FREE AND FRAGMENTING FILLING LENGTH

M. R. BRIDSON AND T. R. RILEY

A . The filling length of an edge-circuit η in the Cayley 2-complex of a finite presentation of a group is the minimal integer length L such that there is a combinatorial null-homotopy of η down to a base point through loops of length at most L. We introduce similar notions in which the null-homotopy is not required to fix a base point, and in which the contracting loop is allowed to bifurcate. We exhibit a group in which the resulting filling invariants exhibit dramatically different behaviour to the standard notion of filling length. We also define the corresponding filling invariants for Riemannian manifolds and translate our results to this setting.

1. I

Consider a vertical cylinder $C \subseteq \mathbb{R}^3$ of height h and diameter $d \ll h$. Let S be the surface formed by the curved portion of C and the disc capping off its top. Topologically, S is a closed 2-disc. The loop ∂S can be homotoped in S to a constant loop through loops of length at most πd by lifting it up the cylinder and then contracting it across the top of C. However, if we insist on keeping a basepoint on ∂S fixed in the course of the null-homotopy then we will encounter far longer loops, some of length at least $2h$.

In this article we will bring to light similar contrasts between basepoint-fixed and basepoint-free null-homotopies for loops in the Cayley 2-complex $Cay^2(\mathcal{P})$ of a finite presentation $\mathcal{P}$ of a group Γ. Words w that represent 1 in Γ (*null-homotopic* words) correspond to edge-circuits η_w in $Cay^2(\mathcal{P})$. The filling length $\mathrm{FL}(w)$ of w was defined by Gromov [13] and in a combinatorial context is the minimal length L such that there is a base point fixing, *combinatorial null-homotopy* of η_w through loops of length at most L. (A closely related notion called LNCH was considered by Gersten in [9].) We define $\mathrm{FFL}(w)$, the *free filling length* of w, likewise but without holding a base point fixed, and $\mathrm{FFFL}(w)$, the *fragmenting free filling length* of w, by also allowing the contracting loops to bifurcate. Detailed definitions are given in Section 2.

Date: December 1, 2005.

2000 *Mathematics Subject Classification.* 2000 Mathematics Subject Classification: 20F65, 20F10, 53C23.

Key words and phrases. filling length, null-homotopy, van Kampen diagram.

Our first goal is to construct a finite presentation in which $\mathrm{FL}(w)$ and $\mathrm{FFL}(w)$ differ dramatically for an infinite sequence of null-homotopic words of increasing length. [Our notational conventions are $[a,b] = a^{-1}b^{-1}ab$, $a^b = b^{-1}ab$ and $a^{-b} = b^{-1}a^{-1}b$. For $f, g : \mathbb{N} \to \mathbb{N}$, we write $f \preceq g$ when there exists $C > 0$ such that for all n we have $f(n) \leq Cg(Cn+C) + Cn + C$, which gives an equivalence relation expressing qualitative agreement of the growths of f and g: write $f \simeq g$ if and only if $f \preceq g$ and $g \preceq f$.]

Theorem A. *Let Γ be the group with aspherical presentation*

$$\mathcal{Q} \ := \ \langle\, a, b, t, T, \tau \ \mid \ a^b a^{-2}, [t,a], [\tau, at], [T,t], [\tau, T] \,\rangle.$$

For $n \in \mathbb{N}$ define $w_n := [T, a^{-b^n} \tau a^{b^n}]$. Then w_n is null-homotopic in $\mathcal{Q}$, has length $\ell(w_n) = 8n+8$, and $\mathrm{FFL}(w_n) \simeq n$, but $\mathrm{FL}(w_n) \simeq 2^n$.

We assume the reader is familiar with *van Kampen diagrams* ([4] is a recent survey); they can be thought of as *combinatorial homotopy discs* for loops in the Cayley 2-complex of a presentation. We show that there are van Kampen diagrams $\hat{\Delta}_n$ for w_n that, owing to geometry like that in our cylinder example, have $\mathrm{FFL}(\hat{\Delta}_n) \preceq n$ and $\mathrm{FL}(\hat{\Delta}_n) \simeq 2^n$. Indeed, we will show that every van Kampen diagram Δ for w_n has intrinsic diameter $\succeq 2^n$. (The intrinsic diameter of Δ is the maximal distance between two vertices as measured in combinatorial metric on $\Delta^{(1)}$) ; it will then follow that $\mathrm{FL}(w_n) \succeq 2^n$.

The area $\mathrm{Area}(w)$ and intrinsic diameter $\mathrm{IDiam}(w)$ of w are the minimal A and D, respectively, such that there is a van Kampen diagram for w with A 2-cells or with intrinsic diameter D. For $\mathrm{M} = \mathrm{Area}, \mathrm{IDiam}, \mathrm{FL}, \mathrm{FFL}$ or FFFL define *filling functions* $\mathrm{M} : \mathbb{N} \to \mathbb{N}$ for finitely presented groups:

$$\mathrm{M}(n) \ := \ \max\{\, \mathrm{M}(w) \mid w \text{ null-homotopic and } \ell(w) \leq n \,\}.$$

(The argument of M determines its meaning; the potential for ambiguity is tolerated as it spares us from a terminology over-load.) In the case $\mathrm{M} = \mathrm{Area}$, the function is the *Dehn function*.

In spite of Theorem A, the filling functions FL and FFL for $\mathcal{Q}$ are $\simeq$-equivalent: we will see in Section 3 that any van Kampen diagram for $w'_n := [T,\ a^{-b^n}\tau a^{b^n}\tau a^{b^n}\tau^{-1}a^{-b^n}]$ has two *peaks* and the savings that can be made by escaping the base point are no longer significant. On the other hand, if we allow our loops to bifurcate then they can pass over peaks independently. So FFFL exhibits markedly different behaviour.

Theorem B. *The filling functions* $\mathrm{IDiam}, \mathrm{FL}, \mathrm{FFL}, \mathrm{FFFL} : \mathbb{N} \to \mathbb{N}$ *for $\mathcal{Q}$ satisfy*

$$\begin{aligned} \mathrm{IDiam}(n) \ \simeq \ \mathrm{FL}(n) \ \simeq \ \mathrm{FFL}(n) \ \simeq \ \mathrm{Area}(n) \ &\simeq \ 2^n \\ \mathrm{FFFL}(n) \ &\simeq \ n. \end{aligned}$$

We remark that Γ has the properties $\text{IDiam}(n) \simeq \text{Area}(n)$ and $\text{FL}(n) \simeq \text{Area}(n)$, which are unusual in non-hyperbolic groups. In contrast, if $\text{Area}(n) \simeq n^{\alpha}$ for some $\alpha \geq 2$ then $\text{IDiam}(n) \preceq n^{\alpha-1}$ – see [11].

Theorems A and B are proved in Section 3, modulo a number of auxiliary propositions postponed to Section 4. The remainder of this article is dedicated to establishing the credentials of FL, FFL and FFFL for inclusion in the pantheon of filling invariants, to relating them to other filling functions, and to interpret them in terms of algorithmic complexity.

As explained in [11], FL can be thought of as a space-complexity measure in that $\text{FL}(w)$ is the minimal L such that w can be converted to the empty word through a sequence of words of length at most L, each obtained from the previous by free reduction, free expansion, or applying a relator. We will show in Section 2 that FFL and FFFL can be interpreted in a similar way: if we also allow conjugation we get $\text{FFL}(w)$ and if we additionally include the move that replaces a word $w = uv$ by a pair of words u, v, we get $\text{FFFL}(w)$. This point of view is useful for calculations and allows us to prove that, as for FL, given a finite presentation, $\text{Area}(n)$ is at most an exponential of FFFL. This and other relationships are spelt out in the following theorem, which shows, in particular, that Q of Theorem A provides an example of $\text{Area}(n)$ outgrowing $\text{FFFL}(n)$ as extremely as is possible.

Theorem C. *Let $\mathcal{P}$ be a finite presentation. There is a constant C, depending only on $\mathcal{P}$, such that for all $n \in \mathbb{N}$ the Dehn, filling length, free filling length, and fragmenting free filling length functions* $\text{Area}, \text{FFL}, \text{FFFL} : \mathbb{N} \to \mathbb{N}$ *of $\mathcal{P}$ satisfy*

$$\begin{aligned} \text{Area}(n) &\leq C^{\text{FFFL}(n)} \\ \text{FFFL}(n) &\leq \text{FFL}(n) \leq \text{FL}(n) \leq C\,\text{Area}(n) + n. \end{aligned}$$

In Section 5 we prove that $\text{FL}(n)$, $\text{FFL}(n)$ and $\text{FFFL}(n)$ are all quasi-isometry invariants (and, in particular, are all group invariants – c.f. [12, Theorem 8.1]) of finitely presented groups, up to $\simeq$-equivalence.

Theorem D. *If $\mathcal{P}$ and $\mathcal{P}'$ are finite presentations of quasi-isometric groups then*

$$\text{FL}_{\mathcal{P}} \simeq \text{FL}_{\mathcal{P}'}, \quad \text{FFL}_{\mathcal{P}} \simeq \text{FFL}_{\mathcal{P}'}, \textit{ and } \quad \text{FFFL}_{\mathcal{P}} \simeq \text{FFFL}_{\mathcal{P}'}.$$

In Section 6 we define analogous filling invariants $\text{FL}_X(n)$, $\text{FFL}_X(n)$ and $\text{FFFL}_X(n)$ describing the geometry of null-homotopies for rectifiable loops in arbitrary metric spaces X, and we prove:

Theorem E. *Suppose a group Γ with finite presentation $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ acts properly and cocompactly by isometries on a simply connected geodesic metric space X for which there exist $\mu, L > 0$ such that every loop of length less than μ admits a based null-homotopy of filling length less than L. Then* $\text{FL}_{\mathcal{P}} \simeq \text{FL}_X$, $\text{FFL}_{\mathcal{P}} \simeq \text{FFL}_X$ *and* $\text{FFFL}_{\mathcal{P}} \simeq \text{FFFL}_X$.

