RANK GRADIENT OF CYCLIC COVERS

JASON DEBLOIS

ABSTRACT. If M is an orientable hyperbolic 3-manifold with finite volume and $\phi\colon \pi_1(M) \twoheadrightarrow \mathbb{Z}$, the family of covers corresponding to $\{\phi^{-1}(n\mathbb{Z}) \,|\, n \in \mathbb{N}\}$ has rank gradient 0 if and only if the Poincaré–Lefschetz dual of the class in $H^1(M;\mathbb{Z})$ corresponding to ϕ is represented by a fiber. This generalizes a theorem of M. Lackenby. If M is closed, we give an explicit lower bound on the rank gradient. The proof uses an acylindrical accessibility theorem due to R. Weidmann and the following result: if M is a closed, orientable hyperbolic 3-manifold and S is a connected incompressible surface that is not a fiber or semi-fiber of M, then the $\pi_1 M$-action on the tree determined by S is $(14g-12)$-acylindrical, where g is the genus of S.

By the *rank* of a manifold M, $\mathrm{rk}\, M$, we will refer to the rank of its fundamental group; that is, the minimal cardinality of a generating set. Given a fixed closed manifold M, the *rank gradient* of a family of covers $\{M_n \to M\}$, each with finite degree, is defined as

$$\mathrm{rg}\,\{M_n\} \doteq \inf_n \frac{\mathrm{rk}\, M_n}{\deg\{M_n \to M\}}.$$

This was defined by M. Lackenby in [14]. Given a generating set for $\pi_1 M$ with cardinality k and a subgroup Γ of index n, the classical *Reidemeister–Schreier process* produces a generating set for Γ with cardinality $n(k-1)+1$. Thus we have:

Fact. (Reidemeister–Schreier) $\mathrm{rg}\,\{M_n \to M\} \leq \mathrm{rk}\, M - 1$.

The rank gradient of $\{M_n \to M\}$ is nonzero precisely when $\mathrm{rk}\, M_n$ grows linearly with the degree of the covering map. The main theorem of this note describes a family of covers for which this may be characterized.

Theorem 0.1. *Let M be an orientable hyperbolic 3-manifold of finite volume with a homomorphism $\phi\colon \pi_1 M \twoheadrightarrow \mathbb{Z}$, and for $n \in \mathbb{N}$ let $M_n \to M$ be the cover corresponding to $\phi^{-1}(n\mathbb{Z}) < \pi_1 M$. Then $\mathrm{rg}\,\{M_n\} = 0$ if and only if the Poincaré dual of the class in $H^1(M;\mathbb{Z})$ corresponding to ϕ is represented by a fiber.*

Since $\mathbb{Z}$ is abelian, ϕ factors through the abelianization of $\pi_1 M$; after identifying this with $H_1(M;\mathbb{Z})$ we obtain from ϕ a class in $H^1(M;\mathbb{Z})$. The Poincaré–Lefschetz dual of this class lies in $H_2(M, \partial M;\mathbb{Z})$; a standard argument implies it is represented by a surface properly embedded in M (see Lemma 1.1 and below it). This surface is a *fiber* if it is a union of point preimages under some submersion $M \to S^1$.

Theorem 0.1 directly generalizes a theorem of Lackenby [15, Theorem 1.11], which asserts the corresponding alternative for the *Heegaard gradient* (see below). One direction is trivial, since a closed surface with genus g has rank $2g$:

Fact. If $\mathrm{PD}(\phi)$ is represented by a fiber surface of genus g, then

$$\mathrm{rk}\, M_n \leq 2g+1$$

Partially supported by NSF grant DMS-0703749.

for each $n \in \mathbb{N}$. In particular, $\mathrm{rg}\,\{M_n\} = 0$.

The cusped case of Theorem 0.1 follows from the closed case and a Dehn surgery argument. In the closed case, our argument yields an explicit bound for the rank gradient depending only on the Thurston norm of ϕ (see [25]).

Theorem 0.2. *Let M be a closed, orientable hyperbolic 3-manifold with a homomorphism $\phi\colon \pi_1 M \twoheadrightarrow \mathbb{Z}$, and for $n \in \mathbb{N}$ let $M_n \to M$ be the cover corresponding to $\phi^{-1}(n\mathbb{Z}) < \pi_1 M$. If $\mathrm{PD}(\phi)$ is not represented by the fiber of a fibration $M \to S^1$, then $\mathrm{rg}\,\{M_n\} \geq \frac{1}{2k}$, where*

$$k = 14\|\phi\| - 12$$

and $\|\phi\|$ is the Thurston norm of ϕ.

Theorem 0.2 is an application of two deep but well known principles. The first, known as "acylindrical accessibility," bounds the cardinality of generating sets of groups acting nicely on trees. The second, that "cylinders have bounded length," is a property of the JSJ decomposition of a manifold obtained from a hyperbolic 3-manifold by cutting along an incompressible surface. We will describe these principles, use them to prove Theorem 0.2, and from this derive Theorem 0.1, in Section 1.

The version of "cylinders have bounded length" that we use here, recorded as Theorem 1.4, to our knowledge has not previously appeared in print. We will describe the JSJ decomposition and prove Theorem 1.4 in Section 3. We use the following consequence of it, which may be of independent interest.

Corollary 1.5. *Let M be a closed, orientable hyperbolic 3-manifold and S a connected, two-sided incompressible surface in M, with genus g, that is not a fiber or semi-fiber. The $\pi_1 M$-action on the tree determined by S is $(14g - 12)$-acylindrical.*

A connected surface $S \subset M$ is a *semi-fiber* if it separates M into a disjoint union of twisted I-bundles over the non-orientable surface double covered by S. If S is a semi-fiber, then there is a twofold cover $\widetilde{M} \to M$ such that S lifts to a fiber of $\widetilde{M}$ in the sense defined below Theorem 0.1. It is easy to see that if S is a fiber or a semi-fiber, the tree T determined by S is homeomorphic to a line, and each element of $\pi_1 S < \pi_1 M$ fixes all of T. Hence it is necessary in the statement of Corollary 1.5 to assume that S is not a fiber or semi-fiber.

Our considerations here are mainly motivated by the "rank versus Heegaard genus" question for 3-manifolds. The *Heegaard genus* of a closed, orientable 3-manifold M, $\mathrm{Hg}\, M$, is the minimum genus of a separating surface S embedded in M so that if $V \subset M$ is a component of $M - S$, then $\overline{V}$ is homeomorphic to a *handlebody*, obtained from a 3-dimensional ball by attaching 1-handles. Such an S is called a *Heegaard surface* for M, and it is a classical theorem that every closed 3-manifold contains one.

If M has a Heegaard surface S of genus g, then the inclusion map $\overline{V} \to M$ induces a surjection at the level of π_1, where $\overline{V}$ is one of the complementary handlebodies to S. Since $\overline{V}$ has g 1-handles, we have $\mathrm{rk}\, M \leq g$. Thus $\mathrm{rk}\, M \leq \mathrm{Hg}\, M$.

Question. Does the rank of a closed, orientable hyperbolic 3-manifold equal its Heegaard genus?

This question was originally asked by Haken [7] and then Waldhausen [27], without any hyperbolicity requirements. However, Boileau–Zieschang described a family of Seifert fibered spaces with rank 2 but Heegaard genus 3 in [2]. More recently, graph manifolds have been described with rank at most $3n$ and Heegaard genus $4n$, for $n \in \mathbb{N}$ [17].

