in T based at v_0 that traverses every edge of T twice, once in each direction, and runs close to the anticlockwise boundary loop of a small neighbourhood of T. For $i = 1, 2, \ldots, 2m$ let γ_i be the edge of T that is the image of $\gamma\,\big|_{[i-1,i]}$, and regard this edge as directed from $\gamma(i-1)$ to $\gamma(i)$. Let τ_i be the geodesic in $T^\star$ from $e_\infty^\star$ to the vertex $e_i^\star$ dual to the 2-cell e_i on the right of γ_i (which may be e_∞). Let $\overline{\tau_i}$ be the union of all the (closed) 2-cells dual to vertices on τ_i.

There may be some 2-cells e^2 in $D \smallsetminus \bigcup_i \overline{\tau_i}$. For such an e^2, let ϵ be the edge of $(\partial e^2) \smallsetminus T$ whose dual edge is closest in $T^\star$ to $e_\infty^\star$. Call such an edge ϵ *stray*. Then ϵ must be an edge-loop because otherwise the subdiagram with boundary circuit made up of ϵ and the geodesics in T from the end vertices of ϵ to $\star$ would contain some e_i^*. It follows that such an ϵ is the boundary of a subdiagram that is of the form of Figure 2 save that shelling away a 2-cell reveals at most $(\lambda - 1)$ more 2-cells instead of two. The shelling discussed in Section 3 generalises to show that such a diagram has filling length at most $(\lambda - 1)\mathrm{Diam}(T^\star)$.

To realise (8), shell the 2-cells of D in the following order: for $i = 1$, then $i = 2$, and so on, shell all the (remaining) 2-cells along $\overline{\tau_i}$, working from

$\star_\infty$ to v_i^*, doing 2-cell collapses to remove successive edges dual to edges on τ_i and the 2-cells they border. Except, whenever a *stray* edge e appears in the boundary circuit, pause and entirely shell away the diagram it contains. In the course of the 2-cell collapses, do a *1-cell-collapse* whenever one is available.

One checks that in the course of this shelling, aside from detours into sub-diagrams enclosed by stray edges, the anticlockwise boundary-circuit starting from v_0 follows a geodesic path in T, then a path embedded in the 1-skeleton of some $\overline{\tau_i}$ before returning to v_0 along ∂D. These three portions of the circuit have lengths at most $\mathrm{Diam}(T)$, $\lambda\,\mathrm{Diam}(T^\star)$ and $\ell(\partial D)$, respectively. Adding $(\lambda-1)\mathrm{Diam}(T^\star)$ for the detour gives an estimate within the asserted bound. ■

In [11] we discuss a number of inter-related conjectures about complementary pairs of trees in finite planar graphs. Conjecture 4.3 claims that GL and DGL are *essentially the same* diagram measurement when one is concerned with diagrams for which there is a uniform bound on the length of the boundary cycles of each 2-cell (with the possible exception of the 2-cell *at infinity*). If this conjecture is true then the theorem simplifies as follows.

Corollary 3.6. *Assuming Conjecture 4.3 of [11] holds, the conclusion of Theorem 3.5 can be replaced by*

$$\mathrm{FL}(D) \ \leq \ K\,(\mathrm{GL}(D) + \mathrm{Perimeter}(D)\,).$$

4 Filling functions for finite presentations

4.1 Definitions

Filling functions for a finite presentation $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ of a group Γ are defined using measurements of van Kampen diagrams D_w (defined below) associated to words w that represent 1 in Γ. These diagrams demonstrate graphically how $w = 1$ is a consequence of the relations $\mathcal{R}$. (Words are strings of letters in $\mathcal{A}$ and their formal inverses, that is, elements of the free monoid $(\mathcal{A} \cup \mathcal{A}^{-1})^\star$.) Therefore one might say that a filling function *captures an aspect of the geometry of the word problem for* $\mathcal{P}$.

Denote the length of w by $\ell(w)$. Let $\mathcal{K}^2$ be the compact 2-complex with fundamental group Γ associated to $\mathcal{P}$: to construct $\mathcal{K}^2$ take a wedge of $|\mathcal{A}|$

circles, label each circle by an element of $\mathcal{A}$ and orient them, and then attach a $\ell(r)$-sided 2-cell for each $r \in \mathcal{R}$ with r describing its attaching map. Let $C(\mathcal{P}) = \widetilde{\mathcal{K}^2}$ denote the *Cayley 2-complex* associated to $\mathcal{P}$. The *Cayley graph* of $\mathcal{P}$ is the 1-skeleton of $C(\mathcal{P})$, and the 0-skeleton is identified with Γ so that the combinatorial metric on $C(\mathcal{P})$ agrees with the word metric $d_{\mathcal{P}}$ on Γ. Each edge of the Cayley graph inherits an orientation from $\mathcal{K}^2$ as well as a label by an element of $\mathcal{A}$.

A word w in Γ such that $w = 1$ in Γ is said to be *null-homotopic*, or is referred to as an *edge-circuit* because it defines a loop in $C(\mathcal{P})^{(1)}$ based at 1 (say).

Definition 4.1. Suppose w is a null-homotopic word. Then a diagram $D_w = S \smallsetminus e_\infty$ with base vertex v_0 is a *van Kampen diagram* for w when there is a combinatorial map $\Phi : (D_w, v_0) \to (C(\mathcal{P}), 1)$ such that $\Phi \mid_{\partial D_w}$ is the edge-circuit w. (A map between complexes is *combinatorial* if for all n it sends n-cells homeomorphically onto n-cells.)

Each edge of D_w inherits a direction from its image in $C(\mathcal{P})^{(1)}$ and also a label by an element of $\mathcal{A}$. So the boundary label of each 2-cell of D_w is a cyclic conjugate of an element of $\mathcal{R}^{\pm 1}$ and, starting at the base vertex v_0, one reads w (by convention anticlockwise) around ∂D.

For an edge-circuit w define $\mathrm{Area}(w)$, $\mathrm{Diam}(w)$, $\mathrm{GL}(w)$, $\mathrm{DGL}(w)$ and $\mathrm{FL}(w)$ by

$$\mathrm{M}(w) \ := \ \min \{ \mathrm{M}(D_w) \mid \ \text{van Kampen diagrams } D_w \text{ for } w \},$$

where M is Area, Diam, GL, DGL and FL, respectively.

We mention some equivalent definitions. It is a consequence of *van Kampen's lemma* [2, 14, 15] that $\mathrm{Area}(w)$ is the least N such that there is an equality

$$w = \prod_{i=1}^{N} u_i^{-1} r_i u_i \tag{9}$$

in the free group $F(\mathcal{A})$ for some $r_i \in R^{\pm 1}$ and $u_i \in \mathcal{A}^\star$. Similarly, modulo an error of $\pm \max \{\ell(r) \mid r \in R\}$, the diameter $\mathrm{Diam}(w)$ is the minimal bound on the length of the conjugating elements u_i in equalities (9). Proposition 1 in [10] says that $\mathrm{FL}(w)$ is the minimal bound on the length of words one encounters in the process of *applying relators* to reduce w to the empty word.

Definition 4.2. (Filling functions.) We define

- the *Dehn function* $\mathrm{Area} : \mathbb{N} \to \mathbb{N}$,
- the *minimal isodiametric function* $\mathrm{Diam} : \mathbb{N} \to \mathbb{N}$,
- the *gallery length function* $\mathrm{GL} : \mathbb{N} \to \mathbb{N}$,
- the *filling length function* $\mathrm{FL} : \mathbb{N} \to \mathbb{N}$,
- and $\mathrm{DGL} : \mathbb{N} \to \mathbb{N}$

of $\mathcal{P}$ by

$$\mathrm{M}(n) \;:= \max \left\{ \mathrm{M}(w) \mid \text{edge-circuits } w \text{ with } \ell(w) \leq n \right\},$$

where M is Area, Diam, GL, DGL and FL, respectively.