As the universal cover of any closed connected Riemmian manifold satisfies these conditions and (as is well known) every finitely presentable group is the fundamental groups of such a manifold, we can use Theorem E and its analogues for Area and IDiam (proved in [4, 6]) to obtain the following results from Theorem A and B.

Corollary F. *There exists a closed connected smooth Riemannian manifold M such that*

$$\begin{array}{rcl} \mathrm{IDiam}_{\widetilde{M}}(n) \ \simeq \ \mathrm{FL}_{\widetilde{M}}(n) \ \simeq \ \mathrm{FFL}_{\widetilde{M}}(n) \ \simeq \ \mathrm{Area}_{\widetilde{M}}(n) & \simeq & 2^n \\ \mathrm{FFFL}_{\widetilde{M}}(n) & \simeq & n. \end{array}$$

Moreover, there is an infinte sequence of loops c_n in $\widetilde{M}$ such that $\ell(c_n) \to \infty$, $\mathrm{FFL}_{\widetilde{M}}(c_n) \simeq n$, *and* $\mathrm{FL}_{\widetilde{M}}(c_n) \simeq 2^n$.

(Strictly speaking, the final part is not a direct corollary of the prior results, but it follows from the methods of Section 6: the w_n of Theorem A can be used to construct *word-like* loops c_n with $\ell(c_n) \simeq \ell(w_n)$; van Kampen diagrams witnessing to the upper bounds $\mathrm{FFL}(w_n) \leq n$ and $\mathrm{FL}(w_n) \leq 2^n$ can be carried to fillings of c_n showing $\mathrm{FFL}_{\widetilde{M}}(c_n) \leq n$ and $\mathrm{FL}_{\widetilde{M}}(c_n) \leq 2^n$; and every null-homotopy of c_n in $\widetilde{M}$ has filling length $\geq 2^n$ because the filling-disc corresponding to a null-homotopy induces a van Kampen diagram filling for w_n with $\simeq$ filling length.)

Natural questions that remain open include the following.

Open problem 1.1. Does there exist a finite presentation for which $\mathrm{FL}(n) \not\simeq \mathrm{FFL}(n)$?

Open problem 1.2. Does there exist a finite presentation for which $\mathrm{IDiam}(n) \not\simeq \mathrm{FL}(n)$?

In reference to the first of these problems we note that for a finitely presentation $\langle \mathcal{A} \mid \mathcal{R} \rangle$ of a group G, the presentation $\langle \mathcal{A} \cup \{t\} \mid \mathcal{R} \rangle$ for $G * \mathbb{Z}$ satisfies $\mathrm{FL}(n) \simeq \mathrm{FFL}(n)$ for the following reason. If w_n is a word over $\mathcal{P}$ with $\mathrm{FL}(w_n) = \mathrm{FL}(n)$ then $\mathrm{FFL}([w_n, t]) \geq 2 + \ell(w_n) + \mathrm{FL}(w_n)$ since any diagram for $[w_n, t]$ is a diagram for w_n joined to a diagram for ${w_n}^{-1}$ by a t-edge, and the most efficient way to shell such a diagram (from the point-of-view of FFL) is to shell the w_n diagram, then collapse the t-edge, and then shell the ${w_n}^{-1}$ diagram.

The second problem was posed by Gromov [13, §5C]. A negative answer would imply that the *double exponential upper bound* [7, 8] on the Dehn (i.e. area) function in terms of $\mathrm{IDiam}(n)$ could be improved to a single exponential using the single exponential upper bound [11, 13] for $\mathrm{Area}(n)$ in terms of $\mathrm{FL}(n)$.

Acknowledgement. The second author is grateful for support from NSF grant 0404767 and for the hospitality of the Centre de Recerca Matemàtica in Barcelona during the writing of this article.

2. T

Let Δ be a *diagram*; that is, a finite, planar, contractible, combinatorial 2-complex; i.e. a van Kampen diagram bereft of any group theoretic decorations. Before discussing filling length, we recall and develop a combinatorial notion of a null-homotopy of Δ from [11] called a *shelling*.

Definition 2.1. A *shelling* $\mathcal{S} = (\Delta_i)$ of Δ is a sequence of diagrams

$$\Delta = \Delta_0, \Delta_1, \ldots, \Delta_m,$$

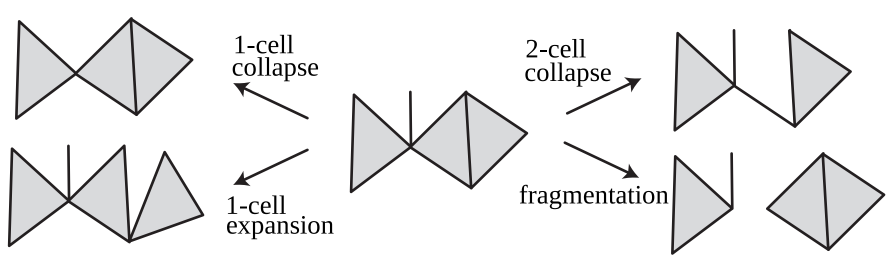
in which each Δ_{i+1} is obtained from Δ_i by one of the *shelling moves* defined below and depicted in Figure 1.

- *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^1 is attached to the rest of Δ_i only by one of its end vertices $\neq e^0$. (We call such an e^1 a *spike*.)
- *1-cell expansion.* Cut along some 1-cell e^1 in Δ_i that has a vertex e^0 in $\partial\Delta_i$, in such a way that e^0 and e^1 are doubled.
- *2-cell collapse.* Remove a pair (e^2, e^1) where e^2 is a 2-cell which has some edge $e^1 \in (\partial e^2 \cap \partial\Delta_i)$. The effect on the boundary circuit is to replace e^1 with $\partial e^2 \smallsetminus e^1$.

We say that the shelling $\mathcal{S}$ is *full* when Δ_m is a single vertex. A *full shelling to a base vertex* $\star = \Delta_m$ on $\partial\Delta$ is a full shelling in which $\star$ is preserved throughout the sequence (Δ_i). In particular, in every *1-cell collapse* $e^0 \neq \star$, and in every *1-cell expansion* on Δ_i where $e_0 = \star$ a choice is made as to which of the two copies of e_0 is to be $\star$ in Δ_{i+1}.

We define a *full fragmenting shelling* $\mathcal{S}$ of Δ by adapting the definition of a full shelling to allow each Δ_i to be a disjoint union of finitely many diagrams, insisting that Δ_m be a set of vertices; we also allow one extra type of move:

- *Fragmentation.* Δ_{i+1} is the disjoint union of Δ_i' and Δ_i'', where $\Delta_i = \Delta_i' \cup \Delta_i''$ and $\Delta_i' \cap \Delta_i''$ is a single vertex.



F 1. Shelling moves.

For a shelling $\mathcal{S}$, define $\ell(\mathcal{S}) := \max_i \ell(\partial\Delta_i)$, where $\ell(\partial\Delta_i)$ denotes the sum of the lengths of the boundary circuits of the components of Δ_i. Then the *filling length* $\mathrm{FL}(\Delta, \star)$, the *free filling length* $\mathrm{FFL}(\Delta)$, and the *fragmented free filling length* $\mathrm{FFFL}(\Delta)$ are the minimum of $\ell(\mathcal{S})$ as $\mathcal{S}$ ranges over all full shellings to $\star$, full shellings, and all full fragmenting free shellings of Δ, respectively. This notation emphasizes the fact that $\mathrm{FL}(\Delta, \star)$ is defined with respect to a base vertex $\star \in \partial\Delta$ but $\mathrm{FFL}(\Delta)$ and $\mathrm{FFFL}(\Delta)$ are not.

Let $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ be a finite presentation of a group Γ. Define the filling length $\mathrm{FL}(w)$ of a null-homotopic word w in a presentation $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ by

$$\mathrm{FL}(w) := \min \{\mathrm{FL}(\Delta, \star) \mid \Delta \text{ a van Kampen diagram for } w\},$$

and $\mathrm{FFL}(w)$ and $\mathrm{FFFL}(w)$ likewise.

A sequence $\mathcal{S} = w_0, \ldots, w_m$ of null-homotopic words is a *null-sequence* if each w_{i+1} is obtained from w_i by one of the following moves:

- *Free reduction.* Remove a subword aa^{-1} or $a^{-1}a$ from w_i, where $a \in \mathcal{A}$.
- *Free expansion.* This is the inverse of a free reduction.
- *Application of a relator.* Replace a subword u of w_i by a word v such that a cyclic conjugate of uv^{-1} is in $\mathcal{R}^{\pm 1}$.

We define two more moves:

- *Cyclic conjugation.* Replace w_i by a cyclic permutation.
- *Fragmentation.* Replace a word $w = uv$ by a pair of words u, v. (In effect, insert a letter into w that represents a blank space.)

To employ the fragmentation move we must generalise our definition of a *null-sequence* so that each w_i is a finite sequence of words, and when we perform any of the operations listed above we execute it on one of the words in w_i.

The following reassuring lemma is straight forward to prove. The “if” part is well known (indeed, freely reducing, freely expanding, and apply relators suffices). The “only if” part can be proved by an easy induction on the number of fragmentation moves used.

Lemma 2.2. *A word w over $\mathcal{P}$ represents the identity if and only if it can be reduced to a sequence of empty words by free reductions, free expansions, applying relators, cyclically conjugating, and fragmenting.*

For a null-sequence $\mathcal{S}$, define $\ell(\mathcal{S}) := \max_i \ell(w_i)$ where, if fragmentation moves are employed, $\ell(w_i)$ is the sum of the lengths of words in the sequence w_i. Proposition 1 in [11] says that for all null-homotopic words w, we have $\mathrm{FL}(w) = \min_{\mathcal{S}} \ell(\mathcal{S})$, quantifying over null-sequences $\mathcal{S}$ for w that employ *free reductions, free expansions* and *applications of relators*. We add the following.