There are no corresponding results for hyperbolic 3-manifolds; on the contrary, there is evidence that at least among some families, the geometry forces rank to equal Heegaard genus. For example, Agol showed that all but finitely many closed hyperbolic 3-manifolds with rank 2 and injectivity radius bounded away from 0 have Heegaard genus 2 (cf. [23, §9]). In this spirit we offer the following conjecture. Below we define the *Heegaard gradient* $\mathrm{Hgr}\{M_n \to M\}$ of a family of finite-degree covers in analogy with the rank gradient, replacing rank by Heegaard genus.

Conjecture. Let M be a closed, orientable hyperbolic 3-manifold and $\{M_n \to M\}$ a family of finite-degree covers. Then $\mathrm{rg}\,\{M_n\} > 0$ if and only if $\mathrm{Hgr}\,\{M_n\} > 0$.

Theorem 0.1 may be considered evidence for this conjecture by comparison with [15, Theorem 1.11]. We will discuss further directions and questions in Section 4.

Acknowledgment

Many thanks to Peter Shalen for describing "cylinders have bounded length" to me, for pointing me to [3], and for helping me extend the vegematic argument to the situation under consideration here.

1. Actions on trees

Given a finitely generated group Γ acting on a tree T, an "accessibility" principle relates the combinatorics of $\Gamma\backslash T$ to the structure of Γ. *Acylindrical* accessibility, introduced by Z. Sela [21], is distinguished from other notions in that it does not require prior knowledge of the structure of vertex or edge stabilizers, but only that their action on T is "nice enough":

Definition. $\Gamma \times T \to T$ is *k-acylindrical* if no $g \in \Gamma - \{1\}$ fixes a segment of length greater than k, and *k-cylindrical* otherwise.

It is a basic consequence of Bass–Serre theory that if a group Γ acts on a tree with a trivial edge stabilizer, Γ is freely decomposable or cyclic. The groups of interest here therefore act at best 1-acylindrically on trees. The acylindrical accessibility theorem that we use here, due to R. Weidmann, has constants that do not depend on the group Γ, a feature that is critical to our rank gradient computations.

Theorem ([28])**.** *Let Γ be a non-cyclic freely indecomposable finitely generated group and $\Gamma \times T \to T$ a minimal k-acylindrical action. Then $\Gamma\backslash T$ has at most $1 + 2k(\mathrm{rk}\,\Gamma - 1)$ vertices.*

We will apply this theorem to an action associated to $\phi\colon \pi_1 M \to \mathbb{Z}$, where M is a compact, orientable 3-manifold. Since 3-manifolds are triangulable and the circle has contractible universal cover, there is a map $p\colon M \to S^1$ such that $p_* = \phi$ (cf. [9, Proposition 1B.9]). Applying [10, Lemma 6.5]), we may homotop p so that for some $x \in S^1$, $p^{-1}(x)$ is a (possibly disconnected) 2-sided incompressible surface S in M.

A surface $S \subset M$ is *incompressible* if it is embedded so that each component is π_1-injective, and no component is a 2-sphere that bounds a 3-ball. Lemma 6.5 of [10] further implies that p is modeled on a closed regular neighborhood $\mathcal{N}$ of S, homeomorphic to $S \times I$, by projection to I followed by a homeomorphism to an interval around x in S^1. Taking S^1 to be the topological space underlying a graph G with a single vertex v and edge e, after a further homotopy we may take p to map $\mathcal{N}$ onto e and $\overline{M - \mathcal{N}}$ to v.

There is a graph G_0, with an edge for each component of $\mathcal{N}$ and a vertex for each component of $\overline{M - \mathcal{N}}$, such that p factors through a projection $G_0 \to G$. Shrinking a maximal tree in G_0, and tracking this with a final homotopy of p, we obtain:

Lemma 1.1. *Suppose M is a compact, orientable 3-manifold and $\phi\colon \pi_1 M \twoheadrightarrow \mathbb{Z}$, and let G be a graph with a single vertex v and edge e. There is a map $p\colon M \to G$ such that $p_* = \phi$ and $p^{-1}(\mathring{e}) = int\,\mathcal{N}$, where $\mathcal{N}$ is a closed regular neighborhood, homeomorphic to $S \times I$, of a 2-sided, nonseparating incompressible surface S. Furthermore, p is modeled on $\mathcal{N}$ by projection to I.*

Above, $\mathring{e}$ refers to the interior of e. The surface S of Lemma 1.1 represents the Poincaré–Lefschetz dual of the class in $H^1(M;\mathbb{Z})$ corresponding to ϕ.

Definition. For a two-sided incompressible surface S in a compact, orientable 3-manifold M, fix a transverse orientation on S and a homeomorphism $f\colon S \times [-1,1] \to \mathcal{N}$, where $\mathcal{N}$ is a closed regular neighborhood of S in M, such that $f(S \times \{0\}) = S$ and orientations and transverse orientations are preserved. We say that $X \doteq M - S \times (-1/2, 1/2)$ is *obtained by cutting M along S*. We let $S_\pm = f(S \times \{\pm 1/2\})$ and take $\mathcal{N}_- = f(S \times (-1,0])$ and $\mathcal{N}_+ = f(S \times [0,1))$.

It is not necessarily the case that the surface S supplied by Lemma 1.1 can be taken to be connected. If it is disconnected, the graph G_0 from above Lemma 1.1 has more than one edge. In this case we have:

Proposition 1.2. *Let M be a compact, orientable 3-manifold and $\phi\colon \pi_1 M \twoheadrightarrow \mathbb{Z}$, let G be a graph with a single vertex v and edge e, and let $p\colon M \to G$ be a map satisfying the conclusions of Lemma 1.1. If $\mathcal{N} = p^{-1}(\mathring{e})$ has k components, then $\mathrm{rg}\,\{M_n\} \geq k-1$, where $M_n \to M$ is the cover corresponding to $\phi^{-1}(n\mathbb{Z}) < \pi_1 M$.*

Proof. The key point of the proof is that $p_* = \phi$ factors through a map onto $\pi_1(G_0)$, a free group on k generators, where G_0 is as described above. Then the rank of $\phi^{-1}(n\mathbb{Z}) < \pi_1 M$ is bounded below by the rank of its image in $\pi_1(G_0)$, which is given by the Reidemeister-Schreier formula.

To show ϕ factors through a map onto $\pi_1(G_0)$, let e_0 be an edge of G_0 and let $\mathcal{N}_0 \cong S_0 \times I$ be the component of $\mathcal{N}$ corresponding to e_0. Choosing a base point $x \in X$ for $\pi_1 M$, let α and β be paths in X, each with initial point x, whose terminal points are of the form $(s,0)$ and $(s,1)$, respectively, for some $s \in S$. Let $\gamma \in \pi_1(M,x)$ be the pointed homotopy class of the path that transverses $\alpha \subset X$ and $\{s\} \times I \subset \mathcal{N}_0$ in the forward direction, then β in reverse.

By Lemma 1.1 — in particular, the fact that p is modeled on $\mathcal{N}$ by projection to I — $p_*(\gamma)$ is the generator for $\pi_1(G)$. Since p_* factors through $\pi_1(G_0) \to \pi_1(G)$, the image of γ in $\pi_1(G_0)$ is the generator corresponding to e_0. Since the choice of e_0 was arbitrary, $\pi_1(M)$ maps onto $\pi_1(G_0)$ and the proposition follows. □

On account of Proposition 1.2, we will assume henceforth that $\mathcal{N} = p^{-1}(\mathring{e})$ is connected, where p is the map supplied by Lemma 1.1. In this case G is the underlying graph of a graph of spaces decomposition of M, with vertex space X and edge space S. (We use the perspective on graphs of groups and spaces from [20]; for definitions see p. 155 there. See also [22] and [26]).