An *isoperimetric function* (resp. *isodiametric function*) for $\mathcal{P}$ is any $f : \mathbb{N} \to \mathbb{N}$ such that $\mathrm{Area}(n) \leq f(n)$ (resp. $\mathrm{Diam}(n) \leq f(n)$) for all n.

There are many references to isoperimetric functions, Dehn functions and isodiametric functions in the literature; [2, 8] are surveys. The filling length function is discussed extensively in [10] and has an important application in [9]. We introduce *gallery length* GL and DGL in this article.

The word problem for $\mathcal{P}$ is solvable if and only if any one (and hence all) of the filling functions Area, Diam, GL, DGL and FL is bounded by a recursive function – see [8].

4.2 Changing the presentation

It is important to note that Area, Diam, GL and FL were all defined for a fixed finite presentation $\mathcal{P}$ of Γ. However Area, Diam and FL are group invariants in the sense that each is well behaved under change of finite presentation. The Dehn functions $\mathrm{Area}_{\mathcal{P}}$ and $\mathrm{Area}_{\mathcal{Q}}$ of two finite presentations $\mathcal{P}$ and $\mathcal{Q}$ of the same group G are $\simeq$-equivalent in the sense defined below. The same can be said of the minimal isodiametric function and of the filling length function. Proofs can be found in [10] in the case of FL and in [12] for Area and Diam.

Definition 4.3. ($\simeq$-equivalence.) For two functions $f, g : \mathbb{N} \to \mathbb{N}$ we say that $f \preceq g$ when there exists $C > 0$ such that $f(n) \leq C g(Cn + C) + Cn + C$ for all n, and we say $f \simeq g$ if and only if $f \preceq g$ and $g \preceq f$.

The situation for GL is more complicated. We need the following definition.

Definition 4.4. (Fat presentations.) One obtains a *fat* presentation $\mathcal{P}' = \langle \mathcal{A}' \mid \mathcal{R}' \rangle$ from a presentation $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ by adjoining an extra generator to $\mathcal{A}$ and extra relations to $\mathcal{R}$ as follows:

$$\mathcal{P}' \ := \ \langle \mathcal{A} \cup \{z\} \mid \mathcal{R} \cup \left\{z, z^2, zz^{-1}, z^3, z^2 z^{-1}\right\} \cup \{[a, z] : a \in \mathcal{A} \cup \{z\}\} \rangle.$$

The added generator z is spurious in so far as $\mathcal{P}$ and $\mathcal{P}'$ present the same group Γ and z represents 1 in Γ. However z will play the following important role in the proof of Theorem 4.5 below and will also be used in §6 and in [11]. If w is an edge-circuit in $\mathcal{P}'$ then we can construct a planar 2-complex A_w that is topologically an annulus, and more specifically is a ring of 2-cells that have boundary labels the commutator relations amidst $\{[a, z]^{\pm 1} \mid a \in \mathcal{A} \cup \{z\} \}$. All the edges in the interior are labelled by z, and both inner and outer anticlockwise boundary labels are w. So if D_w is a van Kampen diagram for w then we can create a larger van Kampen diagram $A_w \cup D_w$ for w by attaching A_w around ∂D_w. Refer to A_w as a *z-collar for w*. The salient fact is that $A_w \cup D_w$ is necessarily a topological disc (whilst in general D_w may not be).

Theorem 4.5. *Suppose that $\mathcal{P}$ and $\mathcal{Q}$ are two finite presentations for the same group Γ. Then the gallery length functions $\mathrm{GL}_{\mathcal{P}'}$ and $\mathrm{GL}_{\mathcal{Q}'}$ of the fat presentations $\mathcal{P}'$ and $\mathcal{Q}'$ are $\simeq$-equivalent.*

Proof. Recall that Tietze's theorem [14] states that if $\mathcal{P}$ and $\mathcal{Q}$ are two finite presentations for the same group then there is a finite sequence of Tietze operations starting at $\mathcal{P}$ and terminating at $\mathcal{Q}$. Suppose $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$. There are four types of Tietze operations $\mathcal{P} \to \mathcal{Q}$: Types I and II, and their inverses –

Type I. Adjoin a new free generator t and a new relator tu^{-1}, where $u \in \mathcal{A}^\star$ (so u does not involve t).

Type II. Adjoin a new relator r which is a consequence of the relations $\mathcal{R}$.

It suffices to check that for all Type I or Type II operations $\mathcal{P} \to \mathcal{Q}$ the gallery length functions $\mathrm{GL}_{\mathcal{P}'}$ and $\mathrm{GL}_{\mathcal{Q}'}$ of the fat presentations $\mathcal{P}'$ and $\mathcal{Q}'$ are $\simeq$-equivalent.

Consider first the Type I operation $\mathcal{P} \to \mathcal{Q}$. In the argument that follows we will assume that u is not the empty word. A separate and easier argument must be given if u is empty, which we omit.

Suppose first that w is a null-homotopic word in the generators of $\mathcal{Q}'$. The length of w increases by at most a constant factor $A := \ell(u)$ when one replaces all occurrences of t in S by u. Thus by at most $\ell(w)$ applications of a relator we obtain a new word $\hat{w}$ in the generators of $\mathcal{P}'$. Now take a van Kampen diagram $D_{\hat{w}} = S_{\hat{w}} \setminus e_\infty$ for $\hat{w}$ over $\mathcal{P}'$ with dual graph $G_{\hat{w}}{}^\star$ satisfying $\mathrm{GL}_{\mathcal{P}'}(\hat{w}) = \mathrm{Diam}(G_{\hat{w}}{}^\star)$. Obtain a van Kampen diagram D_w for w over $\mathcal{Q}'$ by attaching 2-cells labelled by tu^{-1} or ut^{-1} to $\partial D_{\hat{w}}$. The effect on the dual graph is to increase the distance of any given vertex to $e_\infty^\star$ by 1, and therefore $\mathrm{GL}_{\mathcal{Q}'}(w) \leq \mathrm{GL}_{\mathcal{P}'}(\hat{w}) + 2$. So $\mathrm{GL}_{\mathcal{Q}'}(n) \leq \mathrm{GL}_{\mathcal{P}'}(A\,n) + 2$ for all n.

Next suppose that w is a null-homotopic word in the generators of $\mathcal{P}'$. So there is no t in w. Let $D_w = S_w \setminus e_\infty$ be a van Kampen diagram over $\mathcal{Q}'$ for w for which $\mathrm{GL}_{\mathcal{Q}'}(D_w) = \mathrm{GL}_{\mathcal{Q}'}(w)$. The relations in $\mathcal{Q}'$ that involve t are tu^{-1} and $[t^{\pm 1}, z]$. It follows that all occurrences of t in D_w must be in "t-corridors" – concatenations of 2-cells with boundary labels $[t, z]^{\pm 1}$ and with adjacent 2-cells joined across an edge labelled t. There are two types of t-corridor: *linear* t-corridors start with a 2-cell labelled by $(tu^{-1})^{\pm 1}$ and finish with a 2-cell labelled $(tu^{-1})^{\pm 1}$ (see Figure 3), and *annular* t-corridors close up to form annuli.