Proposition 2.3. *Quantifying over all null-sequences $\mathcal{S}$ for a null-homotopic word w, where* free reduction, free expansion, applications of relators, *and* cyclic conjugation *are allowed, we have* $\mathrm{FFL}(w) = \min_{\mathcal{S}} \ell(S)$*. If, additionally, we allow* fragmentations *we get* $\mathrm{FFFL}(w) = \min_{\mathcal{S}} \ell(S)$.

Proof. The proof in [11] that $\min_{\mathcal{S}} \ell(\mathcal{S}) \leq \mathrm{FL}(w)$ is straightforward because the words around the boundary of the van Kampen diagrams in the course of a full shelling form a null-sequence. Each word in the sequence of boundary words in the course of a *free* shelling of a van Kampen diagram is only defined up to cyclic conjugation, as a base-vertex is not kept fixed during the shelling. It is then easy to see that $\min_{\mathcal{S}} \ell(S) \leq \mathrm{FFL}(w)$, quantifying over all all null-sequences $\mathcal{S}$ for w that use *free reduction*, *free expansion*, *applications of relators*, and *cyclic conjugation*. Introducing *fragmentation* moves into the shelling, produces *fragmentations* in the corresponding null-sequence. And we see $\min_{\mathcal{S}} \ell(S) \leq \mathrm{FFFL}(w)$.

The reverse bounds require more care. Given a null-sequence $\mathcal{S}$ for w involving *applications of relations* and *free expansions* and *reductions*, we seek to construct a van Kampen diagram with a shelling during which the lengths of the boundary circuit remain at most $\ell(\mathcal{S})$. This can be done by starting with an edge-loop labelled by w, and filling it in by attaching a 2-cell on every *application of a relator*, by attaching a 1-cell on every *free expansion*, by folding together two adjacent 1-cells on every *free reduction*. However it is possible that the resulting complex will not be planar: 2-spheres or other cycles may be pinched off (for example, when an inverse pair aa^{-1} is inserted and then removed). Removing these cycles gives a van Kampen diagram Δ with $\mathrm{FL}(\Delta) \leq \ell(w)$. This is explained carefully in [11]. If $\mathcal{S}$ also uses *cyclic conjugation* moves then no extra complications are added to the construction of Δ. If $\mathcal{S}$ also uses *fragmentations* then the corresponding move in the course of the construction of Δ is to identify two vertices so that an inner boundary circuit is changed from a topological circle to a figure-eight. □

Remark 2.4. (*Filling length and space complexity*.) Envisaging a null-sequence to be the course of a calculation on a Turing tape, we see that $\mathrm{FL}(w)$ is the non-deterministic space complexity of the following approach to solving the word problem for $\mathcal{P}$: write w on the tape and then exhaustively apply relators and perform free reductions and free expansions. A sequence of moves that converts w into the empty word amounts to a proof that w represents 1 in $\mathcal{P}$ and $\mathrm{FL}(w)$ is the minimal upper bound on the number of places on the tape that have to be used in the calculation (see [10] for more details). If we allow cyclic conjugation then the non-deterministic space complexity is $\mathrm{FFL}(w)$. If we also including fragmentation then the non-deterministic space complexity is $\mathrm{FFFL}(w)$ plus the maximum number of blank spaces separating the words; that is, between $\mathrm{FFFL}(w)$ and $2\,\mathrm{FFFL}(w)$.

The following inequalities are easy

(1) $\qquad \text{FFFL}(w) \leq \text{FFL}(w) \leq \text{FL}(w) \leq K\,\text{Area}(w) + \ell(w),$

(2) $\qquad \text{IDiam}(w) \leq \text{FL}(w),$

where K is a constant depending only on $\mathcal{P}$. (See [11] or [13].) Similar inequalities, for example $\text{IDiam}(n) \leq \text{FL}(n)$ for all n, relate the corresponding filling functions (defined in Section 1).

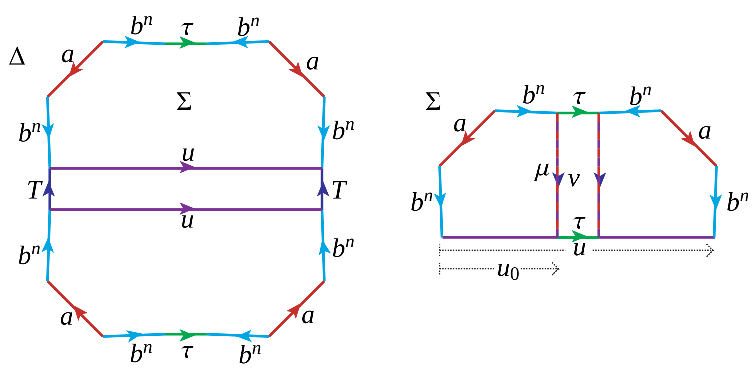
It is well known [11, 13] that there is a constant C, that depends only on $\mathcal{P}$, such that the Dehn function $\text{Area} : \mathbb{N} \to \mathbb{N}$ of $\mathcal{P}$ satisfies

$$\text{Area}(n) \leq C^{\text{FL}(n)}$$

for all n. This is essentially the bound Time-Space bound of algorithmic complexity: the number of different words of length $\text{FL}(n)$ on an alphabet of size C is $C^{\text{FL}(n)}$, and as w admits a null-sequence that does not include repeated words, $C^{\text{FL}(n)}$ is an upper bound on the number of times relations are applied in the null-sequence and hence on $\text{Area}(w)$. The same proof applies to FFFL and so with (1) gives Theorem C.

3. P T A B

Proof of Theorem A. We exploit a technique due to the first author in [3] to show that the intrinsic diameter $\text{IDiam}(w_n)$ of $w_n = [T, a^{-b^n}\tau a^{b^n}]$ is at least 2^n.



F 2. A van Kampen diagram Δ for w_n with a subdiagram Σ.

Suppose $\pi : \Delta \to Cay^2(\mathcal{Q})$ is a van Kampen diagram for w_n. Figure 2 is a schematic depiction of Δ and Figure 3 shows an explicit example when $n = 3$. We seek an edge-path μ in Δ, along which reads a word in which the exponent sum of the letters t is 2^n. A T-corridor through Δ connects the two letters T in w_n. Along each side of this corridor we read a word u in $\left\{t^{\pm 1}, \tau^{\pm 1}\right\}^{\star}$. Let Σ be a subdiagram of Δ with boundary made up of one side of the T-corridor and a portion of $\partial\Delta$ labelled $a^{-b^n}\tau a^{b^n}$. A τ-corridor in Σ joins the τ in $a^{-b^n}\tau a^{b^n}$ to some edge-labelled τ in u. Let u_0 be the prefix of u such that the letter immediately following u_0 is this τ, and then let μ be the edge-path along the side of the τ-corridor running from the vertex at the end of a^{-b^n} to the vertex at the end of u_0. Let $v \in \left\{(at)^{\pm 1}\right\}^{\star}$ be the word along μ. Then $u_0 = a^{b^n}v$ in Q. Killing T, t and τ, retracts Q onto the subpresentation $\langle a, b \mid a^b = a^2\rangle$. So $\overline{v} = a^{b^n}$ in $\langle a, b \mid a^b = a^2\rangle$, where $\overline{v} \in \left\{a^{\pm 1}\right\}^{\star}$ is v with all letters $t^{\pm 1}$ removed. It follows that the exponent sum of the letters in $\overline{v}$ is 2^n and hence that μ has the asserted property.

Killing all generators other than t defines a retraction ϕ of Q onto $\langle t\rangle \cong \mathbb{Z}$ that is distance decreasing with respect to word metrics. But the image of $\phi \circ \pi : \Delta \to \mathbb{Z}$ has diameter at least 2^n on account of μ. So $\mathrm{IDiam}(w_n) \geq 2^n$, as claimed.

It is easy to check that the van Kampen diagram $\hat{\Delta}_n$ for w_n constructed below admits a shelling down to their base vertex that realises the bound $\mathrm{FL}(w_n) \preceq 2^n$. So, as $\mathrm{IDiam}(w_n) \leq \mathrm{FL}(w_n)$, we deduce that $\mathrm{IDiam}(w_n) \simeq \mathrm{FL}(w_n) \simeq 2^n$.

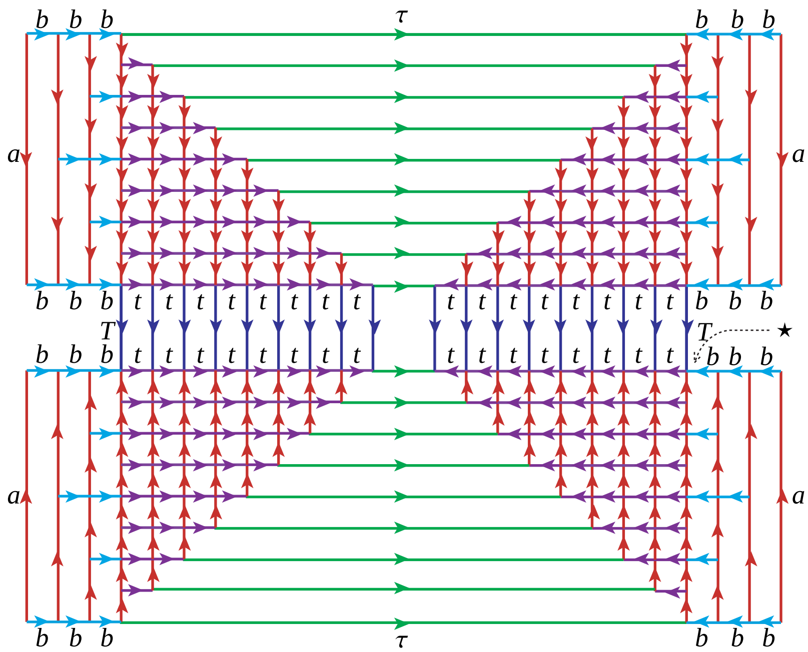
The bound $\mathrm{FFL}(w_n) \preceq n$ follows from Proposition 4.4. None-the-less we will sketch a proof since the salient ideas, developing the cylinder example from Section 1, appear here more transparently than in the more general contexts of Section 4.