Associated to such a decomposition is an action of $\pi_1 M$ on a tree T, without involutions, such that each vertex stabilizer is conjugate into the vertex group $\pi_1(X)$ and each edge stabilizer into $\pi_1 S$. (This is a fundamental result in Bass-Serre theory; see [20, pp. 166–167].) The terminology "k-acylindrical" is justified by Lemma 1.3 below, which relates k-cylindricity to homotopies through M of curves in S. We must first establish some precise notation.

Definition. Let X and Y be topological spaces. A *homotopy with domain* X *and target* Y is a map $H\colon X \times I \to Y$. The *time-t map of* H, $H_t\colon X \to Y$, is defined by $H_t(x) = H(x,t)$. For a map $f\colon X \to Y$, a *homotopy of* f is a homotopy H with $H_0 = f$. A map $g\colon X \to Y$ is *homotopic to* f if there is a homotopy H of f with $H_1 = g$.

Let $H^1, \dots, H^n$ be homotopies with domain X and target Y. A homotopy H with domain X and target Y is the *composition of* $H^1, \dots, H^n$ if there exist numbers $0 = x_0 < x_1 \cdots < x_n = 1$ and monotone increasing linear homeomorphisms $\alpha_i\colon [x_{i-1}, x_i] \to [0,1]$ such that $H(x,t) = H^i(x, \alpha_i(t))$ whenever $t \in [x_{i-1}, x_i]$.

For $Z \subset Y$, we say $f\colon X \to Y$ is *homotopic into* Z if f is homotopic to a map g with $g(X) \subset Z$. If $W \subset X$, a *homotopy of* W is a homotopy of the inclusion map $W \to X$. A map of pairs $f\colon (X,W) \to (Y,Z)$ is *essential* if f is not homotopic through maps $(X,Y) \to (Z,W)$ to a map into W.

The definitions above are standard, although we have borrowed them from [3]. This is the same source for the definitions below, which are perhaps not as standard.

Definition. Let M be a closed 3-manifold and $S \subset M$ a transversely oriented incompressible surface. A *homotopy in* (M,S) *with domain* K is a homotopy H with domain K and target M such that $H(K \times \partial I) \subset S$. Such a homotopy is *basic* if $H^{-1}(S) = K \times \partial I$. For $\epsilon \in \{+,-\}$, we say a basic homotopy *starts* (or *ends*) *on the ϵ-side* if $H(K \times [0,\delta]) \subset \mathcal{N}_\epsilon$ (or, respectively, if $H(K \times [1-\delta, 1]) \subset \mathcal{N}_\epsilon$).

Let X be obtained by cutting M along S. If H is a basic homotopy in (M,S) with domain K, after straightening in a neighborhood of S and reparametrizing, the restriction of H to $H^{-1}(X)$ determines a homotopy H' in $(X, \partial X)$ with domain K. We say H is *essential* if H' is essential as a map of pairs $(K \times I, K \times \partial I) \to (X, \partial X)$.

A homotopy H in (M,S) with domain K is *reduced with length* k if there exist basic essential homotopies $H^1, \dots, H^k$ and $\epsilon_i \in \{+,-\}$ for $1 \le i \le k$ such that H is the composition of $H^1, \dots, H^k$, for each $i < k$ H^i starts on the ϵ_i- and ends on the $-\epsilon_{i+1}$-side, and H^k starts on the ϵ_k-side.

The following lemma draws the connection between k-acylindricity (or its absence) and homotopies of curves.

Lemma 1.3. *Let M be a closed, orientable hyperbolic* 3*-manifold and $\pi_1 M \times T \to T$ the action on a tree T associated to the graph of spaces decomposition of M determined by a connected, two-sided incompressible surface S. If the action is k-cylindrical for some $k > 1$, then there is a reduced homotopy in (M,S) with length k and domain S^1.*

The proof of this lemma is an elementary but somewhat lengthy digression into Bass–Serre theory. As we do not know of it appearing in the literature, we write a proof in an appendix. In Section 3, we will describe an argument that uses the JSJ decomposition to prove the following:

Theorem 1.4 (Cylinders have bounded length)**.** *Let M be a closed, orientable hyperbolic 3-manifold and $S \subset M$ a connected, two-sided incompressible surface of genus g that is not a fiber or semi-fiber. Then the length of any reduced homotopy in (M,S) is bounded above by $k = 14g - 12$.*

The combination of Lemma 1.3 and Theorem 1.4 immediately yield the following bound.

Corollary 1.5. *Let M be a closed, orientable hyperbolic 3-manifold and S a connected, two-sided incompressible surface in M, with genus g, that is not a fiber or semi-fiber. The $\pi_1 M$-action on the tree determined by S is $(14g-12)$-acylindrical.*

Together with Weidmann's theorem, this quickly implies Theorem 0.2.

Proof of Theorem 0.2. Let M be a closed, orientable hyperbolic 3-manifold with an epimorphism $\phi\colon \pi_1 M \twoheadrightarrow \mathbb{Z}$, and let $p\colon M \to G$ be a map satisfying the conclusion of Lemma 1.1, where G is a graph with a single vertex v and edge e. Fixing $x \in \mathring{e}$, let $S = p^{-1}(x)$ and suppose S is not a fiber in a fibration $M \to S^1$.

If S is not connected then by Proposition 1.2, the rank gradient of $\{M_n\}$ is at least 1, where $M_n \to M$ is the cover corresponding to $\phi^{-1}(n\mathbb{Z})$. Otherwise, Corollary 1.5 implies that the action $\pi_1 M \times T \to T$ associated to ϕ is $(14g-12)$-acylindrical, where g is the genus of S. Furthermore, $\pi_1 M_n$ inherits a $(14g-12)$-acylindrical action on T from that of $\pi_1 M$, with quotient a graph G_n covering G n-to-1. (This follows from the fact that S_ϕ has n distinct homeomorphic lifts to the cover M_n, cyclically permuted by deck transformations.) Weidmann's acylindrical accessibility theorem therefore implies

$$\operatorname{rk} M_n \geq \frac{n-1}{2(14g-12)} + 1.$$

We thus find that $\operatorname{rg}\{M_n\} \geq 1/2(14g-12)$. Since S represents $\mathrm{PD}(\phi)$, and the Thurston norm $\|\phi\|$ is simply the minimal genus of such a surface, the theorem is proved. □

Proof of Theorem 0.1. Let M be an orientable hyperbolic 3-manifold of finite volume that is not compact, so that M is homeomorphic to the interior of a compact 3-manifold $\overline{M}$ with boundary consisting of a nonempty disjoint union of incompressible tori. Now suppose $\phi\colon \pi_1 M \twoheadrightarrow \mathbb{Z}$. Our first claim limits the behavior of $\partial\overline{M}$ under ϕ.

Claim. If there is a component T of $\partial\overline{M}$ such that the inclusion-induced homomorphism $\pi_1 T \to \pi_1 M$ maps into $\ker \phi$, then $\operatorname{rg}\{M_n\} \geq 1$.

This claim follows from the "half lives, half dies" lemma for 3-manifolds with boundary, which asserts that the image of the inclusion-induced map

$$H_1(\partial\overline{M}; F) \to H_1(\overline{M}; F)$$

has dimension $n/2$, where F is a fixed coefficient field and $n = \dim\ H_1(\partial M; F)$. If $T \subset \partial M$ is a component satisfying the hypotheses of the claim, then T has n distinct, homeomorphic lifts to the cover $M_n \to M$ corresponding to $\phi^{-1}(n\mathbb{Z})$. Therefore by half lives, half dies, $H_1(M_n;\mathbb{Q})$ has rank at least n (since $H_1(T;\mathbb{Q})$ has rank 2). Since $\pi_1(M_n)$ maps onto $H_1(M_n;\mathbb{Q})$, the claim follows.