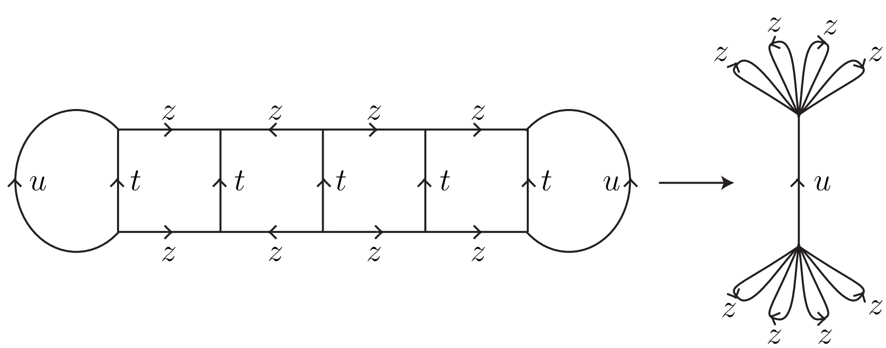

Figure 3: Collapsing a t-corridor.

We want to eradicate all instances of t from D_w and arrive at a diagram over $\mathcal{P}'$. In the case of the annular t-corridors we can simply change the edges labelled by t to edges labelled by z because $[z, z]$ is one of the defining relators in $\mathcal{P}'$. In the case of the linear t-corridors we collapse the diagrams in the way illustrated in Figure 3. However the gallery length could be increased on account of possible loses of efficient roots through the dual graph of D_w that run along linear t-corridors. To deal with this problem we insert z-collars around the linear t-corridors before collapsing them: we cut along the boundary edge-circuit of each linear t-corridors and we glue in a z-collar. (It may be that this edge-circuit does not embed in D_w, however this does not represent an obstruction to inserting a collar.) The effect of gluing in the z-collars is to increase the gallery length by at most a factor of two. Then on collapsing the linear t-corridors we get a van Kampen diagram for w over $\mathcal{P}'$ and the gallery length does not increase further. So $\mathrm{GL}_{\mathcal{P}'}(n) \leq 2\,\mathrm{GL}_{\mathcal{Q}'}(n)$.

Next assume that $\mathcal{P} \to \mathcal{Q}$ is a type II Tietze operation, where we adjoin a new relator r. It is clear that $\mathrm{GL}_{\mathcal{Q}'} \leq \mathrm{GL}_{\mathcal{P}'}$ because every van Kampen diagram over $\mathcal{P}'$ is *a fortiori* a van Kampen diagram over $\mathcal{Q}'$.

For the bound in the other direction let D_w be a van Kampen diagram over $\mathcal{Q}'$ for an edge-circuit of length n. The word r is an edge-circuit in $\mathcal{P}'$ and so has a van Kampen diagram D_r over $\mathcal{P}'$. It will be important that this diagram is a topological disc and this can be ensured by gluing attaching a z-collar for r around its boundary if necessary. Let B be the diameter of the dual graph to D_r. Now let $D_{r^{-1}}$ be the van Kampen diagram for r^{-1} over $\mathcal{P}'$ obtained by reflecting D_r. Cut out of D_w every 2-cell that is labelled by r or by r^{-1} and glue in copies of D_r and $D_{r^{-1}}$ respectively. Call the resulting diagram $\hat{D}_w$. That $\mathrm{GL}(\hat{D}_w) \leq B\,\mathrm{GL}(D_w)$ follows from the fact that D_r is a topological disc. ∎

In fact, an analysis of the change in areas of the diagrams in the proof above allow us to see how *simultaneously realisable* area and gallery length bounds are affected by a change in presentation. The following definition provides the language to make this precise.

Definition 4.6. We say that a pair (f, g) of functions $f,\, g : \mathbb{N} \to \mathbb{N}$ is an $(\mathrm{Area}, \mathrm{GL})$-*pair* for the finite presentation $\mathcal{P}$ when, for every edge-circuit w in the Cayley graph of $\mathcal{P}$, there exists a van Kampen diagram D_w with $\mathrm{Area}(D_w) \leq f(\ell(w))$ and $\mathrm{GL}(D_w) \leq g(\ell(w))$.

Scholium 4.7. *Suppose that $\mathcal{P}$ and $\mathcal{Q}$ are two finite presentations for the same group Γ, and that $\mathcal{P}'$ and $\mathcal{Q}'$ are their fattenings. If $(f_{\mathcal{P}'}, g_{\mathcal{P}'})$ is an $(\mathrm{Area}, \mathrm{GL})$-pair for $\mathcal{P}'$ then there is an $(\mathrm{Area}, \mathrm{GL})$-pair $(f_{\mathcal{Q}'}, g_{\mathcal{Q}'})$ for $\mathcal{Q}'$ such that $f_{\mathcal{P}'} \simeq f_{\mathcal{Q}'}$ and $g_{\mathcal{P}'} \simeq g_{\mathcal{Q}'}$.*

5 The Double Exponential Theorem

In this section we use gallery length to give a new proof of a theorem of D.E.Cohen [3] known as *the Double Exponential Theorem*, that is a cornerstone amongst results about how filling functions interrelate. It asserts that the Dehn function of any finite presentation $\mathcal{P}$ of a group is bounded above by a double exponential in any isodiametric function for $\mathcal{P}$. Cohen's proof involves an analysis of the Nielsen reduction process. A further proof, using *Stallings folds*, was given by the first author in [7].

The essence of our new proof is that there is at most an exponential leap from the gallery length filling function up to the Dehn function, and from the minimal isodiametric function up to the gallery length.

Theorem 5.1. *Let $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ be a finite presentation and w be an edge-circuit in its Cayley graph. Suppose D_w is a minimal diameter van Kampen diagram for w, and moreover is of minimal area amongst all minimal diameter diagrams for w. Then there exists $A > 0$ depending only on $\mathcal{P}$, such that*

$$\mathrm{GL}(D_w) \quad \leq \quad A^{1 \,+\, 2\,\mathrm{Diam}(D_w)}.$$

It follows that $\mathrm{GL}(n) \leq A^{1 \,+\, 2\,\mathrm{Diam}(n)}$ for all n.

Proof. Let v_0 denote the base vertex of D_w and let T be a maximal geodesic tree in $G := D_w^{(1)}$ based at v_0. To every edge e in $G \smallsetminus T$ we can associate an anticlockwise edge-circuit γ_e in G based at v_0 by connecting the vertices of e to v_0 by geodesic paths in T. Let w_e be the word one reads along γ_e. Then $\ell(w_e) \leq 1 + 2\,\mathrm{Diam}(D_w)$.

Let $T^\star$ be the complementary tree to T (as defined in §2.3). Let $e_\infty^\star$ be the vertex of $T^\star$ dual to the face e_∞.

Suppose that e_1 and e_2 are two distinct edges of $G \smallsetminus T$ dual to edges $e_1^\star$ and $e_2^\star$ of $T^\star$ on some geodesic in $T^\star$ from $e_\infty^\star$ to a leaf. For definiteness assume $e_2^\star$ is further from $e_\infty^\star$ along the geodesic than $e_1^\star$. Then γ_{e_1} and γ_{e_2} bound subcomplexes $C(\gamma_{e_1})$ and $C(\gamma_{e_2})$ of D_w in such a way that $C(\gamma_{e_1})$ is a subcomplex of $C(\gamma_{e_2})$. Suppose that the words w_{e_1} and w_{e_2} are the same. Then we could cut $C(\gamma_{e_2})$ out of D_w and glue $C(\gamma_{e_1})$ in its place, producing a new van Kampen diagram for w with strictly smaller area and no increase in diameter. This would be a contradiction.