The van Kampen diagram $\hat{\Delta}_3$ for w_3 is depicted in Figure 3. The analogous construction of the diagram $\hat{\Delta}_n$ for w_n should be clear. Within $\hat{\Delta}_n$ there are four triangular subdiagrams over the subpresentation $\langle a, t \mid [a, t]\rangle$, in which strings of a-edges run vertically (in the sense of Figure 3). Cut along each of these strings (except those of length 2^n at the left and right sides of the diagrams) and insert back-to-back copies of the $\langle a, b \mid a^b = a^2\rangle$-diagrams Ω_k that *shortcut* a^k to a word u_k of length $\sim \log_2 k \leq n$. These *shortcut diagrams* Ω_k are constructed in Proposition 4.1 and the way they are inserted is shown (in a more general context) in Figure 8. Call the resulting diagram Δ_n.

For $0 \leq k \leq 2^n$ let $\hat{\rho}_k$ be the edge-paths in $\hat{\Delta}_n$, forming concentric squares in Figure 3, labelled by $a^kTa^{-k}\tau^{-1}a^kT^{-1}a^{-k}\tau$. Next, for $0 \leq k \leq 2^n$ define ρ_k to be the edge-path in Δ_n that is obtained from $\hat{\rho}_k$ by replacing each subword $a^{\pm k}$ by its shortcut ${u_k}^{\pm 1}$. (In particular $\rho_{2^n} = \partial\Delta_n$.) Note that for all k, the length of ρ_k is $\preceq n$.

We now briefly describe a full shelling of Δ_n that realises the bound $\mathrm{FFL}(w_n) \preceq n$. In the course of this shelling we encounter the subdiagrams of Δ_n that have ρ_k as their boundary loops. The following lemma concerning the diagrams Ω_k of Proposition 4.1 is the key to shelling the subdiagram bounded by ρ_{k+1} down to that bounded by ρ_k. The length of the boundary loop is kept within $\preceq n$ since $\log_2 k \leq n$.

Lemma 3.1. *Let Π be the van Kampen diagram comprising a copy of Ω_{k+1} and a copy Ω_k joined to each side of a t-corridor along sides labelled a^{k+1}. Let $\star$ be a*



F 3. The van Kampen diagram $\hat{\Delta}_3$ for w_3.

vertex on $\partial\Pi$ located at the start of either of the paths labelled a^k along the sides of the t-corridor. Then $\mathrm{FL}(\Pi, \star) \leq \log_2 k$.

The proof of this lemma becomes clear when one considers concurrently running the shellings of Ω_{k+1} and Ω_k of Proposition 4.1, and a shelling of the t-corridor.

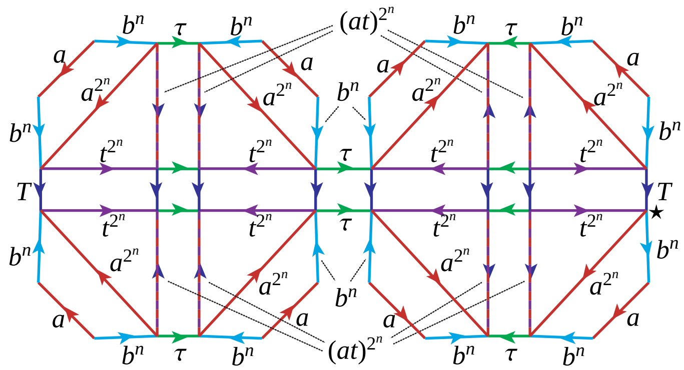
We complete the proof of Theorem A by noting that the lower bound $\mathrm{FFL}(w_n) \geq n$ is trivial as $\mathrm{FFL}(w_n) \geq \ell(w_n) = 8n + 8$, and so $\mathrm{FFL}(w_n) \simeq n$. $\square$

Proof of Theorem B. The lower bound of 2^n on $\mathrm{IDiam}(w_n)$ established above proves that $2^n \preceq \mathrm{IDiam}(n)$. So by (1) in Section 2 we see that $2^n \preceq \mathrm{FL}(n) \preceq \mathrm{Area}(n)$ and $\mathrm{FFL}(n) \preceq \mathrm{Area}(n)$ for all $n \in \mathbb{N}$.

To establish $2^n \preceq \mathrm{FFL}(n)$ we show that the words

$$w_n' := [T,\ a^{-b^n} \tau a^{b^n} \tau a^{b^n} \tau^{-1} a^{-b^n}]$$

satisfy $\mathrm{FFL}(w_n') \geq 2^n$. Figure 4 shows a van Kampen diagram for w_n' built up of a copy of the diagram $\hat{\Delta}_n$ for w_n, a $[T, \tau]$-2-cell, and a mirror image of $\hat{\Delta}_n$. Thus w_n' is null-homotopic.



F 4. A van Kampen diagram for w_n'.

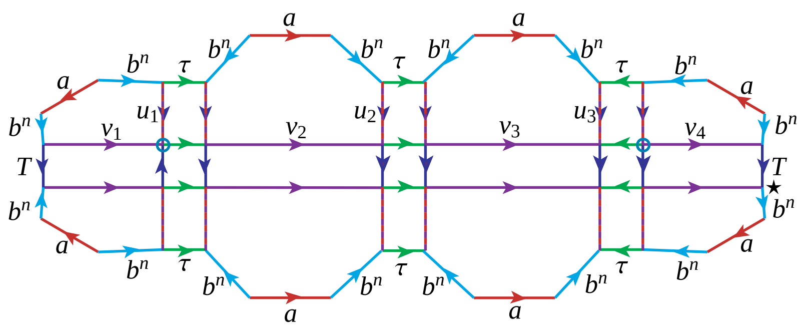
To show that $\mathrm{FFL}(\Delta') \geq 2^n$ for *all* van Kampen diagrams Δ' for w_n' we develop the argument used to establish the lower bound on diameter in the proof of Theorem A. A T-corridor runs through Δ', and as there are three occurrences of τ in w_n' and three of τ^{-1}, each τ is joined to a τ^{-1} by a τ-corridor that crosses the T-corridor. This is illustrated in Figure 5; note that generically the behaviour of the corridors could be more complex because each τ-corridor could cross the T-corridor multiple times. As shown in Figure 5, let u_1, u_2 and u_3 be the words along the sides of the initial portions of these three τ-corridors running from the first, second and third τ in w_n' and ending where the corridor first meets the T-corridor. (So $u_1 = u_3 = (at)^{2^n}$ and u_2 is the empty word in the example of Figure 4.)

Retracting $\mathcal{Q}$ onto the subpresentation $\langle a, b \mid b^{-1}ab = a^2 \rangle$ by killing T, t and τ we see that the exponent sum of the letter a, and hence also of t, in u_1 and in u_3 is 2^n, while the exponent sum in u_2 is 0. The word along the side of the T-corridor is of the form $v_1 \tau v_2 \tau v_3 \tau v_4$ where $v_i \in \left\{t^{\pm 1}, \tau^{\pm 1}\right\}^*$ and v_i runs to the vertex at the end of u_i for $i = 1, 2, 3$ (see Figure 5). By, considering the retraction of $\mathcal{Q}$ onto $\langle t \rangle$ in which a, b, τ, T are killed, we see that the exponent sum of t in v_i is 2^n for $i = 1, 4$ and is -2^n for $i = 2, 3$.

Suppose $\mathcal{S} = (\Delta_i')$ is a full shelling of Δ', in the course of which the base vertex is not required to be kept fixed. The retraction ϕ in which all generators other than t are killed, defines a distance decreasing map ϕ of $\mathcal{Q}$ onto $\langle t \rangle \cong \mathbb{Z}$. And the edge-circuit $\phi \circ \pi \big|_{\partial \Delta_i'} : \partial \Delta_i' \to \mathbb{Z}$ has length at most $\ell(\partial \Delta_i')$. There are natural combinatorial maps $\psi_i : \Delta_i' \to \Delta'$ (only prevented from being injective by *1-cell expansion* moves) under which $\psi_i(\partial \Delta_i')$ forms a contracting sequence of edge-circuits. Let i be the least integer such that $\psi_i(\partial\Delta'_i)$ includes either the vertex at the end of v_1 or at the start of v_4 – these are ringed by small circles in Figure 5. We will explain why $\ell(\partial\Delta'_i)$ is at least 2^n when this vertex x is at the end of v_1. A similar argument will show the same result to hold when x is the vertex at the start of v_4. Some vertex y on v_4 must be included in $\psi_i(\partial\Delta'_i)$ because the contracting edge-circuit cannot have yet crossed the vertex at the start of v_4. So

$$\mathrm{FFL}(w'_n) \;\geq\; \ell(\partial\Delta'_i) \;\geq\; |\phi(x) - \phi(y)| \;\geq\; 2^n,$$

as required.



F 5. A van Kampen diagram Δ' for w'_n.

It is clear that $n \geq \mathrm{FFFL}(n)$ because for all null-homotopic words w we have $\mathrm{FFFL}(w) \geq \ell(w)$. Proposition 4.7, which is the culmination of a sequence of propositions in Section 4, will give $\mathrm{FFFL}(n) \simeq n$. Theorem C will then give us $\mathrm{Area}(n) \preceq 2^n$ and the proof of the theorem will be complete. □

4. A

In this section we provide a number of results which build up to Proposition 4.7 where we prove a linear upper bound on $\mathrm{FFFL}(n)$ in the presentation $\mathcal{Q}$ of Theorem A. We begin in Proposition 4.1 with a technical result giving carefully controlled shellings of diagrams Ω_k that *shortcut* a^k in $\langle a, b \mid a^b = a^2\rangle$ to a word of length $\leq \log_2 k$. Next Proposition 4.2 claims the filling length of $\mathcal{Q}_0 := \langle a, b, t \mid a^b a^{-2}, [a, t]\rangle$ admits a linear upper bound.