Given the claim, we will assume henceforth that for each component T of $\partial\overline{M}$, ϕ maps $\pi_1 T$ onto a finite-index subgroup of $\mathbb{Z}$. Since $\overline{M}$ has hyperbolic interior, it is again a $K(\Gamma, 1)$ for its fundamental group, and applying [10, Lemma 6.5] and arguing as above, we may assume there is a map $p\colon \overline{M} \to G$ satisfying the conclusion of Lemma 1.1. The only change is that we now take the surface S described in Lemma 1.1 to have boundary and be properly embedded, intersecting the boundary components in essential simple closed curves. We further take each component of $\mathcal{N} \cap T$, $T \subset \partial\overline{M}$, to be a closed subset, homeomorphic to $\mu \times I$, of a regular neighborhood in T of a component μ of $S \cap T$.

For each component T of $\partial\overline{M}$, since $\pi_1 T$ is mapped nontrivially by ϕ it follows that $S \cap T$ is a nonempty collection of parallel simple closed curves on T. Let $\mu_1, \ldots, \mu_k$ enumerate the *boundary slopes* of S in M, each a simple closed curve on a component T of $\partial\overline{M}$ parallel to the components of $S \cap T$. We will use the shorthand $M(\mu)$ for $M(\mu_1, \ldots, \mu_k)$, produced by identifying a solid torus V_i with each component T_i of $\partial\overline{M}$ along its boundary in such a way that the corresponding slope μ_i bounds a disk in V_i. The embedding $S \to \overline{M}$ thus extends to an embedding $S(\mu) \to M(\mu)$, where $S(\mu)$ is the closed surface obtained from S by capping off all boundary components.

Since each boundary slope μ_i maps to a single point under p, any element of $\pi_1 M$ representing μ lies in the kernel of $\phi = p_*$. It follows that there is a map $\bar{\phi}\colon \pi_1 M_\mu \to \mathbb{Z}$ such that ϕ equals the composition of $\bar{\phi}$ with the quotient map $\pi_1 M \to \pi_1 M(\mu)$. Thus for $n \in \mathbb{N}$, $\phi^{-1}(n\mathbb{Z})$ maps onto $\bar{\phi}^{-1}(n\mathbb{Z}) < \pi_1 M(\mu)$, so the rank gradient of the family of covers $\{(M(\mu))_n \to M(\mu)\}$ corresponding to $\bar{\phi}^{-1}(n\mathbb{Z})$ bounds the rank gradient of $\{M_n\}$ below.

It is easy to see that p extends over the filling tori V_i to a map $\bar{p}\colon M(\mu) \to G$ such that $\bar{p}_* = \bar{\phi}$. If S is not connected, then neither is $S(\mu)$, and Proposition 1.2 implies that $\{(M(\mu))_n\}$ has rank gradient at least 1. (Recall that the hypotheses of Proposition 1.2 do not require M hyperbolic.) We will thus assume henceforth that S, and hence $S(\mu)$, is connected. If $M(\mu)$ is hyperbolic, then Theorem 0.2 implies that $\{(M(\mu))_n\}$ has positive rank gradient.

In general it may happen that $M(\mu)$ is not hyperbolic; we will deal with this by passing to a regular cover $M' \to M$. By a theorem of Hodgson-Kerckhoff [11], there is a universal constant K such that for any cusped hyperbolic 3-manifold M and every slope μ on $\partial\overline{M}$ with "normalized length" greater than K, the filled manifold $M(\mu)$ is hyperbolic. Normalized length has the property that the normalized length of the kth power of μ is k times the normalized length of μ. It is well known that $\Gamma = \pi_1 M$ is residually finite; thus let $\Gamma' < \Gamma$ be a finite index normal subgroup not containing μ_i^k for any k such that μ_i^k has normalized length less than K, and let $M' \to M$ be the corresponding cover.

For some fixed n, consider the cover of M' corresponding to $\Gamma'_n := \Gamma_n \cap \Gamma'$, where $\Gamma_n = \phi^{-1}(n\mathbb{Z})$. This choice of notation is somewhat misleading, since the index $[\Gamma' : \Gamma'_n]$ may be less than n. However, the quotient Γ'/Γ'_n is cyclic, since by the diamond isomorphism theorem it is isomorphic to $\Gamma'\Gamma_n/\Gamma_n$, and this is a subgroup of $\Gamma/\Gamma_n \cong \mathbb{Z}_n$. Furthermore, since S lifts to M_n we have $\pi_1 S < \Gamma_n$, and therefore $\pi_1 S' = \pi_1 S \cap \Gamma' < \Gamma'_n$. Thus the cover $M'_n \to M'$ corresponding to Γ'_n is a cyclic cover dual to S'. By above, this family of covers has positive rank gradient $(\Gamma', \{\Gamma'_n\})$, and it follows from Lemma 3.1 of [14] that $(\Gamma, \{\Gamma_n\})$ is also positive. □

2. Essential surfaces and essential intersections

This section is explicitly modeled on [3, Section 4], where it is remarked that the notion of essential intersection of subsurfaces "has appeared implicitly in much of the literature on the characteristic submanifold of a Haken manifold," and such a theory is developed, adapted to "large" subsurfaces. Here we extend this theory to allow certain annular components. Most results of [3, Section 4] extend directly to this context using similar proof strategies, although some require important caveats.

We will work in the PL category throughout the next two sections. In particular, a *polyhedron* is a topological space that admits the structure of a simplicial complex. It is well known that the class of such spaces includes surfaces and 3-manifolds.

Definition. Let S be an orientable surface with no 2-sphere components. If K is a polyhedron, we will say a map $f\colon K \to S$ is π_1*-injective* if on each component K_0 the induced map on $\pi_1 K_0$ is injective, and *large* if this map has nonabelian image.

If $A \subset S$ is a subsurface, we will say A is *incompressible* if no component of A is a disk and the inclusion map $A \to S$ is π_1-injective. A *redundant* component A_0 of an incompressible subsurface A is homotopic in S into another component of A. We say $A \subset S$ is *irredundant* if it is incompressible and has no redundant components.

If A is a compact orientable surface, we will refer to the union of the components of A with negative Euler characteristic as the *large part* $A_{\mathcal{L}}$, and to the union of the cores of the remaining annular components as the *small part* $A_{\mathcal{S}}$ of A.

Remark. If A and B are orientable surfaces and $h\colon A \to B$ is a π_1-injective map, then $h(A_{\mathcal{L}}) \subset B_{\mathcal{L}}$.

The kind of argument we will use in this section is illustrated by a sketch proof for the following assertion: if A is an incompressible subsurface of an orientable surface S with no 2-sphere components, each redundant component of A is homeomorphic to an annulus. For suppose A_0 is such a component, homotopic in S into another component A_1, which we may assume lies in $int\, S$ after pushing off the boundary. Choosing a basepoint in A_1, let $\tilde{S} \to int\, S$ be the cover corresponding to $\pi_1 A_1$. The inclusion $A_1 \to S$ lifts to an embedding to a subsurface $\tilde{A}_1 \subset \tilde{S}$ that carries $\pi_1 \tilde{S}$. Therefore each component of $\tilde{S} - int\, \tilde{A}_1$ is homeomorphic to a half-open annulus. Since A_0 is homotopic into A_1, its inclusion map also lifts to an embedding in $\tilde{S}$. This embedding is π_1-injective, since A_0 is, and does not intersect the preimage of A_1. The latter fact implies that its image is contained in a half-open annulus, so A_0 is an orientable surface with cyclic fundamental group; hence an annulus.

The lemma below extends Lemma 4.1 of [3].