Now the number of distinct words of length at most $1 + 2\,\mathrm{Diam}(D_w)$ is fewer than

$$(2\,|\mathcal{A}| + 1)^{1\,+\,2\,\mathrm{Diam}(D_w)}. \tag{10}$$

Therefore there exists $A > 0$ such that

$$\mathrm{GL}(D_w) \;\leq\; 2\,\mathrm{Diam}_{e^\star_\infty}(T^\star) \;\leq\; A^{1\,+\,2\,\mathrm{Diam}(D_w)}.$$

∎

Proposition 5.2. *Let $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ be a finite presentation and w be an edge-circuit in its Cayley graph. Define $B := \max\{\ell(r) \mid r \in \mathcal{R}\}$. If D_w is a van Kampen diagram for w then*

$$\mathrm{Area}(D_w) \leq n(B-1)^{\mathrm{GL}(D_w)}, \tag{11}$$

where $n := \ell(\partial D_w)$. It follows that $\mathrm{Area}(n) \leq n(B-1)^{\mathrm{GL}(n)}$ for all n.

Proof. Note that B is an upper bound on the number of 1-cells in the boundary of each 2-cell in van Kampen diagram D_w over $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$. So inequality (5) of Proposition 2.4 gives us (5.2). ∎

Theorem 5.1 and Proposition 5.2 combine to give:

Theorem 5.3 (The Double Exponential Theorem). *Given a finite presentation $\mathcal{P}$ there exists $C > 0$ such that the Dehn function $\mathrm{Area}(n)$ and the minimal isodiametric function $\mathrm{Diam}(n)$ for $\mathcal{P}$ satisfy*

$$\mathrm{Area}(n) \leq n\, C^{C^{\mathrm{Diam}(n)}}$$

for all $n \in \mathbb{N}$.

Scholium 5.4. *Given a finite presentation $\mathcal{P}$ there exists a constant $C > 0$ such that for all edge-circuit w in $\mathcal{P}$ there exists a van Kampen diagram D_w with*

$$\begin{aligned} \mathrm{Area}(D_w) &\leq n\, C^{C^{\mathrm{Diam}(n)}}, \\ \mathrm{GL}(D_w) &\leq C^{\mathrm{Diam}(n)}, \\ \mathrm{Diam}(D_w) &\leq \mathrm{Diam}(n), \end{aligned}$$

where $n := \ell(w)$.

Open questions 5.5.

5.5.1 **(The single exponential problem.)** Given a finite presentation $\mathcal{P}$, does there exist a constant $C_{\mathcal{P}} > 0$, such that ${C_{\mathcal{P}}}^{\mathrm{Diam}(n) + n}$ is an isoperimetric function for $\mathcal{P}$?

5.5.2 **(The strong form of the single exponential problem.)** Given any finite presentation $\mathcal{P}$, does there exist a constant $C_{\mathcal{P}} > 0$, such that $\left({C_{\mathcal{P}}}^{\mathrm{Diam}(n) + n}, \mathrm{Diam}(n)\right)$ is an $(\mathrm{Area}, \mathrm{Diam})$-function for $\mathcal{P}$? (Here we use the notation of Definition 4.6 with Diam in place of GL.)

5.5.3 Given any finite presentation $\mathcal{P}$, does there exist a constant $C_{\mathcal{P}} > 0$ such that $\mathrm{FL}(n) \leq C_{\mathcal{P}}(\mathrm{Diam}(n) + n)$ for all $n \in \mathbb{N}$?

5.5.4 One can ask 5.5.3 with GL in place of FL.

Question 5.5.1 was, to our knowledge, first raised by Stallings. Question 5.5.3 was asked by Gromov [13, page 100].

In §3 we constructed a family of diagrams D_n to show that the bound of 5.5.3 fails on the level of diagrams. We cannot immediately deduce a negative answer to 5.5.3 because, *a priori*, it may not be possible to realise the behaviour exhibited by the diagrams D_n in minimal filling length van Kampen diagrams for some sequence of null homotopic words in some finite presentation $\mathcal{P}$ – indeed it may be that minimal diameter and minimal filling length diagrams for a given word are different.

Gromov points out in [13, §5.C] that an affirmative answer to 5.5.3 would imply an affirmative answer to 5.5.1 (and, in fact, to 5.5.2) since it is known that a single exponential of filling length provides a simultaneously realisable isoperimetric function for $\mathcal{P}$:

Proposition 5.6. *Let $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ be a finite presentation of a group. Let $K := 2\,|\mathcal{A}| + 1$. Then $\mathrm{Area}(n) \leq K^{\mathrm{FL}(n)}$ for all n. Moreover, if D is a van Kampen diagram for an edge-circuit w in $\mathcal{P}$ and has $\mathrm{FL}(D) \leq \mathrm{FL}(w)$, and in addition is of minimal area amongst all such diagrams then $\mathrm{Area}(D) \leq K^{\mathrm{FL}(D)}$.*

This proposition is the "*space-time*" bound of [10, Corollary 2] or [13, 5.C]. The fact that $K^{\mathrm{FL}(n)}$ is an upper bound on the number of words of length at most $\mathrm{FL}(n)$ is at the heart of its proof.

The following theorem says firstly that if (5.5.4) has a positive answer then so does (5.5.2), and secondly provides a near converse.

Theorem 5.7. *Let $\mathcal{P}$ be a finite presentation. Suppose that there is a constant $C_{\mathcal{P}}$ such that* $\mathrm{GL}(n) \leq C_{\mathcal{P}}(\mathrm{Diam}(n) + n)$ *for all* $n \in \mathbb{N}$. *Then*

$$\Pi \;:=\; \left({C_{\mathcal{P}}}^{\mathrm{Diam}(n) + n}, \mathrm{Diam}(n) \right)$$

is an $(\mathrm{Area}, \mathrm{Diam})$*-pair for* $\mathcal{P}$.

Conversely if Π *is an* $(\mathrm{Area}, \mathrm{Diam})$*-pair for* $\mathcal{P}$ *for some constant* $C_{\mathcal{P}}$ *then* $\mathrm{GL}(n) \preceq \mathrm{Diam}(n)$.

Proof. The first part of this theorem follows from Proposition 5.2.

The second part is a consequence of Theorem 7.1 of [11], which says that for *fat* finite presentations, GL is $\simeq$-equivalent to another filling function called DlogA. Let $\mathcal{P}'$ be the fattening (in the sense of Definition 4.4) of $\mathcal{P}$. Then, with a similar proof to Scholium 4.7, there is an $(\mathrm{Area}, \mathrm{Diam})$-pair $(f(n), g(n))$ for $\mathcal{P}'$ for some $f, g : \mathbb{N} \to \mathbb{N}$ satisfying $f(n) \simeq {C_{\mathcal{P}}}^{\mathrm{Diam}(n) + n}$ and $g(n) \simeq \mathrm{Diam}(n)$. Now DlogA concerns the sum of the logarithm of the area of a diagram with its diameter, and in this case is therefore

$$\mathrm{DlogA}(n) \preceq g(n) + \log f(n) \preceq \mathrm{Diam}(n).$$

∎

We conjecture that 5.5.3 and 5.5.4 are, in fact, equivalent on account of a close relationship between FL and GL that is discussed further in §7 and in [11].