Proposition 4.4 gives a linear upper bound on $\mathrm{FFL}(w)$ for null-homotopic words in $\mathcal{Q}$ that have exactly one pair of letters T, T^{-1} and one pair τ, τ^{-1}, with the occurrences of the $T^{\pm 1}$ alternating with those of the $\tau^{\pm 1}$. We show that such words have diagrams with one T- and one τ-corridor with these corridors crossing only once. These diagrams can be thought of as *towers*, and are exemplified the diagrams Δ_n for w_n constructed in the proof of Theorem A. In the proof of Proposition 4.4 we refer back to Proposition 4.3 which provides controlled shellings for the four subdiagrams that *tower* diagram breaks up into if we remove the T- and τ-corridors.

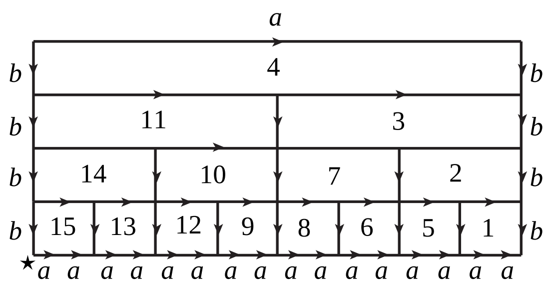
Next, in Proposition 4.5, we establish an upper bound on $\mathrm{FFFL}(w)$ that is linear in $\ell(w)$, for null-homotopic words w in Q that have exactly one pair of letters T, T^{-1}. The essential idea here is to find a diagram that can be fragmented into a number of subdiagrams, one for each τ-corridor that crosses the T-corridor, and apply Proposition 4.4 to each of these subdiagrams. Proposition 4.6 takes care of the case where there is no letter T in w. Finally, we prove Proposition 4.7: given a word that is null-homotopic in Q we construct a diagram that can be fragmented into subdiagrams each of which contain at most one T-corridor, and then we apply Propositions 4.5 and 4.6.

Proposition 4.1. *Let $\mathcal{P}$ be the presentation $\langle a, b \mid a^b = a^2\rangle$. Fix $k > 0$. There is a word u_k of length at most $12 + 4\log_2 k$ such that $u_k = a^k$ in $\mathcal{P}$. Moreover, there is a $\mathcal{P}$-van Kampen diagram Ω_k with boundary word $a^k {u_k}^{-1}$ reading anticlockwise starting from a vertex $\star$ satisfying the following. Let μ be the subarc of $\partial\Omega_k$ along which one reads a^k. There is a shelling $\mathcal{S} = \Omega_{k0}, \ldots, \Omega_{kp}$ that collapses $\Omega_k = \Omega_{k0}$ to $\star = \Omega_{kp}$, such that each Ω_{ki} is a subdiagram of Ω_k and, expressing the boundary circuit of Ω_{ki} as μ_i followed by ν_i, where μ_i is the maximal length subarc of $\partial\Omega_{ki}$ that starts at $\star$ and follows μ, we have $\ell(\nu_i) \leq 12 + 4\log_2 k$.*

Proof. Let m be the least integer such that $2^m \geq k$. So $m < 1 + \log_2 k$. Let Ξ be the standard van Kampen diagram that demonstrates the equality $a^{b^m} = a^{2^m}$ and is depicted in Figure 6 in the case $m = 4$. Let (Ξ_i) be the shelling of Ξ in which each Ξ_i is a subdiagram of Ξ and Ξ_{i+1} is obtained from Ξ_i as follows. A 1-cell collapse is performed on a *spike* of Ξ_i if possible. Otherwise, of the rightmost 2-cells in Ξ_i, let e^2 be the lowest in the sense of Figure 6. Let e^1 be the right-most edge of the lower horizontal side of e^2. Then $e_1 \in \partial\Xi_i$. Perform a 2-cell collapse on (e^2, e^1). As an illustration, the numbers in the 2-cells in Figure 6 show the order in which the 2-cells are collapsed.

When there is no spike in Ξ_i, its anticlockwise boundary circuit, starting from $\star$, follows *horizontal* (in the sense of Figure 6) edges labelled by a, then travels *upwards* through the boundary of m 2-cells (visiting at most three 1-cells of each) and then *descends* back to $\star$ along edges labelled b. Removing the initial horizontal path, the number of edges traversed is at most $4m$. And, as the total length of the 1-dimensional portions of $\partial\Xi_i$ is at most 8, and $m < 1 + \log_2 k$, we deduce that the length of $\partial\Phi_i$ minus the length of the initial horizontal arc, is at most $12 + 4\log_2 k$.

In the case $k = 2^m$ we find that defining $\Omega_i := \Xi_i$ for all $i \geq 0$ gives the asserted result. For the case $k \neq 2^m$, let c be the maximum i such that a^k is a prefix of the word w one reads anticlockwise around $\partial\Xi_i$, starting from $\star$. Then defining $u_k := {w_0}^{-1}$, where $w = a^k w_0$ and $\Omega_{ki} := \Xi_{c+i}$ for all $i \geq 0$, we have our result. $\square$



F 6. The van Kampen diagram Ξ for $a^{2^4}a^{-b^4}$.

Proposition 4.2. *The filling length function* FL : $\mathbb{N} \to \mathbb{N}$ *of*

$$\mathcal{Q}_0 \quad := \quad \langle a, b, t \mid a^b a^{-2}, [a, t] \rangle$$

admits a linear upper bound.

Proof. Corollary E1 of [2] says that an HNN extension of a finitely generated free group with finitely many stable letters, in which the associated subgroups are all finitely generated, is asynchronously automatic. This applies to $\mathcal{Q}_0$. Theorem 3.1 in [9] says (in different language) that if a group is asynchronously combable then its filling length function admits a linear upper bound. □

Proposition 4.3. *Suppose $(at)^{-k}wt^j$ is a null-homotopic word in $\mathcal{Q}_0 = \langle a, b, t \mid a^b a^{-2}, [t, a] \rangle$. Let Λ be the 1-dimensional van Kampen diagram for*

$$t^j (at)^{-k} (at)^{k-1} at^{-(j-1)}$$

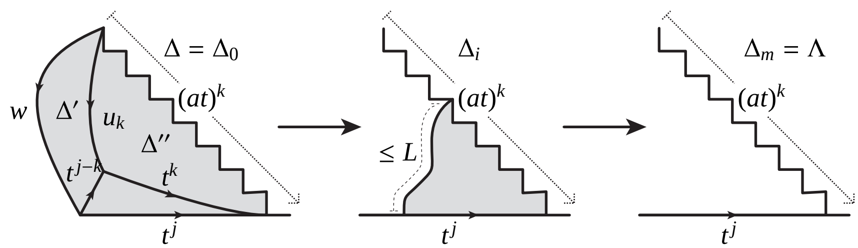
constructed by assembling 1-cells in $\mathbb{R}^2$ as depicted at the right in Figure 7. There is a $\mathcal{Q}_0$-van Kampen diagram Δ for $(at)^{-k}wt^j$ with the following properties. There is a shelling of Δ through a sequence of diagrams $\Delta = \Delta_0, \Delta_1, \ldots, \Delta_m = \Lambda$ with the portion $t^j(at)^{-k}$ of $\partial\Delta$ left undisturbed throughout (see Figure 7). Let ν_i be the maximal length arc of the boundary circuit of Δ_i contained entirely in Λ. There exists $C > 0$, depending only on $\mathcal{Q}_0$ such that $L := \max_i(\ell(\partial\Delta_i) - \ell(\nu_i)) \leq C\ell(w)$.

Proof. We consider first the case $k \geq 0$. In $\mathcal{Q}_0$ we find that $w^{-1} = t^j(at)^{-k} = t^{j-k}a^{-k} = t^{j-k}{u_k}^{-1}$, where u_k is the word of Proposition 4.1 that has length at most $12 + 4\log_2 k$. These equalities are displayed in the left-most diagram of Figure 7, which shows the framework of the van Kampen Δ: a union of a diagram Δ' for $t^{j-k}{u_k}^{-1}w$, a diagram Δ'' for $(at)^{-k}u_k t^k$ and a tripod; the lower triangular region in the figure folds up to give a tripod, the exact configuration of which depends on the relative signs of j, k and $j - k$.

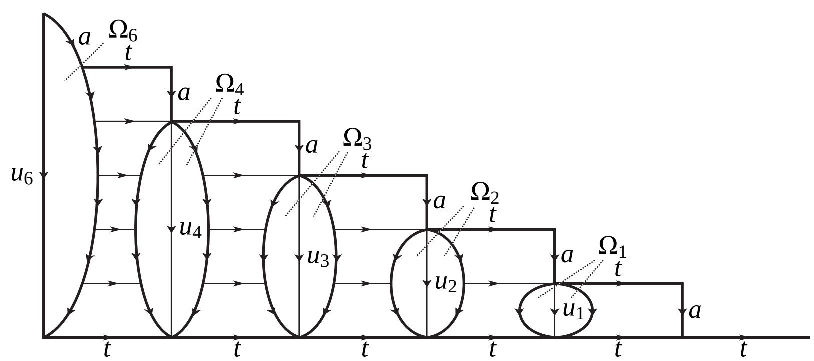
Note that $j - k = \ell_t(w) \leq \ell(w)$ as $\mathcal{Q}_0$ retracts onto $\langle t \rangle \cong \mathbb{Z}$. And $k \leq 2^{\ell(w)}$ because killing t retracts $\mathcal{Q}_0$ onto $\mathcal{P}$. It follows that $\ell(u_k) \leq 12 + 4\ell(w)$ and $\ell(t^{j-k}{u_k}^{-1}w) \leq$

$2\ell(w) + \ell(u_k) \leq 12 + 6\ell(w)$. By Proposition 4.2 we can take Δ' to be a van Kampen diagram for $t^{j-k}{u_k}^{-1}w$ with filling length at most a constant times $12 + 6\ell(w)$. We can cut along the edge-path in Δ labelled by $u_k t^{-(j-k)}$, leaving Δ' attached to the rest of the diagram at only one vertex, and then shell Δ', and in the process the length of the non-$t^j(at)^{-k}$-portion of the boundary curve has length at most a constant times $\ell(w)$.