Lemma 2.1. *Suppose A and B are irredundant subsurfaces of a compact, orientable surface S with no 2-sphere or torus components, and A is homotopic into B.*

1. *A is isotopic in S to a subsurface of B.*
2. *If B is homeomorphic to an irredundant subsurface of A, then A and B are isotopic subsurfaces of S.*
3. *If B is homotopic into A, then A and B are isotopic subsurfaces of S.*

Proof. We follow the outline of the proof of [3, Lemma 4.1]; as there, assume without loss of generality that S is connected. If S is an annulus, then any irredundant subsurface of S is also an annulus, and the conclusions of the lemma follow quickly. We thus assume below that S has negative Euler characteristic.

We first prove (1). We initially consider only $A_{\mathcal{L}} \cup A_{\mathcal{S}}$; that is, the disjoint union of the large part $A_{\mathcal{L}}$ of A and the 1-submanifold $A_{\mathcal{S}}$ consisting of the cores of the annular components. Isotop this object so that $\partial A_{\mathcal{L}} \cup A_{\mathcal{S}}$ meets ∂B transversely and in the minimal number of points possible. (We will not make an assumption analogous to the second one in the first paragraph of the proof of Lemma 4.1.)

For a component A_0 of $A_{\mathcal{L}}$, let B_0 be the component of B into which it is homotopic. Fixing a base point in B_0, we let $p\colon \tilde{S} \to int\, S$ be the cover corresponding to $\pi_1 B_0$ and let $\tilde{A}_0$ and $\tilde{B}_0$ be components of $p^{-1}(A_0)$ and $p^{-1}(B_0)$, respectively, projecting homeomorphically under p. Since B_0 is incompressible, the inclusion $\tilde{B}_0 \to \tilde{S}$ induces an isomorphism at the level of π_1, and so each component of $X = \tilde{S} - int\, \tilde{B}_0$ is homeomorphic to a half-open annulus.

If A_0 is large and $\partial\tilde{A}_0$ meets $\partial\tilde{B}_0$, or if A_0 is a simple closed curve and $\tilde{A}_0$ meets $\partial\tilde{B}_0$, then the argument of the paragraph that spans pp. 2405–2406 in [3] yields a contradiction to our minimality assumption. Since this does not occur, every large component A_0 has boundary disjoint from ∂B_0, and every small component A_0 is disjoint from ∂B_0, where B_0 is the component into which A_0 is homotopic.

There is a further isotopy of $A_{\mathcal{L}} \cup A_{\mathcal{S}}$, which is constant on $A_{\mathcal{S}}$ and has image entirely contained in $A_{\mathcal{L}} \cup A_{\mathcal{S}}$, after which $A_{\mathcal{L}} \subset B_{\mathcal{L}}$. This is accomplished by pushing across annuli in A as described in the second full paragraph on p. 2406 of [3]. Such an isotopy does not change the cardinality of $(\partial A_{\mathcal{L}} \cup A_{\mathcal{S}}) \cap \partial B$.

Now suppose $\mathfrak{a}$ is a component of $A_{\mathcal{S}}$, and let B_0 be the component of B into which $\mathfrak{a}$ is homotopic. Then either $\mathfrak{a}$ is contained in $int\, B_0$ or disjoint from B_0. In the latter case, since $\mathfrak{a}$ is a simple closed curve homotopic into B_0 it bounds an annulus C in S with a component $\mathfrak{b}$ of ∂B_0. Since A is irredundant, $int\, C$ does not contain any component of $A_{\mathcal{L}} \cup A_{\mathcal{S}}$; thus if they are not disjoint there is an arc contained in $\partial A_{\mathcal{L}}$ or $A_{\mathcal{S}}$ that bounds a disk in C with a subarc of $\mathfrak{b}$. Pushing across an innermost such disk reduces the number of intersections between $\partial A_{\mathcal{L}} \cup A_{\mathcal{S}}$ and ∂B, a contradiction. Thus $int\, C$ is disjoint from $A_{\mathcal{L}} \cup A_{\mathcal{S}}$, and $\mathfrak{a}$ can be pushed across C into $int\, B_0$. After a sequence of such isotopies we have $A_{\mathcal{L}} \cup A_{\mathcal{S}} \subset B$.

In order to establish (1), fix a hyperbolic metric with convex boundary on S, and choose $\epsilon > 0$ so that for each component $\mathfrak{a}$ of $A_{\mathcal{S}}$, the following hold:

(1) The ϵ neighborhood $\mathcal{N}_\epsilon(\mathfrak{a})$ is regular and contained in the component A_0 of A containing $\mathfrak{a}$.
(2) Throughout the isotopy described above, $\mathcal{N}_\epsilon(\mathfrak{a})$ remains regular, and $\mathfrak{a}$ has distance at least 2ϵ from every other component of $A_{\mathcal{L}} \cup A_{\mathcal{S}}$.
(3) After the isotopy described above, $\mathcal{N}_\epsilon(\mathfrak{a}) \subset B$.

By the first criterion above A deformation retracts to the union of $A_{\mathcal{L}}$ with $\bigcup_{\mathfrak{a}} \mathcal{N}_\epsilon(\mathfrak{a})$ over the components $\mathfrak{a}$ of $A_{\mathcal{S}}$. By the second criterion, the isotopy of $A_{\mathcal{L}} \cup A_{\mathcal{S}}$ extends to this union, and by the third, it takes it into B. This establishes (1).

We now turn to the proof of (2). Using (1), we will assume that $A \subset int\, B$. In particular, $A_{\mathcal{L}} \subset int\, B_{\mathcal{L}}$. Since π_1-injective maps preserve large parts $B_{\mathcal{L}}$ is homeomorphic to a large subsurface of $A_{\mathcal{L}}$. The last 3 paragraphs on [3, p. 2406] thus imply that each component of $\overline{B_{\mathcal{L}} - A_{\mathcal{L}}}$ is an annulus with exactly one boundary component in $A_{\mathcal{L}}$. In particular, we note that $\chi(B_{\mathcal{L}}) = \chi(A_{\mathcal{L}})$, where $\chi(S)$ refers to the Euler characteristic of S.

Since A is irredundant it follows that each annular component of A is contained in an annular component of B, and that no two are contained in the same component. Therefore $B_{\mathcal{S}}$ has at least as many components as $A_{\mathcal{S}}$. If $B_{\mathcal{S}}$ had more components than $A_{\mathcal{S}}$, then the homeomorphic embedding $B \to A$ would either take two annular components into the same annular component of A, contradicting irredundancy of the image, or would take an annular component of B into a component of $A_{\mathcal{L}}$. But since the image of $B_{\mathcal{L}}$ is a large subsurface of $A_{\mathcal{L}}$ with the same Euler characteristic, each component of its complement is an annulus, and the latter possibility above again contradicts irredundancy of the image of B.

We thus find that each annular component of B contains a unique component of A as an incompressible sub-annulus. Together with the assertions above regarding $A_{\mathcal{L}} \subset B_{\mathcal{L}}$, this implies (2).

To establish (3), we note that if B is homotopic into A, then by (1) it is isotopic to a subsurface of A. This subsurface is necessarily irredundant, since B is, hence the desired conclusion follows from (2). □

The following proposition extends [3, Proposition 4.2]. Below we reference the "large intersection" $A \wedge_{\mathcal{L}} B$ of large surfaces A and B from [3, Definition 4.3].

Proposition 2.2. *Suppose A and B are irredundant subsurfaces of an orientable compact surface S with no 2-sphere or torus components. Then up to non-ambient isotopy there is a unique irredundant subsurface C of S with the following property:*

(∗) *$C_{\mathcal{L}} = A_{\mathcal{L}} \wedge_{\mathcal{L}} B_{\mathcal{L}}$, and a π_1-injective map $f\colon K \to S$ is homotopic into each of A and B if and only if f is homotopic into C.*

Furthermore, there are subsurfaces $A_0 \subset S$ and $B_0 \subset S$, isotopic to A and B, respectively, such that ∂A_0 meets ∂B_0 transversely and a union C of components of $A_0 \cap B_0$ satisfies (∗) above.