Remark 5.8. The group theoretic input to this proof of the Double Exponential Theorem is that the bound (10) on the number of words of length at most $1 + 2\,\mathrm{Diam}(w)$. This amounts to using the finite valence of the 1-skeleton of the Cayley graph to obtain a bound on the number of distinct edge-paths in the Cayley graph that have at most this length and emanate from a given vertex. Thus the methods above can be used to prove the Double Exponential Theorem in the more general setting of filling edge-circuits in a simply connected 2-complex with uniformly bounded geometry (that is, a uniform bound on the number of vertices, edges or face that are incident each given vertex, edge or face) – see [16], where Papasoglu uses the method of *Stallings' folds* to give a proof of the double exponential in this more general context.

6 The gallery length of combable groups

In this section we give an upper bound on the gallery length function of combable groups as well as a simultaneously realisable isoperimetric function. This is a wide and much studied class of groups that includes all automatic groups [5].

We recall the relevant definitions (*cf.* [1] or [8]). A *combing* $\sigma : \Gamma \to (\mathcal{A} \cup \mathcal{A}^{-1})^{\star}$ is a section of the natural surjection $(\mathcal{A} \cup \mathcal{A}^{-1})^{\star} \twoheadrightarrow \Gamma$. Denote the image of $g \in \Gamma$ by σ_g and view this as a continuous path $[0, \infty) \to C(\Gamma, \mathcal{A})$ in the Cayley graph from the identity to g, moving at a constant speed from the identity for time $\ell(\sigma_g)$ before stopping at g. We refer to σ_g as the *combing line* of g. The *length function* $L : \mathbb{N} \to \mathbb{N}$ of σ is defined by:

$$L(n) := \max \{\ell(\sigma_g) \mid d(1, g) \leq n\} .$$

We say that σ satisfies the *synchronous k-fellow-traveller property* when

$$\forall g, h \in \Gamma, \quad (d(g, h) = 1 \implies \forall t \in \mathbb{N}, \; d(\, \sigma_g(t), \sigma_h(t)\,) \; \leq \; k)$$

Define a *reparametrisation* ρ to be an unbounded function $\mathbb{N} \to \mathbb{N}$ such that $\rho(0) = 0$ and $\rho(n+1) \in \{\rho(n), \rho(n)+1\}$ for all n. Then say that σ satisfies the *asynchronous k-fellow traveller property* when for all $g, h \in \Gamma$ with $d(g, h) = 1$, there exist reparametrisations ρ_g and ρ_h such that

$$\forall t \in \mathbb{N}, \quad d(\, \sigma_g(\rho_g(t)), \sigma_h(\rho_h(t))\,) \leq k.$$

Our result is phrased in terms of an $(\mathrm{Area}, \mathrm{GL})$-pair – see Definition 4.6.

Proposition 6.1. *Suppose a group Γ with finite generating set $\mathcal{A}$ admits a combing σ that satisfies a (synchronous or asynchronous) k-fellow-traveller property. Let $L : \mathbb{N} \to \mathbb{N}$ be the length function of σ. Then there exists a finite presentation $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ for Γ for which $(n\, L(n), L(n)\,)$ is an $(\mathrm{Area}, \mathrm{GL})$-pair for the fattened presentation $\mathcal{P}'$, up to a common multiplicative constant. (That is, there exists some $\mu > 0$ that depends only on $\mathcal{P}$ such that $(\mu\, n\, L(n), \; \mu\, L(n)\,)$ is an $(\mathrm{Area}, \mathrm{GL})$-pair for $\mathcal{P}'$.)*

In particular, if L admits a linear bound then so does GL, and (n^2, n) is an $(\mathrm{Area}, \mathrm{GL})$-pair for $\mathcal{P}'$, up to a common multiplicative constant.

Proof. Suppose σ is a combing of Γ that satisfies the synchronous k-fellow-traveller property. For an edge-circuit w construct the *cockleshell diagram*

D_w: start with a loop of 1-cells labelled by w reading anticlockwise from the base vertex $\star$ and then join each vertex to $\star$ by a combing line; next, between each pair of adjacent combing lines σ_g and σ_h, the synchronous k-fellow-traveller property allows one to construct a *ladder* whose *rungs* are paths of length at most k between $\sigma_g(t)$ and $\sigma_h(t)$ for all t in

$$\{1, 2, \ldots, \max(\ell(\sigma_g), \ell(\sigma_h))\}.$$

The number of 2-cells in this *ladder* is $\max(\ell(\sigma_g), \ell(\sigma_h))$ and each has boundary word of length at most $2k+2$. It follows that if $\mathcal{R}$ is the set of null-homotopic words of length at most $2k+2$ then $\mathcal{P} := \langle \mathcal{A} \mid \mathcal{R} \rangle$ is a finite presentation for Γ and D_w is a van Kampen diagram for w with respect to $\mathcal{P}$. Define G to be the graph $D^{(1)}$.

We now seek to bound the gallery length of D_w and we notice that paths in the dual graph $G^\star$ that run along the *ladders* between adjacent combing lines have lengths $\preceq L$. However it may be the case that some *rungs* have zero length, and so it may not be possible to run along the ladders to reach the base vertex. To remedy this we *fatten the combing lines*: we fatten the presentation $\mathcal{P}$ to $\mathcal{P}' = \langle \mathcal{A}' \mid \mathcal{R}' \rangle$ as per Definition 4.4; then we cut along each combing line σ_g in D_w and glue in the diagram depicted in Fig. 4 which has boundary made up of two copies of σ_g and has all interior edges labelled by z. The result is a van Kampen diagram over $\mathcal{P}'$ that we call D'_w.

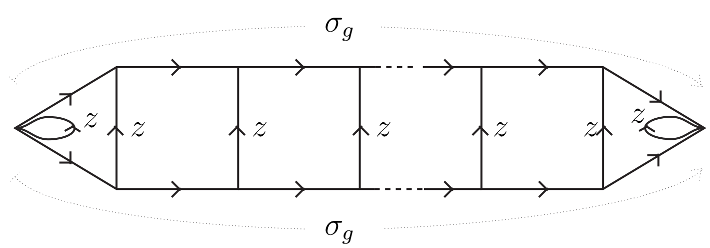

Figure 4: A *fattened* combing line

One checks that every vertex in the dual graph $G'^\star$ to D'_w (including $e^\star_\infty$) is within a bounded distance of a vertex dual to some 2-cell of a fattened combing line. So $\mathrm{GL}(D'_w)$ is bounded by $L(n)$ up to an additive constant. The area of the D'_w is at most $2nL(n)$. Now w does not involve the generator z. If w' is a word in $(\mathcal{A}' \cup \mathcal{A}'^{-1})^\star$ then let w be the word obtained by deleting all instances of $z^{\pm 1}$. Then construct a van Kampen diagram D'_w for w as above and obtain a van Kampen diagram $D'_{w'}$ for w' by attaching monogons with edges labelled by z. Then $\mathrm{GL}(D'_{w'}) \leq \mathrm{GL}(D'_w) + 2$. Deduce that $(nL(n), L(n))$ is an $(\mathrm{Area}, \mathrm{GL})$-pair for $\mathcal{P}'$ up to a common multiplicative constant.

(We remark that it follows from Scholium 4.7 that this same $(\mathrm{Area}, \mathrm{GL})$-pair applies to any finite fat presentation for $\mathcal{P}$, up to $\simeq$-equivalence.)