The word $t^k(at)^{-k}a^k$ admits an obvious diagram with vertical t-corridors (as shown in Figure 8) of height $k-1, k-2, \ldots, 1$. We cut along the vertical paths labelled by powers of a and insert copies of the diagrams Ω_i of Proposition 4.1 and their mirror images, as shown in Figure 8 (illustrated in the case $k = 6$). The shellings of Lemma 3.1 can be composed to give a shelling down to Λ that realises the asserted bound on L. □



F 7. Shelling Δ down to Λ.



F 8. Δ''.

Proposition 4.4. *Suppose w is a null-homotopic word over*

$$Q := \langle\, a, b, t, T, \tau \mid a^b a^{-2}, [t,a], [\tau, at], [T,t], [\tau, T] \,\rangle$$

such that $\ell_T(w) = \ell_\tau(w) = 2$ *and the occurrences of* $T^{\pm1}$ *and* $\tau^{\pm1}$ *alternate in* w*. Then* $\mathrm{FFL}(w) \leq C\ell(w)$ *where* $C > 0$ *depends only on* $\mathcal{Q}$*.*

Proof. In any van Kampen diagram Δ for w there is one T-corridor and one τ-corridor. The condition that the occurrences of $T^{\pm1}$ and $\tau^{\pm1}$ in w alternate is equivalent to saying that these corridors cross at least once in Δ.

Take Δ to be a minimal area diagram. We argue that the two corridors cross exactly once. The words along the sides of the T and τ-corridors are of the form $\alpha_1 t^{\epsilon_1}\alpha_2 t^{\epsilon_2}\ldots t^{\epsilon_{p-1}}\alpha_p$ and $\beta_1(at)^{\mu_1}\beta_2(at)^{\mu_2}\ldots(at)^{\mu_{q-1}}\beta_q$, where $\epsilon_i, \mu_i = \pm1$ and $\alpha_i \in \left\{t^{\pm1}\right\}^\star$ and $\beta_i \in \left\{(at)^{\pm1}\right\}^\star$ for all i. Further, the α_i and β_i must be reduced because otherwise Δ would not be a reduced diagram and hence not be of minimal area. Suppose, for a contradiction, that the T and τ-corridors cross more than once. Then there is a subdiagram between the two corridors with boundary word $w_0 = uv$ where $u \in \left\{t^{\pm1}\right\}^\star$ and $v \in \left\{(at)^{\pm1}\right\}^\star$. Killing all the generators other than t retracts $\mathcal{Q}$ onto $\langle t\rangle = \mathbb{Z}$ and so $\ell_t(w_0) = 0$. It follows that $\ell_a(w_0) = 0$ because killing t, τ and T retracts $\mathcal{Q}$ onto the subpresentation $\mathcal{P}$ in which a has infinite order. So v is not freely reduced and we have a contradiction.

An additional feature of a minimal area diagram is that it contains no T or τ-annulus. This can be proved by a similar method to the above.

Conclude that Δ consists of a T-corridor, a τ-corridor and four subdiagrams of the form where Proposition 4.3 applies. Produce a new van Kampen diagram Δ' for w by replacing the four subdiagrams that minimise the length L of Proposition 4.3. A shelling of Δ' realising the asserted bound is obtained by running shellings of the four subdiagrams and the two corridors concurrently in the obvious way so that the diagram is eventually shelled to the $[\tau, T]$-2-cell, and then to a single vertex. □

Proposition 4.5. *Suppose* w *is a null-homotopic word in* $\mathcal{Q}$ *(defined above) and* $\ell_T(w) = 2$*. Then* $\mathrm{FFFL}(w) \leq C\ell(w)$ *where* $C > 0$ *depends only on* $\mathcal{Q}$*.*

Proof. Let Δ be a reduced van Kampen diagram for w. We will use the layout of the τ-corridors and the one T-corridor in Δ as a template for the construction of another van Kampen diagram Δ_2 for w that will admit a shelling realising the asserted bound.

Suppose C is a τ-corridor in Δ that does not cross the T-corridor. The word w_C along the sides of C is in $\{(at)\}^\star$ and is reduced because Δ is a reduced diagram, and so must be $(at)^k$ for some $k \in \mathbb{Z}$. Killing all defining generators other than t retracts $\mathcal{Q}$ onto $\langle t\rangle$. So $k \leq \ell(\lambda)$, where λ is a portion of the boundary circuit of Δ connecting the end points of a side of C.

If follows that if we remove any number of τ-corridors that do not cross the T-corridor from Δ, then the length of the boundary circuit of each connected component is at most $2\ell(w)$.

Suppose we remove *all* of the τ-corridors that do not cross the T-corridor from Δ. Define Δ_0 to be the connected component that contains the T-corridor. All of the other connected components have boundary words that are null-homotopic in

$$Q_0 = \langle a, b, t \mid a^b a^{-2}, [t, a] \rangle,$$

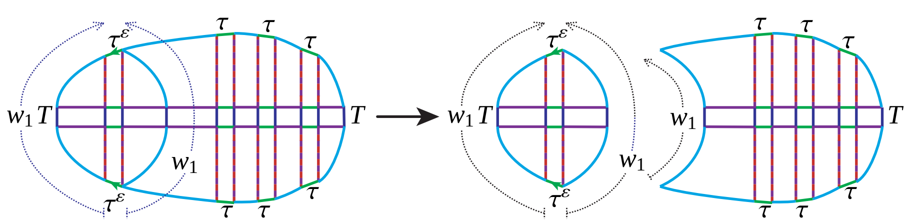
which is both a retract and a subpresentation of Q. Obtain Δ_1 from Δ by replacing all these subdiagrams by Q_0-diagrams of minimal FFL.

Repeating the following gives a shelling of Δ_1 down to Δ_0 in the course of which the boundary circuit has length at most $C_0 \ell(w)$, where C_0 is a constant that depends only on Q. Choose a τ-corridor C in Δ_1 such that, of the tow components we get by removing C, that which does not contain the T-corridor contains no τ-corridor. Cut along one side of C using *1-cell expansion* moves, and one *fragmentation* move. Next use *1-cell collapse* and *2-cell collapse* moves to remove the 2-cells along C. By the remarks above, both connected components have boundary circuits of length at most $2\ell(w)$. Collapse the component that does not contain the T-corridor down to a single vertex using a minimal FFL shelling, in the course of which the boundary circuit has length at most a constant times $2\ell(w)$ by Proposition 4.2. (In fact, a shelling down to Δ_0 within the required bound, can be achieved without the fragmentation move if care is taken over base points.)

It remains to show that the word w_0 around $\partial \Delta_0$ admits a van Kampen diagram with a full, fragmenting, free shelling in which the sum of the lengths of the boundaries of the components are at most a constant times $\ell(w_0)$. For then we can take Δ_2 to be Δ_1 with Δ_0 replaced by this diagram.

First suppose $\ell_\tau(w_0) = 0$. Then w_0 is null-homotopic in the retract $Q_1 := \langle\, a, b, t, T \mid a^b a^{-2}, [t, a], [T, t] \,\rangle$, and the length of the T-corridor in any reduced Q_1-diagram Δ_0' for w_0 is at most $\ell(w_0)/2$ on account of the retraction onto $\langle t \rangle$. Assume that the two components of Δ_0' we get on removing the T-corridor are Q_0-diagrams of minimal FFL. Then we can collapse Δ_0' by shelling each of these components and the T-corridor in turn, and using Proposition 4.2 it is easy to check that the length of the boundary circuit remains at most a constant times $\ell(w_0)$.

Next suppose $\ell_\tau(w_0) = 2$. Then Proposition 4.4 applies and gives us the result we need.



F 9. Shelling away one τ-corridor.

Finally, suppose $\ell_\tau(w_0) > 2$. Then there is a subword $\tau^\varepsilon w_1 \tau^{-\varepsilon}$ in w_0, where $\varepsilon = \pm 1$, $\ell_\tau(w_1) = 0$ and $\ell_T(w_1) = 1$. As $\tau^\varepsilon w_1 \tau^{-\varepsilon} = w_1$ in Q, there is a Q-van Kampen diagram for w_0 that we can shell by cutting along an edge-path labelled by w_1 to cut the diagram into two, as shown in Figure 9, and the shelling the two components. One of these components is a diagram for $\tau^\varepsilon w_1 \tau^{-\varepsilon} {w_1}^{-1}$, and this we shell first as per Proposition 4.4. The remaining component has boundary length $\ell(w_0) - 2$ and includes two fewer letters $\tau^{\pm 1}$, and so by continuing inductively we can find a shelling for which FFFL is at most a constant times $\ell(w_0)$. □

Proposition 4.6. *Suppose w is a null-homotopic word in*

$$Q_2 := \langle\, a, b, t, \tau \mid a^b a^{-2}, [t, a], [\tau, at] \,\rangle.$$

Then $\mathrm{FL}(w) \leq C\ell(w)$ *where $C > 0$ depends only on Q_2.*

Proof. The method used in the proof of Proposition 4.4, to reduce to the case where all the τ-corridors cross the T-corridor, gives this result. □

Proposition 4.7. *Suppose w is a null-homotopic word in Q. Then* $\mathrm{FFFL}(w) \leq C\ell(w)$ *where $C > 0$ depends only on Q.*

Proof. The cases where $\ell_T(w) = 0$ and $\ell_T(w) = 2$ are dealt with by Propositions 4.6 and 4.4, respectively. For the case $\ell_T(w) > 2$ we take a similar approach to that used in the proof of Proposition 4.4 to control $\mathrm{FFFL}(w_0)$.

There is a subword $T^\varepsilon w_1 T^{-\varepsilon}$ in w, where $\varepsilon = \pm 1$ and $T^\varepsilon w_1 T^{-\varepsilon} = w_1$ in Q. So we can find a Q-van Kampen diagram for w that can be severed into two components, one of which is a diagram for $T^\varepsilon w_1 T^{-\varepsilon} {w_1}^{-1}$, and the other of which is a diagram for a word of length $\ell(w) - 2$ that has two fewer letters $T^{\pm 1}$. The former of these two components can be shelled as per Proposition 4.4. Continuing inductively we see that the other component can be taken to be a diagram that admits a shelling in which the boundary circuit has length at most a constant times $\ell(w)$. □

5. Q -

In this section we prove Theorem D. Our approach is to monitor how filling length, in its three guises, behaves in the standard proof that finite presentability is a quasi-isometry invariant [5, page 143]. As careful quantified versions of this proof are well established ([1], addressing Area, is the first in print), our exposition here will be brief.