Before we prove the proposition, we record a helpful observation.

Lemma 2.3. *Let S be a compact, orientable surface with no torus or 2-sphere components and K a connected polyhedron. If $f\colon K \to S$ is a π_1-injective map that is not large, then there is a curve $\mathfrak{c} \subset K$ such that f is homotopic in S to $f|_{\mathfrak{c}}$.*

Proof. Since K is connected, we may assume S is connected. Since S is not the 2-sphere, the universal cover of S is contractible. Therefore by a standard argument, if $\pi_1 K$ is trivial, f is homotopic to a constant map. Since the restriction of f to any curve in K has this property as well, the lemma holds in this case. We thus assume $\pi_1 K$ is nontrivial. After pushing off the boundary, we also assume f maps into *int* S.

If S has negative Euler characteristc, it is well known that abelian subgroups of $\pi_1 S$ are cyclic. (This follows for instance from the fact that S admits a hyperbolic metric with geodesic boundary.) Therefore there is a curve $\mathfrak{c} \subset K$ such that the map on fundamental group induced by $f|_{\mathfrak{c}}$ carries the image in $\pi_1 S$ of $\pi_1 K$. Let $\tilde{S} \to$ *int* S be the cover determined by $f_*(\pi_1 K)$. Since $\tilde{S}$ has cyclic fundamental group, it is homeomorphic to an open annulus.

Let $\tilde{f}$ be a lift of f to $\tilde{S}$. Since $\tilde{f}|_{\mathfrak{c}}$ induces an onto map at the level of π_1, it is homotopic to the core curve of $\tilde{S}$. Since $\tilde{S}$ deformation retracts to its core, there is a homotopy of $\tilde{f}$ to $\tilde{f}|_{\mathfrak{c}}$; hence, after projecting to S, a homotopy of f to $f|_{\mathfrak{c}}$.

If S is an annulus, the argument proceeds as in the paragraph above. □

We will frequently take Lemma 2.3 for granted below, and if a map is not large only consider its restriction to some curve.

Proof of Proposition 2.2. We assume without loss of generality that $A, B \subset int\, S$. If C and C' are surfaces with property $(*)$, then C is homotopic into C' and vice-versa. Hence Lemma 2.1(2) implies that they are isotopic, establishing uniqueness.

Now let B_0 be a representative of the isotopy class of B in S with the property that ∂B_0 meets ∂A transversely in the smallest possible number of points, and let C_0 be the union of the components of $A \cap B_0$ that are large. (In the language of [3], $C = \mathcal{L}(A \cap B_0)$.) The proof of [3, Proposition 4.2] implies that every large map $f\colon K \to S$ homotopic into A and B is homotopic into C_0. We will construct C by adding annular components to C_0.

Suppose $f\colon \mathfrak{c} \to S$ is a homotopically nontrivial closed curve homotopic into A and B but not into C_0. Let A_1 be a component of A such that f is homotopic into A_1, let $p\colon \tilde{S} \to int\, S$ be the covering space corresponding to $\pi_1 A_1$, and let $\tilde{A} \subset \tilde{S}$ be a component of $p^{-1}(A)$ mapping homeomorphically under p. Since A_1 is π_1-injective in S, the inclusion-induced homomorphism $\tilde{A} \to \tilde{S}$ is an isomorphism and hence every component of $X = \tilde{S} - int\, \tilde{A}$ is a half-open annulus.

Note that since f is homotopic into B, it is homotopic into B_0. Since f is homotopic into A_1, it admits a lift $\tilde{f}$ to $\tilde{S}$; furthermore, the homotopy into B_0 lifts to a homotopy of $\tilde{f}$ to a map g with image in $p^{-1}(B_0)$. Let $\tilde{B}_0$ be the component of $p^{-1}(B_0)$ containing $g(\mathfrak{c})$.

Unlike in the proof of [3, Proposition 4.1], it is not necessarily the case that $\tilde{B}_0$ intersects $\tilde{A}_1$, but suppose for now that it does. Then the argument that begins in the paragraph of [3] spanning pp. 2407–2408 establishes that $\tilde{B}_0$ deforms in $\tilde{S}$ into $\tilde{B}_0 \cap \tilde{A}_1$. After projecting to S it follows that f is homotopic into a component of $p(\tilde{B}_0 \cap \tilde{A}_1) \subset B_0 \cap A_1$. Since f is π_1-injective, this component is not a disk, and since f is not homotopic into C_0, it is an annulus C_1 which furthermore is not parallel to any component of C_0.

Suppose now that $\tilde{B}_0 \cap \tilde{A}_1 = \emptyset$, and let Z be the component of X containing $\tilde{B}_0$. Since $\tilde{B}_0$ contains the homotopically nontrivial curve $g(\mathfrak{c})$, it has a boundary component $\mathfrak{b}_0$ that bounds an annulus $Z_0 \subset Z$ together with $\mathfrak{a}_0 = \partial Z$. If any component of the frontier in $\tilde{S}$ of $p^{-1}(B_0)$ intersected $\mathfrak{a}_0$, there would thus be a disk in Z_0 with boundary $\alpha \cup \beta$, where $\alpha \subset \mathfrak{a}_0$ and $\beta \subset \partial(p^{-1}(B_0))$. If this did occur then B_0 could be isotoped to reduce the number of intersections with A, by the argument of the paragraph of [3] spanning pp. 2405–2406. Thus $\mathfrak{a}_0 \cap p^{-1}(B_0) = \emptyset$.

Since p projects A_1 homeomorphically, it sends $\mathfrak{a}_0$ homeomorphically to a component of ∂A_1 in S. Since $\tilde{B}_0$ is a component of $p^{-1}(B_0)$, p restricts on $\mathfrak{b}_0$ to a k-to-1 covering map to a component $\mathfrak{b}$ of ∂B_0 for some $k \geq 1$. By the paragraph above, $\mathfrak{b}$ does not intersect $\mathfrak{a}$. Furthermore, the annulus Z_0 bounded by $\mathfrak{a}_0$ and $\mathfrak{b}_0$ projects under p to a free homotopy in S between $\mathfrak{a}$ and the kth power of $\mathfrak{b}$. Since S is an orientable surface, by [5, Lemma 2.4] $k = 1$ and $\mathfrak{a}$ and $\mathfrak{b}$ bound an annulus Z_1 in S.

Now B_0 may be further isotoped across Z_1 to overlap A in an annulus $C_1 \subset A_1$ containing $\mathfrak{a}$. There are at most a finite number of such isotopies, since A has only a finite number of boundary components, after which the discussion above shows that every π_1-injective map of a curve $\mathfrak{c}$ is homotopic into a subsurface C consisting of the union of C_0 with finitely many annular components of $A \cap B_0$. Property $(*)$ for C now follows from Lemma 2.3, and C satisfies the final conclusion of the proposition by construction. □

Definition. If A and B are irredundant subsurfaces of an orientable compact surface S, we say an irredundant surface C that satisfies condition $(*)$ of Proposition 2.2 *represents the essential intersection* $A \cap_{ess} B$ *of* A *and* B.

Proposition 2.2 implies in particular that each of A and B contains a subsurface that represents $A \cap_{ess} B$, and that these subsurfaces are isotopic in S.

The result below extends [3, Proposition 4.4], but note the added complication. We will use it in the proof of Proposition 3.1. Below, for a subset S of a topological space X, we refer to the *frontier* of S in X as $\mathrm{fr}\, S \doteq \overline{S} \cap \overline{X - S}$.