The proof in the case that σ satisfies the more general asynchronous k-fellow-traveller property is similar except that the *ladders* between the combing lines have *rungs* organised more haphazardly: for each $t \in \mathbb{N}$ there is a *rung* between $\sigma_g(\rho_g(t))$ and $\sigma_h(\rho_h(t))$ (for some reparametrizations ρ_g and ρ_h). The 2-cells are now *triangular* or *rectangular*; each has boundary made up of two *rungs* and either one or two 1-cells, and so has boundary length at most $2+2k$. We again take $\mathcal{R}$ to be the set of null-homotopic words of length at most $2+2k$. Then $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ is a finite presentation and the fattening of $\mathcal{P}$ admits the claimed $(\mathrm{Area}, \mathrm{GL})$-pair. ∎

7 Relating gallery length and filling length

In §4 we defined the filling length function $\mathrm{FL} : \mathbb{N} \to \mathbb{N}$ for a finite presentation $\mathcal{P}$ of a group Γ by

$$\mathrm{FL}(n) := \max \{ \mathrm{FL}(w) \mid \text{words } w \text{ with } \ell(w) \leq n \text{ and } w = 1 \text{ in } \Gamma \},$$

where $\mathrm{FL}(w)$ is the minimal value of $\mathrm{FL}(D_w)$ over all van Kampen diagrams for w. In §2 the diagram measurement $\mathrm{FL}(D_w)$ was defined as the minimal bound on the length of the boundary curve in the course of a shelling (a combinatorial null-homotopy) of the diagram.

An alternative definition for the filling length of a word w that equals 1 in $\mathcal{P}$ is that $\mathrm{FL}(w)$ is the minimal bound on the length of words one encounters when reducing the w down to the empty word by successively applying relators [10, Proposition 1]. This reveals $\mathrm{FL}(n)$ to be the space complexity of the naïve approach to solving the word problem for $\mathcal{P}$ by exhaustively applying relators – more details are in [9].

The main result of this section is that FL and DGL are qualitatively the same filling functions; that is, $\mathrm{FL} \simeq \mathrm{DGL}$. Thus filling length is a ubiquitous concept: not only is it both a space complexity measure and a differential geometric invariant controlling the length of curves in null-homotopies, but, in addition, it has a geometric/combinatorial interpretation in terms of the diameter of graphs embedded in 2-spheres and their duals.

Theorem 7.1. *Let* $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ *be a finite fat presentation of a group* Γ*. The filling functions* DGL *and* FL *for* $\mathcal{P}$ *satisfy* $\mathrm{GL} \leq \mathrm{DGL} \simeq \mathrm{FL}$.

Proof. That $\mathrm{FL} \preceq \mathrm{DGL}$ is immediate from the inequality of Theorem 3.5 relating the corresponding diagram measurements, and that $\mathrm{GL} \leq \mathrm{DGL}$ is an easy consequence of the definitions. (In fact, the hypothesis that $\mathcal{P}$ is *fat* is superfluous here.) We turn our attention to establishing the bound $\mathrm{DGL} \preceq \mathrm{FL}$.

Let D be a van Kampen diagram for an edge-circuit w in $\mathcal{P}$ of length n, that admits a shelling $\mathcal{S} = (C_0, C_1, \ldots, C_m)$ with $\mathrm{FL}(\mathcal{S}) = \mathrm{FL}(w) = \mathrm{FL}(n)$. Define $k := \mathrm{Area}(D)$ and $M := \max\{\ell(r) \mid r \in \mathcal{R}\}$. Only interior 1-cells of D can be involved in *1-cell expansion* move in $\mathcal{S}$. There are at most $M\,k$ such 1-cells. So the total number of *1-cell collapse moves* in $\mathcal{S}$ is at most the total number of 1-cells in the 1-dimensional portions of D plus twice the number in the 2-dimensional portions, that is at most $n + 2\,M\,k$. Adding these bounds to the number k of *2-cell collapse* moves we have

$$m \leq n + (2\,M + 1)\,k. \qquad (12)$$

We construct a van Kampen diagram $\overline{D}$ for w from D for which we will be able to use the length of the boundary circuits of the C_i to estimate $\mathrm{DGL}(\overline{D})$. Recalling Definition 4.4, there is one generator $z \in \mathcal{A}$ for which the words $z, z^2, zz^{-1}, z^3, z^2z^{-1}$ as well as the commutators $[z, a]$ for all $a \in \mathcal{A}$ are found amongst the defining relations. Our first step is to insert z-corridors into the diagram D along the (images of) boundaries of the C_i.

If $k = 0$ the bound $\mathrm{DGL}(n) \leq \mathrm{FL}(n)$ is immediate. Assume $k > 0$. For all $0 \leq i \leq m$ there is a natural combinatorial map $C_i \to D$, which is only prevented from being an embedding by the *1-cell expansion* moves in the shelling. Define $i_1 \leq i_2 \leq \ldots \leq i_k$ to be the k integers for which C_{i_j+1} is obtained from C_{i_j} by a *2-cell collapse* move. In the move $C_{i_j} \to C_{i_j+1}$ a 2-cell $e^2_{i_j}$ and an edge $e^1_{i_j} \subseteq \partial C_{i_j} \cap \partial e^2_{i_j}$ are removed from C_{i_j}.

Let ϕ_i be the edge-path in D that we obtain by mapping the anticlockwise boundary circuit ∂C_i into D using the map $C_i \to D$. Let w_i be the word one reads along ϕ_i. Then $\ell(w_i) = \ell(\partial C_i) \leq \mathrm{FL}(w) = \mathrm{FL}(n)$.

For a word u, let C_u denote the z-corridor for u; that is, the van Kampen diagram for the word $[z, u]$ that is a concatenation of 2-cells whose boundary words are commutators $[z, a^{\pm 1}]$ for letters a in u. The two portions of the boundary of C_u along which one reads u and u^{-1} are the *sides* of the corridor.

We produce a diagram $\hat{D}$ from D by gluing copies of C_{w_i} into D along the edge paths ϕ_i for $1 \leq i \leq m$ so as to introduce the word $z^{-m} z^m$ into the boundary circuit. So $\hat{D}$ is a van Kampen diagram for $z^{-m} z^m w$.

We obtain $\overline{D}$ from $\hat{D}$ by attaching two van Kampen diagrams A_{-m} and A_m, defined in the next paragraph, along the portions of the boundary circuit labelled by the words z^m and z^{-m} respectively. Figure 5 is a schematic picture of $\overline{D}$ in the case $m = 8$ with the two diagrams A_m and A_{-m} shown shaded.

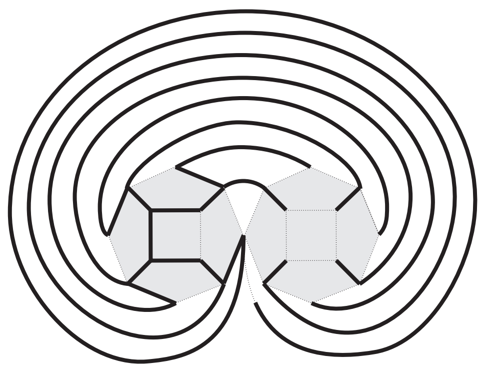

Figure 5: The diagram $\overline{D}$ for w and the maximal tree $\overline{T}$ in its 1-skeleton.