We have quasi-isometric groups Γ and Γ' with finite presentations $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ and $\mathcal{P}' = \langle \mathcal{A}' \mid \mathcal{R}' \rangle$, respectively. So there is a quasi-isoemtry $f : (\Gamma, d_{\mathcal{A}}) \to (\Gamma', d_{\mathcal{A}'})$ with quasi-inverse $g : (\Gamma', d_{\mathcal{A}'}) \to (\Gamma, d_{\mathcal{A}})$.

We begin by showing $\mathrm{FL}_{\mathcal{P}} \simeq \mathrm{FL}_{\mathcal{P}'}$.

Suppose ρ' is an edge-circuit in the Cayley graph of $\mathcal{P}'$, visiting vertices $v_0, v_1, \ldots, v_n = v_0$ in order. Consider a circuit ρ in the Cayley graph $Cay^1(\mathcal{P})$

of $\mathcal{P}$ obtained by joining the successive vertices of $g(v_0), g(v_1), \ldots, g(v_n)$ by geodesics. Note that $\ell(\rho)$ is at most a constant times $\ell(\rho')$. Fill ρ with a van Kampen diagram Δ over $\mathcal{P}$ admitting a shelling $\mathcal{S}$ of filling length $\mathrm{FL}_{\mathcal{P}}(\ell(\rho))$ and use f to map $\Delta^{(0)}$ to $\mathcal{P}'$. Then join $f(a)$ to $f(b)$ by a geodesic whenever a and b are the end points of an edge in Δ. The result is a combinatorial map $\pi_0' : {\Delta_0'}^{(1)} \to Cay^1(\mathcal{P}')$ filling a loop ρ_0', where Δ_0' is obtained from Δ by subdividing each of its edges into edge-paths of length at most some constant.

Interpolate between ρ' and ρ_0' by joining v_i to $f(g(v_i))$ for every i, to build a map $\pi_1' : {\Delta_1'}^{(1)} \to Cay^1(\mathcal{P}')$ where Δ_1' is obtained from Δ_0' by attaching an annulus A of n 2-cells around the boundary.

A shelling $\mathcal{S}_1'$ of Δ_1' down to the base vertex with $\ell(\mathcal{S}_1')$ at most a constant times $(1 + \mathrm{FL}_{\mathcal{P}}(\ell(\rho)))$ is obtained by first shelling away A, to leave just a stalk from v_0 to $f(g(v_0))$, and then running a shelling of Δ_0' modelled on $\mathcal{S}$: where $\mathcal{S}$ demands the collapse of a 1-cell in Δ collapse all the 1-cells in the corresponding edge-path in Δ_0'.

However, whilst the words around the 2-cells of Δ_1' are null-homotopic they may fail to be in $\mathcal{R}'$, so Δ_1' may not be a van Kampen diagram over $\mathcal{P}'$. To rectify this we will replace the 2-cells of Δ_1' by van Kampen diagrams over $\mathcal{P}'$, each of area at most some uniform constant. A concern here is that van Kampen diagrams can be singular 2-discs and so gluing them in place of 2-cells may destroy planarity. This is dealt with by replacing the 2-cells of Δ_1' one at a time; on each occasion, if the 2-cell e^2 to be replaced has non-embedded boundary circuit then we discard all the edges inside the simple edge-circuit σ in ∂e^2 such that no edge of ∂e^2 is outside σ, and then we fill σ. The result is a van Kampen diagram Δ_2' for ρ' over $\mathcal{P}'$. We obtain a shelling $\mathcal{S}_2'$ for Δ_2' by altering $\mathcal{S}_1'$: each time we discarded some connected component of the set of edges inside some σ we contract it (more strictly, its pre-image) and all the 2-cells it encloses to a single vertex in every one of the diagrams comprising the shelling, and each time we fill some 2-cell e^2 with a van Kampen diagram D we shell out all of D when we had been due to perform a 2-cell collapse move on e^2. [This could fail to give a shelling when a 2-cell collapse move in $\mathcal{S}_1'$ removes a pair (e^2, e^1), where e^1 is one of the now contracted edges – but then a 1-cell expansion followed by a 2-cell collapse producing the same effect can be used instead.] The difference between $\ell(\mathcal{S}_1')$ and $\ell(\mathcal{S}_2')$ is then at most some additive constant. Deduce that $\mathrm{FL}_{\mathcal{P}'} \preceq \mathrm{FL}_{\mathcal{P}}$. Interchanging $\mathcal{P}$ and $\mathcal{P}'$ we have $\mathrm{FL}_{\mathcal{P}} \preceq \mathrm{FL}_{\mathcal{P}'}$ and so $\mathrm{FL}_{\mathcal{P}} \simeq \mathrm{FL}_{\mathcal{P}'}$.

The proof that $\mathrm{FFL}_{\mathcal{P}} \simeq \mathrm{FFL}_{\mathcal{P}'}$ is essentially the same, except we consider free shellings, we discard the stalk between v_0 and $f(g(v_0))$, and we replace $\mathrm{FL}_{\mathcal{P}}(\ell(\rho))$ by $\mathrm{FFL}_{\mathcal{P}}(\ell(\rho))$. To show $\mathrm{FFFL}_{\mathcal{P}} \simeq \mathrm{FFFL}_{\mathcal{P}'}$ we additionally allow free and fragmenting shellings and we use $\mathrm{FFFL}_{\mathcal{P}}(\ell(\rho))$ in place of $\mathrm{FL}_{\mathcal{P}}(\ell(\rho))$; no further technical concerns arise.

6. R

Suppose $c : [0,1] \to X$ is a loop in a metric space X.

A *based null-homotopy* H of c is a continuous map $H : [0,1]^2 \to X$ for which $H(0,t) = H(1,t) = c(0)$ for all t and, defining $H_t : [0,1] \to X$ by $H_t(s) = H(s,t)$, we have $H_0 = c$ and $H_1(t) = c(0)$ for all t.

A *free null-homotopy* H of c is a continuous map $H : [0,1]^2 \to X$ such that $H(0,t) = H(1,t)$ for all t, and H_0 and H_1 are c and a constant function, respectively.

Let $\mathcal{S}$ be the set of subspaces S of $[0,1]^2$ that have $[0,1] \times \{0\} \subset S$ and are the union of a finite family of solid closed triangles $\Delta_i = [(a_i, 0), (b_i, 0), (c_i, d_i)]$ with $c_i \in [a_i, b_i]$ and $d_i \in (0,1]$. The fibres S_t of the projection mapping points in S to their second co-ordinate are disjoint unions $\bigsqcup_{i=1}^{k_t} I_{t,i} \times \{t\}$ of closed intervals $I_{t,i}$. At finitely many critical t-values τ_i, intervals comprising the fibres bifurcate or collapse to a point.

A *free and fragmenting null-homotopy* H of c is a continuous map $H : S \to X$ for some $S \in \mathcal{S}$ where, defining H_t to be the restriction of H to S_t, we find that $H_0 = c$, that H_1 is constant on $I_{1,i}$ for all i, and that $H_t(x) = H_t(y)$ for all t, whenever x and y are the end points of some $I_{t,i}$. We define $\ell(H_t)$ to be the sum of the lengths of the k_t loops in X defined by H_t. Note that taking S to be a single triangle reduces to the case of a free null-homotopy.

In each of the three settings above define $\ell(H) := \sup_{t \in [0,1]} \ell(H_t)$, and then

$$\begin{aligned}
\mathrm{FL}(c) &= \inf \{\, \ell(H) \mid \text{ based null-homotopies } H \text{ of } c \,\} \\
\mathrm{FFL}(c) &= \inf \{\, \ell(H) \mid \text{ free null-homotopies } H \text{ of } c \,\} \\
\mathrm{FFFL}(c) &= \inf \{\, \ell(H) \mid \text{ free and fragmenting null-homotopies } H \text{ of } c \,\}.
\end{aligned}$$

For $\mathrm{M} = \mathrm{FL}, \mathrm{FFL}$ or FFFL, define $\mathrm{M}_X : [0,\infty) \to [0,\infty)$ by

$$\mathrm{M}_X(l) = \sup \{\, \mathrm{M}(c) \mid \text{null-homotopic loops } c \text{ with } \ell(c) \leq l \,\}.$$

The following lemma gives sufficient conditions for FL_X, FFL_X and FFFL_X to be well-defined – conditions enjoyed by the universal cover of any closed connected Riemannian manifold, for example.

Lemma 6.1. *Suppose X is the universal cover of a compact geodesic space Y for which there exist $\mu, L > 0$ such that every loop of length less than μ admits a based null-homotopy of filling length less than L. Then FL_X, FFL_X and FFFL_X are well-defined functions $[0,\infty) \to [0,\infty)$.*

Proof. The proof of Lemma 2.2 in [6] can readily be adapted to this context. In brief, we first show that every rectifiable loop c in X admits a based null-homotopy with finite filling length – apply a compactness argument to an arbitrary based null-homotopy for c to partition c into finitely many loops of length at most μ; by hypothesis each such loop has finite filling length and it follows that c has finite filling length.