Proposition 2.4. *Let B be an irredundant subsurface of a compact, orientable surface S with no 2-sphere components and $f\colon K \to B$ a π_1-injective map. If $g\colon K \to B$ is homotopic to f in S, then:*

1. *Either f and g are homotopic in B; or*
2. *There are distinct components $\mathfrak{a}$ and $\mathfrak{b}$ of the frontier of B in S, and a component K_0 of K, such that $f|_{K_0}$ is homotopic into $\mathfrak{a}$, and $g|_{K_0}$ into $\mathfrak{b}$, in B.*

In the second case above, $\mathfrak{a}$ and $\mathfrak{b}$ are in particular parallel in S.

We will use the lemma below in the proof of Proposition 2.4, and also in the following section.

Lemma 2.5. *Let B be a compact subsurface of a surface S, and suppose $\mathfrak{c} \subset B$ is a closed curve homotopic into $S - B$. Then $\mathfrak{c}$ is homotopic in B into* $\mathrm{fr}\, B$.

Proof. After pushing off of boundaries, we will assume that $B \subset int\, S$ and $\mathfrak{c} \subset int\, B$. Let $H\colon S^1 \times I \to S$ be a homotopy with $H_0 = \mathfrak{c}$ and $H_1(S^1) \subset S - B$. After a small perturbation relative to $S^1 \times \partial I$, we may assume H is transverse to $\mathrm{fr}\, B$. Then $H^{-1}(\mathrm{fr}\, B) \subset S^1 \times I$ is a disjoint union of simple closed curves separating $S^1 \times \{0\}$ from $S^1 \times \{1\}$. Any such a subset contains a curve $\mathfrak{a}$ ambiently isotopic to $S^1 \times \{1/2\}$, since a disjoint union of contractible curves does not separate. Restricting H to $\overline{U}$, where U is the component of $S^1 \times I - \mathfrak{a}$ containing $S^1 \times \{0\}$, yields the desired homotopy. □

Proof of Proposition 2.4. Assume $B \subset int\, S$. We may also assume that K is connected, since the homotopy may be adjusted component-by-component. Thus let B_0 be the component of B containing $f(K)$. If K is simply connected, then since S is not a 2-sphere and B is incompressible, a standard argument implies that f and g are both homotopic in B to constant maps, so the result follows from path-connectedness of B_0. We therefore assume K is not simply connected.

Choosing a base point in B_0, we let $p\colon \tilde{S} \to int\, S$ be the cover corresponding to $\pi_1(B_0) < \pi_1(S)$. By construction, the inclusion map $B_0 \to S$ lifts to an embedding to $\tilde{S}$ with image a subsurface we denote $\tilde{B}_0$, that carries the fundamental group of $\tilde{S}$. Since B_0 is π_1-injective in S, each component of $\tilde{S} - int\, \tilde{B}_0$ is homeomorphic to a half-open annulus. In particular, there is a deformation retraction $r\colon \tilde{S} \to \tilde{B}_0$.

Since f maps K into B_0, it has a lift $\tilde{f}\colon K \to \tilde{S}$ with $\tilde{f}(A) \subset \tilde{B}_0$; furthermore, the homotopy from f to g lifts to a homotopy H from $\tilde{f}$ to a lift $\tilde{g}$ of g with image in $p^{-1}(B)$. If $\tilde{g}$ has image in $\widetilde{B}_0$, then $H_1 = p \circ r \circ H$ is a homotopy between f and g with image in B_0.

Suppose $\tilde{g}$ does not have image in $\tilde{B}_0$, and let Z be the component of $\tilde{S} - int\, \tilde{B}_0$, a half-open annulus, containing $\tilde{g}(K)$. Since f, hence also g, is π_1-injective and

$\pi_1 K$ is non-trivial, we find that the component $\tilde{B}_1$ of $p^{-1}(B)$ containing $\tilde{g}(A)$ has a boundary component $\tilde{\mathfrak{b}}_1$ parallel to $\tilde{\mathfrak{b}}_0 = \partial Z$. Furthermore, $\pi_1 K = \mathbb{Z}$, so by Lemma 2.3, there is a curve $\mathfrak{c} \subset K$ such that $\tilde{f}$ is homotopic in $\tilde{B}_0$ to $\tilde{f}|_{\mathfrak{c}}$ and $\tilde{g}$ is homotopic in $\tilde{B}_1$ to $\tilde{g}|_{\mathfrak{c}}$. Since $\tilde{H}$ takes $\tilde{g}|_{\mathfrak{c}}$ out of $\tilde{B}_1$, Lemma 2.5 implies that $\tilde{g}|_{\mathfrak{c}}$ is homotopic in $\tilde{B}_1$ into $\tilde{b}_1$. Similarly, $\tilde{f}|_{\mathfrak{c}}$ is homotopic in $\tilde{B}_0$ into $\tilde{b}_0$.

By our hypotheses p projects $\tilde{\mathfrak{b}}_0$ homeomorphically to a component $\mathfrak{b}_0$ of $\operatorname{fr} B_0$, and it restricts on $\tilde{\mathfrak{b}}_1$ to a k-to-1 covering to another component $\mathfrak{b}_1$. Since the annulus in Z bounded by $\tilde{\mathfrak{b}}_0$ and $\tilde{\mathfrak{b}}_1$ projects to a free homotopy in S between $\mathfrak{b}_0$ and the kth power of $\mathfrak{b}_1$, [5, Lemma 2.4] implies that $k = 1$ and $\mathfrak{b}_0$ and $\mathfrak{b}_1$ bound an annulus in S. The conclusion now follows from the paragraph above. □

3. Cylinders have bounded length

This section is dedicated to proving Theorem 1.4. The proof is a "vegematic argument" — a well known strategy that uses properties of the characteristic submanifold of the manifold X obtained by cutting M along S. The first step of this argument identifies a sequence $\Psi_1 \supset \Psi_2 \supset \cdots$ of subsurfaces of S with minimal complexity such that the following holds: for each k, if H is a reduced homotopy in (M, S) with length k, then H_0 is homotopic into Ψ_k in S. The second step uses the fact that M is hyperbolic and S is not a fiber or semi-fiber to show that Ψ_k is not homotopic into Ψ_{k+2} in S as long as $\Psi_k \neq \emptyset$. Therefore eventually $\Psi_k = \emptyset$, and homotopies expire in finite time.

Previous versions of this argument have appeared in [3, §5], [4, §4], and [16]. For various reasons, none of these proofs requires accounting for solid torus components of the characteristic submanifold. However in our case many examples have the property that the longest homotopies must pass through solid tori; thus our main task is to extend the vegematic argument to accomodate such components. Much of our proof relies on and directly extends work in [3]. We have indicated when this so and attempted to give appropriate cross-references.

If M is a closed, orientable hyperbolic 3-manifold containing an incompressible surface S, the manifold X obtained by cutting M along S is *simple* in the following precise sense: X is compact, connected, orientable, irreducible, and ∂X is incompressible; no subgroup of $\pi_1 X$ is isomorphic to $\mathbb{Z} \times \mathbb{Z}$; and X is not a closed manifold with finite fundamental group.

The JSJ decomposition of a simple 3-manifold with boundary describes a "best" way to cut it apart along *essential* annuli; that is, along annuli that are properly embedded, π_1-injective, and not boundary-parallel [12], [13]. Below, if P is an I-bundle over a surface F, we let $\partial_h P$ denote the associated ∂I-bundle, the *horizontal* boundary of P, and denote by $\partial_v P$ (the *vertical* boundary) the I-bundle over ∂F. A submanifold of a Seifert fibered space is *saturated* if it is a union of fibers.