For $m = 1, 2, 3$ we define A_m to be the van Kampen diagram for z^m with one 2-cell. Define A_4 to be the diagram with two triangular 2-cells joined along one 1-cell to make a square. When $m \geq 5$, let q be the least integer greater than or equal to $\log_2 m$. The diagram A_m is built out of $q - 2$ concentric annular 2-complexes arranged around A_4. The construction is illustrated in Figure 6 in the case $m = 13$. These annuli are built out of five-sided and four-sided 2-cells, that are then triangulated (the dotted lines in Figure 6 are the edges inserted to triangulate the 2-cells). With the possible exception of the outermost, the annuli are built out of five-sided 2-cells in such a way that their outer-boundary is twice the length of their inner boundary. In the outermost annulus four-sided 2-cells are used in place of $2^q - m$ of the five-sided 2-cells in order to give the outermost boundary length m.

All the 2-cells in A_m are triangular and all 1-cells are labelled by z. The orientation of the interior edges is chosen arbitrarily and those in the boundary are oriented in such a way that the boundary word (reading anticlockwise) is z^m. As $\mathcal{P}$ is *fat* the defining relations include z^3 and $z^2 z^{-1}$ and so A_m is a van Kampen diagram with respect to $\mathcal{P}$. The diagram A_{-m} for z^{-m} is obtained from A_m by reversing the orientation of the boundary edges. (It is significant, although not to this proof, that all the vertices of A_m have uniformly bounded valence – see Remark 7.3.)

Refer to the boundary components of the concentric annuli as *rings* in $A_m^{(1)}$ and refer to the edges of the four or five-sided 2-cells (before triangulation) that are not in the *rings* as *radial* edges.

Both the diameter and gallery length of the diagrams $A_{\pm m}$ are $\preceq q \leq 1 + \log_2 m$. More particularly, define U to be a tree in the 1-skeleton of A_{-m} consisting of all the *radial* edges, as well as all but one edge of the innermost *ring* and alternate edges on the remaining *rings* in the interior of A_{-m} (see Figure 6, where U is shown with heavy lines in the left-hand diagram). Also define V to be the forest of trees in the 1-skeleton of A_m made up of all the *radial* edges (see the righthand diagram of Figure 6). Then let $U^\star$ and $V^\star$ be the *internal complementary forests* of U and V in $A_{\mp m}$. (The *internal complementary forest* of U in A_{-m} has a vertex dual to every 2-cell in A_{-m}, and an edge dual to every edge in the interior of A_{-m} that is not in U.) Then $U^\star$ is a forest and $V^\star$ is a tree, and one checks that the diameter of any connected component of U, $U^\star$, V and $V^\star$ is at most

$$6\,q \;\leq\; 6 + 6\,\log_2 m \;\leq\; K + K\,n + K\,\mathrm{FL}(n), \tag{13}$$

for a constant $K > 0$ that depends only on $\mathcal{P}$ – the second inequality is obtained from (12) and the following consequence of the *"space-time bound"* of Proposition 5.6:

$$k \;\leq\; (2\,|\mathcal{A}| + 1)^{\mathrm{FL}(D)} \;=\; (2\,|\mathcal{A}| + 1)^{\mathrm{FL}(n)},$$

which holds because we can assume that D has minimal area amongst all van Kampen diagrams for w with $\mathrm{FL}(w) = \mathrm{FL}(n)$.

We now come to the bound on $\mathrm{DGL}(\overline{D})$. We specify a maximal tree $\overline{T}$ in the graph $\overline{G} = \overline{D}^{(1)}$ and examine its diameter and that of the complementary tree $\overline{T}^\star$ in $\overline{G}^\star$. A maximal tree $\hat{T}$ in $\hat{D}^{(1)}$ is obtained by taking the union of all the sides of the corridors C_{w_i} and the z^m portion of the boundary and then removing (the images in $\hat{D}$ of) the edges $e^1_{i_j}$ between the 2-cells $e^2_{i_j}$ and the corridors $C_{w_{i_j}}$.

We then obtain $\overline{T}$ from $\hat{T}$ by removing the z^m portion of the boundary word of $\hat{D}$, by including the forests U and V of the attached diagrams A_{-m}

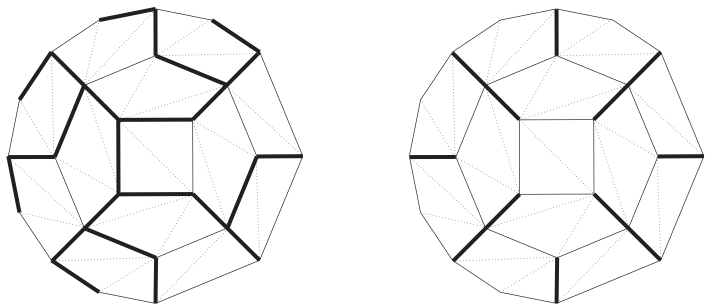

Figure 6: The diagrams (A_{-13}, U) and (A_{13}, V).

and A_m, and by removing the first edge of $\partial\overline{D}$. One checks that $\overline{T}$ is a maximal tree in $\overline{D}^{(1)}$ (or, equivalently, that the complementary tree $\overline{T}^{\star}$ is a maximal tree) – see Figure 5 where $\overline{T}$ is illustrated with heavy lines.

We claim that

$$\mathrm{Diam}(\overline{T}),\ \mathrm{Diam}(\overline{T}^{\star}) \quad \leq \quad 3(\mathrm{FL}(n) + M + K + K\,n + K\,\mathrm{FL}(n)\,)\,,$$

We first explain the contributions to the bound on $\mathrm{Diam}(\overline{T})$. One can reach the side of one of the corridors $C_{w_{i_j}}$ from any vertex of V within a distance at most $K + K\,n + K\,\mathrm{FL}(n)$ by (13). One can then follow the side of a corridor $C_{w_{i_j}}$ (length at most $\mathrm{FL}(n)$), but possibly making a diversion around a 2-cell $e^2_{i_j}$ (length less than $\ell(\partial e^2_{i_j}) \leq M$), and reach the subtree U. The diameter of U is at most $K + K\,n + K\,\mathrm{FL}(n)$ by (13).

For the bound on $\mathrm{Diam}(\overline{T}^{\star})$ notice that from a vertex dual to any 2-cell not in either of the subdiagrams $A_{\pm m}$ one can reach the interior of a 2-cell in a corridor within a distance 1. Further, from a vertex in A_{-m} one can reach the interior of a corridor by following a component of $U^{\star}$. Then one can follow a path along the interior of a corridor to the tree $V^{\star}$.

Summing the inequalities for $\mathrm{Diam}(\overline{T})$ and $\mathrm{Diam}(\overline{T}^{\star})$ we have a bound on $\mathrm{DGL}(\overline{D})$ from which we can conclude that $\mathrm{DGL} \preceq \mathrm{FL}$. ■

Conjecture 4.3 of [11] claims a close relationship between the diagram measurements GL and DGL. If this conjecture holds then an implication for the filling functions of a given presentation is that $\mathrm{GL} \simeq \mathrm{DGL}$. This would lead to the following attractively simple interpretation of the filling length function as a measure of the diameter of dual graphs to van Kampen diagrams.

Corollary 7.2. *Let $\mathcal{P}$ be a finite fat presentation. Assuming Conjecture 4.3 of* [11] *holds, the filling functions* $\mathrm{FL}, \mathrm{GL} : \mathbb{N} \to \mathbb{N}$ *for $\mathcal{P}$ satisfy* $\mathrm{FL} \simeq \mathrm{GL}$.

Thus assuming Conjecture 4.3 of [11], Corollary 6.6 in [11] would also reproduce the result of [9] that the filling length function of a finite presentation of a nilpotent group admits a linear upper bound.

Remark 7.3. The maximum valence of vertices in the diagram $\overline{D}$ constructed in the proof of Theorem 7.1 is at most 11. This observation is not used in the proof above, but is important in the proof of an inequality in [11, Theorem 9.3].