Next suppose c has length l and assume (by shrinking μ if necessary) that balls of radius μ in Y lift to X. Cover Y with a maximal collection of disjoint balls of radius $\mu/10 > 0$; let $\Lambda \subset X$ be the set of lifts of their centres. Subdivide c into $m \leq 1 + 10\ell/\mu$ arcs with end-points v_i, each of length at most $\mu/10$; each v_i lies within $\mu/5$ of some $u_i \in \Lambda$; form a piece-wise geodesic loop c' approximating c by connecting-up these u_i. Loops made up of the portion of c from v_i to v_{i+1} and geodesics $[u_i, u_{i+1}]$, $[u_i, v_i]$ and $[u_{i+1}, v_{i+1}]$ have length at most μ, and so homotopy discs for these loops together form a *collar* between c and c'. By passing across these discs one at a time, it is possible to homotop c across the collar to a loop made up of c' and a stalk of length $\mu/5$ from $c(0)$ to a u_0, encountering loops only of length at most a constant (depending on L and μ) times l en route. Modulo the action of $\pi_1 Y$, there are only finitely many such piecewise geodesic loops such as c' and, by our earlier argument, each one admits a filling of finite filling length. It follows that FL_X, and hence FFL_X and FFFL_X, are well-defined functions. □

Proof of Theorem E. Fix a basepoint $p \in X$. Define a quasi-isometry Φ mapping the Cayley graph of $\mathcal{P} = \langle \mathcal{A} \mid \mathcal{R} \rangle$ to X by choosing a geodesic from p to its translate $a \cdot p$ for each $a \in \mathcal{A}$, and then extending equivariantly. Let Ψ be a quasi-isometry from X to Γ sending $x \in X$ to some γ such that $\gamma.p$ is a point of $\Gamma.p$ closest to x.

A path in X is called *word-like* (following [4]) if it is the image in X of an edge-path in the Cayley graph. For each $r \in \mathcal{R}$, let c_r denote the word-like loop in X, based at p that is the image of an edge-circuit in the Cayley graph labelled r. Map the Cayley 2-complex of $\mathcal{P}$ to X by choosing a disc-filling arising from a based null-homotopy of finite filling length for each c_r.

We will show first that $\mathrm{FL}_X \preceq \mathrm{FL}_{\mathcal{P}}$, $\mathrm{FFL}_X \preceq \mathrm{FFL}_{\mathcal{P}}$ and $\mathrm{FFFL}_X \preceq \mathrm{FFFL}_{\mathcal{P}}$. As in the proof of Lemma 6.1, a collar between an arbitrary rectifiable loop c in X and a word-like loop c', can be used to show there is no change in the $\simeq$ classes of FL_X, FFL_X or FFFL_X if one takes the suprema in their definitions to be over fillings of word-like loops only: for FL_X one notes that c can be homotoped across the collar to a loop based at $c(0)$ that is obtained from c' by attaching a stalk from $c(0)$ to $c'(0)$, and one need pass through loops of length no more than $C\ell(c) + C$ en route, where C is a constant independent of c; for FFL_X and FFFL_X, the stalk is abandoned and the homotopy is between c and c'.

One gets an upper bound on the filling length of a word-like loop c in X by taking the image in X of a minimal filling length van Kampen diagram Δ. The progress of the boundary circuit in the course of a shelling of Δ dictates a sequence of stages in a null-homotopy of c. Using Lemma 6.1, we can interpolate between these stages in a way that increases the length of the curve by no more than an additive constant, and so we get $\mathrm{FL}_X \preceq \mathrm{FL}_{\mathcal{P}}$. The proof that $\mathrm{FFL}_X \preceq \mathrm{FFL}_{\mathcal{P}}$ and $\mathrm{FFFL}_X \preceq \mathrm{FFFL}_{\mathcal{P}}$ can be completed likewise.

Now we address $\mathrm{FL}_{\mathcal{P}} \preceq \mathrm{FL}_X$. Consider a word-like loop $c : [0,1] \to X$ corresponding to a null-homotopic word w over $\mathcal{P}$ of length n. Fix a constant $\lambda > \max_{a \in \mathcal{A}} d_X(p, a.p)$. Then $\ell(c) \leq \lambda n$. Let $H : [0,1]^2 \to X$ be a based null-homotopy of c with filling length at most $1 + \mathrm{FL}_X(\lambda n)$. By uniform continuity, there exists $\varepsilon > 0$ such that $\varepsilon^{-1} \in \mathbb{Z}$ and $d_X(H(a), H(b)) \leq 1$ for all $a, b \in [0,1]^2$ with $d_{\mathbb{E}^2}(a, b) \leq \varepsilon$.

Subdivide $[0,1]^2$ into ε^{-1} rectangles separated by the lines $t = t_j$ where $t_j = j\varepsilon$ and $j = 0, 1, \ldots, \varepsilon^{-1}$. For all such $t = t_j$, take $0 = s_{t,0} < s_{t,1} < \ldots < s_{t,k_t} = 1$ in such a way that for all i, the restriction of H_t to $[s_{t,i}, s_{t,i+1}]$ is an arc of length at most λ and $k_t \leq 1 + \ell(H_t)/\lambda$. Mark the points $s_{t,0}, \ldots, s_{t,k_t}$ onto each of the lines $t = t_j$. Then, for all $j = 0, 1, \ldots, \varepsilon^{-1} - 1$ and all $i = 1, 2, \ldots, k_{t_j} - 1$, join $(s_{t,i}, t_j)$ to $(s_{t,i'}, t_{j+1})$ by a straight-line segment where $(s_{t,i'}, t_{j+1})$ is the first marked point reached from by $(s_{t,i}, t_{j+1})$ by increasing the s-co-ordinate. Note that

$$d(H_{t_j}(s_{t,i}), H_{t_{j+1}}(s_{t,i'})) \ \leq \ 1 + \lambda. \tag{3}$$

In the same way, for all $j = 1, \ldots, \varepsilon^{-1}$ and for all $i = 1, \ldots, k_{t_j} - 1$ such that $(s_{t,i}, t_j)$ is not the terminal vertex of one of the edges we just connected, join $(s_{t,i}, t_j)$ to some $(x_{t,i'}, t_{j-1})$ for which

$$d(H_{t_j}(s_{t,i}), H_{t_{j-1}}(s_{t,i'})) \ \leq \ 1 + \lambda, \tag{4}$$

so as to produce a diagram Δ in which every 2-cell has boundary circuit of combinatorial length at most 4.

Orient every edge e of Δ arbitrarily and define

$$g_e \ := \ \Psi(H_{t'}(s'))\Psi(H_t(s))^{-1}$$

when the initial and terminal points of e are (s, t) and (s', t'), respectively. It follows from (3) and (4) that g_e has word length $|g_e|$ at most some constant K with respect to $\mathcal{A}$. Subdivide e into a path of $|g_e|$ edges; give each of these new edges an orientation and a labelling by a letter in $\mathcal{A}$ so that one reads a word representing g_e along the path. Make all the choices in the construction above in such a way that w labels the line $t = 0$ and all the other edges in $\partial\Delta$ are labelled by e.

The shelling (Δ_i) of Δ which strips away the rectangles from left to right, shelling each in turn from top to bottom, has

$$\max_i \ell(\Delta_i) \ \leq \ K(1 + \mathrm{FL}_X(\lambda n)) + 4K.$$

Let w_i be the word one reads around the boundary circuit of Δ_i. Each 2-cell in each Δ_i has boundary circuit labelled by a null-homotopic word that may not be in $\mathcal{R}$, but has length at most $4K$. So it is possible to interpolate between the w_i to produce a null-sequence (see Section 2) with respect to $\mathcal{P}$ for $w = w_0$ in which every word has length at most $K(1 + \mathrm{FL}_X(\lambda n)) + 4K$ plus a constant. Thus $\mathrm{FL}_{\mathcal{P}} \preceq \mathrm{FL}_X$, as required.

That $\mathrm{FFL}_{\mathcal{P}} \preceq \mathrm{FFL}_X$ can be proved in the same way. The argument needs to be developed further to show that $\mathrm{FFFL}_{\mathcal{P}} \preceq \mathrm{FFFL}_X$. Given the word-like loop c, one takes a free and fragmenting null-homotopy $H : S \to X$ of c with $\ell(H)$ at most $\mathrm{FFFL}_X(\ell(c)) + 1$ *and* with the property that whenever loops of length less than some prior fixed constant appear, those loops are contracted to points before any further bifurcations occur. This implies that for all t, the number of connected components k_t in the fibre $S_t = \bigsqcup_{i=1}^{k_t} I_{t,i} \times \{t\}$ is at most a constant times $(1 + \ell(H_t))$. For the construction of Δ we inscribe S with the arcs of its intersection with the lines $t = t_j$ and with the additional lines $t = \tau_i$, where τ_i are the critical t-values of H. Using H_t and Ψ as before, we subdivide the fibre S_t into edges and label each of its k_t connected components by a null-homotopic word – this works as before, except we additionally insist that the end points of the closed intervals $I_{t,i}$ comprising S_t be included amongst the s_i – this may add k_t to the total length the words along S_t, but the argument given above ensures that this additional cost is no more than a constant times $(1 + \ell(H_t))$. □

R

[1] J. M. Alonso. Inégalitiés isopérimétriques et quasi-isométries. *C.R. Acad. Sci. Paris Ser. 1 Math.*, 311:761–764, 1990.

[2] G. Baumslag, S. M. Gersten, M. Shapiro, and H. Short. Automatic groups and amalgams. *J. Pure Appl. Algebra*, 76(3):229–316, 1991.

[3] M. R. Bridson. Asymptotic cones and polynomial isoperimetric inequalities. *Topology*, 38(3):543–554, 1999.

[4] 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.

[5] M. R. Bridson and A. Haefliger. *Metric Spaces of Non-positive Curvature*. Number 319 in Grundlehren der mathematischen Wissenschaften. Springer Verlag, 1999.

[6] M. R. Bridson and T. R. Riley. Extrinsic versus intrinsic diameter for Riemannian filling-discs and van Kampen diagrams. preprint, 2005.

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

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

[9] S. M. Gersten. Asynchronously automatic groups. In Charney, Davis, and Shapiro, editors, *Geometric group theory*, pages 121–133. de Gruyter, 1995.

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

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

[12] 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.

[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.

M R. B
Mathematics, Huxley Building, Imperial College, London, SW7 2AZ, UK
m.bridson@imperial.ac.uk, http://www.ma.ic.ac.uk/~mbrids/

T R. R
Mathematics Department, 310 Malott Hall, Cornell Univ., Ithaca, NY 14853-4201, USA
tim.riley@math.cornell.edu, http://www.math.cornell.edu/~riley/