One checks that vertices in the interior of the $A_{\pm m}$ have valence at most 7, and vertices on $\partial A_{\pm m}$ have valence at most 4. Also vertices in the interior of $\hat{D}$ and on $\partial \hat{D}$ have valence at most 6 and 4 respectively. When A_m and A_{-m} are attached to $\hat{D}$ to form $\overline{D}$ the maximum valence of the identified vertices is at most 11 – the most complicated vertex to check is the common vertex of A_m and A_{-m}, where five vertices are identified: one from each of A_m and A_{-m} (each of valence ≤ 4) and the three vertices of $\partial \hat{D}$ at the ends of and in the middle of the $z^{-m} z^m$ portion of the boundary word (valences at most 2, 2 and 3 respectively).

A The length of expressions for integers as sums $\pm p^i$ for fixed prime p

In this appendix we prove a result about representing integers by sums of terms $\pm p^i$, for a prime p. It is well known that the minimal number of terms $\rho(n)$ in an expression for a positive integer n as a sum of $\{p^i \mid i \in \mathbb{N}\}$ is $\lambda_0 + \lambda_1 + \ldots + \lambda_m$ where $n = \lambda_0 + \lambda_1 p + \cdots + \lambda_m p^m$ is the base p representation of n. Therefore $\{\rho(n) \mid n \in \mathbb{N}\}$ is unbounded. However we are concerned with the minimal number $\|n\|$ of terms in such sums when one is also allowed terms $\{-p^i \mid i \in \mathbb{N}\}$.

Word metrics provide a convenient language in which to formalise the definitions of $\rho(n)$ and $\|n\|$. Fix a prime p. Take the infinite set $\mathcal{A} := \{p^i \mid i \geq 0\}$ as a generating set for the additive monoid $\mathbb{N}$, and $\mathcal{A}^{\pm} := \{\pm p^i \mid i \geq 0\}$ as a symmetric generating set for $\mathbb{Z}$. These define two word metrics $d_{\mathcal{A}}$ and $d_{\mathcal{A}^{\pm}}$; we define $\rho(n) := d_{\mathcal{A}}(0, n)$ and $\|n\| := d_{\mathcal{A}^{\pm}}(0, n)$. We remark that the metrics are different on $\mathbb{N}$ in general: if $n = p^i - 1$ then $\rho(n) = i$ but $\|n\| = 2$.

Theorem A.1. *If $\{a_n\,;\ n \geq 1\}$ is a sequence of rational integers which converges in the p-adic integers $\hat{\mathbb{Z}}_p$ to an element which is not a rational integer (i.e. the limit is in $\hat{\mathbb{Z}}_p \smallsetminus \mathbb{Z}$), then the sequence $\{\|a_n\|\,;\ n \geq 1\}$ is unbounded. Hence $\|\ \|$ is unbounded on $\mathbb{Z}$ (and so also on $\mathbb{N}$).*

Proof. Assume to the contrary that $\|\ \|$ is bounded on the sequence $\{a_n\}$ and let $\|a_n\| \leq M$ for all n. We may write $a_n = b_n - c_n$, where

$$\begin{array}{rcl} b_n & = & p^{e_{n1}} + p^{e_{n2}} + \cdots + p^{e_{ns_n}} \\ c_n & = & p^{f_{n1}} + p^{f_{n2}} + \cdots + p^{f_{nt_n}} \end{array}$$

with

$$\begin{array}{c} e_{n1} \leq e_{n2} \leq \cdots \leq e_{ns_n} \\ f_{n1} \leq f_{n2} \leq \cdots \leq f_{nt_n}, \end{array}$$

so that $\|a_n\| = s_n + t_n \leq M$.

We express the following argument informally to avoid the notation getting out of control. Depending on whether or not the sequence $\{e_{n1}\}_{n\in\mathbb{N}}$ is bounded, we can extract a subsequence that is either constant or tends to infinity. In the second case the b_n tend to the rational integer 0 on the subsequence, whereas in the first case the first summand $p^{e_{n1}}$ is constant. Now restrict to the subsequence. And, provided that infinitely many e_{n2} are defined on the subsequence, argue that there is a further subsequence for which the e_{n2} either are constant or tend to infinity. In the second case the b_n tend to the rational integer $p^{e_{n1}}$ on this new subsequence. Continue similarly extracting subsequences examining the e_{n3}, then the e_{n4} and so on, in all at most M times. The result is a subsequence on which the b_n tend to a rational integer. Next look at the f_{n1}, then f_{n2}, and so on, and extract further subsequences at most M times. This produces a subsequence on which the c_n also tend to a rational integer.

So, after extracting subsequences a total of at most $2M$ times, we have a subsequence of the a_n which tends to a rational integer in the limit. This contradicts our hypothesis.

To deduce that $\|\ \|$ is unbounded it suffices to give a sequence of rational integers which is convergent in $\hat{\mathbb{Z}}_p$ but whose limit is not a rational integer. Well, $a_n := 1 + p^2 + \cdots + p^{2n}$ tends to $-1/(p^2-1)$, which is not a rational integer for any p. ∎

References

[1] M. R. Bridson. On the geometry of normal forms in discrete groups. *Proc. London Math. Soc.*, 67(3):596–616, 1993.

[2] M. R. Bridson. The geometry of the word problem. In M. R. Bridson and S. Salamon, editors, *Invitations to Geometry and Topology*, pages 33–94. O.U.P., 2002.

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

[4] R. Diestel. *Graph Theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag, 2nd edition, 2000.

[5] D. B. A. Epstein, J. W. Cannon, D. F. Holt, S. V. F. Levy, M. S. Paterson, and W. P. Thurston. *Word Processing in Groups*. Jones and Bartlett, 1992.

[6] S. Frankel and M. Katz. The Morse landscape of a Riemannian disc. *Ann. Inst. Fourier, Grenoble*, 43(2):503–507, 1993.

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

[8] S. M. Gersten. Isoperimetric and isodiametric functions. In G. Niblo and M. Roller, editors, *Geometric group theory I*, number 182 in LMS lecture notes. Camb. Univ. Press, 1993.

[9] S. M. Gersten, D. F. Holt, and T. R. Riley. Isoperimetric functions for nilpotent groups. *GAFA*, 13:795–814, 2003.

[10] S. M. Gersten and T. R. Riley. Filling length in finitely presentable groups. *Geom. Dedicata*, 92:41–58, 2002.

[11] S. M. Gersten and T. R. Riley. Some duality conjectures for finite graphs and their group theoretic consequences. To appear in Proc. Edin. Math. Soc., `http://www.math.utah.edu/~gersten/grouptheory.htm/`, 2003.

[12] S. M. Gersten and H. Short. Some isoperimetric inequalities for free extensions. *Geom. Dedicata*, 92:63–72, 2002.

[13] 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.

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

[15] A. Yu. Ol′shanskii. *Geometry of defining relations in groups.* Kluwer Academic Publ., 1991. [Russian original, Nauka Publ., 1989].

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

Steve M. Gersten
Mathematics Department, 155S. 1400E., Rm. 233, Univ. of Utah, Salt Lake City, UT 84112, USA
gersten@math.utah.edu, http://www.math.utah.edu/∼sg/

Tim R. Riley
Mathematics Department, 10 Hillhouse Avenue, P.O. Box 208283, New Haven, CT 06520-8283, USA
tim.riley@yale.edu, http://www.math.yale.edu/users/riley/