Theorem (Jaco–Shalen, Johansson). *Let X be a simple 3-manifold with nonempty boundary. Up to ambient isotopy, its characteristic submanifold Ω is the unique compact submanifold of X with the following properties.*

(1) *Every component of Ω is either an I-bundle P over a surface such that $P \cap \partial X = \partial_h P$, or a Seifert fibered space S such that $S \cap \partial X$ is a saturated 2-manifold in ∂S.*
(2) *Every component of the frontier of Ω is an essential annulus or torus in X.*

(3) *No component of* Ω *is ambiently isotopic in* X *to a submanifold of another component of* Ω.
(4) *If* Ω_1 *is a compact submanifold of* X *such that (1) and (2) hold with* Ω_1 *in place of* Ω*, then* Ω_1 *is ambiently isotopic in* X *to a submanifold of* Ω.

If K *is a polyhedron and* $H\colon (K \times I, K \times \partial I) \to (X, \partial X)$ *is an essential, nondegenerate map, then* H *is homotopic into* $(\Omega, \Omega \cap \partial X)$.

If X is a simple 3-manifold with no torus boundary components, then each Seifert fibered component of its characteristic submanifold Ω is homeomorphic to a solid torus. We will refer to $\Omega \cap \partial X$ as the *characteristic set*. A consequence of the JSJ theorem is that the characteristic set carries a homotopic image of the time-0 map of any homotopy H with length 1. The first main result of this section extends Proposition 5.2.8 of [3], identifying a sequence of subsurfaces that play a role analogous to the characteristic set for homotopies with length $k \geq 1$.

Proposition 3.1. *Let* M *be a closed, orientable hyperbolic* 3*-manifold and* $S \subset M$ *a two-sided incompressible surface such that* $M \setminus\setminus S$ *has two components* $X^{\pm 1}$*. For each* $\epsilon \in \{\pm 1\}$ *there is a sequence of essential (possibly empty) subsurfaces* $(\Psi_k^\epsilon)_{k \in \mathbb{N}}$ *of* S*, such that* $\Psi_1^\epsilon \subset \Omega^\epsilon \cap \partial X^\epsilon$*, where* Ω^ϵ *is the characteristic submanifold of* X^ϵ*, and for each* $k \in \mathbb{N}$ *we have:*
(1) $\Psi_k^\epsilon \supset \Psi_{k+1}^\epsilon$*; and*
(2) *A* π_1*-injective map* $f\colon K \to S$ *is homotopic in* S *to a map with image in* Ψ_k^ϵ *if and only if there exists a reduced homotopy* $H\colon K \times I \to M$ *of length* k*, starting on the* ϵ *side, with* $H_0 = f$*; and*
(3) $(\Psi_k^\epsilon)_{\mathcal{L}} = \Phi_k^\epsilon$*, where* Φ_k^ϵ *is the surface identified in* [3, Proposition 5.2.8].

A surface with the properties above is determined up to isotopy in S *by the requirement that it be irredundant.*

We will assume the results of [3] throughout this section, and briefly review below some of its relevant terminology and references, taking M, S, and $X^{\pm 1}$ to be as in Proposition 3.1 above. The results of [3] that we will use were proven there with the hypothesis that M is a simple *knot manifold*; that is, with a single torus boundary component. However, they depend only on the results on large intersection developed in [3, §4] and basic facts about I-bundles, and so carry over to our context without alteration.

Section 5.2.1 of [3] defines $(\Sigma^\epsilon, \Phi^\epsilon)$ to be the $(I, \partial I)$-bundle pair that is the union of all I-bundle components of the characteristic submanifold of X^ϵ. Proposition 5.2.8 of [3] identifies a sequence $\Phi_1^\epsilon \supset \Phi_2^\epsilon \supset \ldots$ of subsurfaces of Φ^ϵ, with $\Phi_1^\epsilon = (\Phi^\epsilon)_{\mathcal{L}}$, determined by the analog of Property (2) from Proposition 3.1, where "π_1-injective" has been replaced by "large". Proposition 5.3.1 of [3] supplies a further tool for describing homotopies, defining homeomorphisms $h_k^\epsilon\colon \Phi_k^\epsilon \to \Phi_k^{(-1)^{k+1}\epsilon}$, such that if $H\colon K \times I \to M$ is a reduced homotopy of length k starting on the ϵ-side with large time-0 map, then there exists $f\colon K \to \Phi_k^\epsilon$ such that H_0 is homotopic to f and H_1 to $h_k^\epsilon \circ f$.

We will find the definition and some additional properties of the maps h_k^ϵ useful below. There is a fixed-point free involution τ_ϵ of Φ^ϵ that exchanges the endpoints of I-fibers. Then h_1^ϵ is defined to be the restriction of τ_ϵ to Φ_1^ϵ, and for $k > 1$, h_k^ϵ is defined recursively by composing $\tau_{\pm\epsilon}$ with a homotope of the restriction of h_{k-1}^ϵ. In particular, [3, Proposition 5.3.4]) implies that for $k > 1$ there is an embedding

$g^{\epsilon}_{k-1}\colon \Phi^{\epsilon}_k \to \Phi^{(-1)^{k-1}\epsilon}_1$ such that h^{ϵ}_k is homotopic in S to $\tau_{(-1)^{k-1}\epsilon} \circ g^{\epsilon}_{k-1}$, and [3, Proposition 5.3.5] implies that $h^{\epsilon}_{k-1}(\Phi^{\epsilon}_k)$ is isotopic in S to the large part of $\Phi^{(-1)^k\epsilon}_{k-1} \cap_{ess} \Phi^{(-1)^{k-1}\epsilon}_1$.

The standard lemma below is a weak version of "essential annuli in I-bundles are homotopic to vertical." The proof of [3, Lemma 5.2.4] implies it without revision, although the lemma is only stated there for homotopies with large time-0 maps.

Lemma 3.2. *For $\epsilon \in \{\pm 1\}$, let $H\colon (S^1 \times I, S^1 \times \partial I) \to (P, \partial_h P)$ be an essential basic homotopy, where P is an I-bundle component of Ω^{ϵ}. Then H_1 is homotopic in P to $\tau_{\epsilon} \circ H_0$.*

The lemma below extends the conclusion of [3, Proposition 5.3.1] to apply to reduced homotopies with time-0 maps that are not large.

Lemma 3.3. *For $k \in \mathbb{N}$ and $\epsilon \in \{\pm 1\}$, let $H\colon S^1 \times I \to M$ be a reduced homotopy of length k that starts on the ϵ side, such that H_0 is homotopic in S to a curve $\mathfrak{c} \subset \Phi^{\epsilon}_k$ that is not homotopic into $\partial\Phi^{\epsilon}_k$. Then H_1 is homotopic in S to $h^{\epsilon}_k \circ \mathfrak{c}$.*

Proof. We prove this first for $k = 1$; thus assume that $H\colon (S^1 \times I, S^1 \times \partial I) \to (X^{\epsilon}, S^{\epsilon})$ is an essential basic homotopy. Let P be the component of Ω^{ϵ} containing $\mathfrak{c}$, and note that $P \cap S^{\epsilon}$ contains the component of Φ^{ϵ}_k containing $\mathfrak{c}$. The JSJ theorem implies that H is homotopic as a map of pairs to $H'\colon (S^1 \times I, S^1 \times \partial I) \to (\Omega^{\epsilon}, \Omega^{\epsilon} \cap S^{\epsilon})$. If the image of H' were not contained in P, then $\mathfrak{c}$ would be homotopic in S^{ϵ}) out of $P \cap S^{\epsilon}$. But by Lemma 2.5, this would contradict our assumption that $\mathfrak{c}$ is not homotopic into $\partial\Phi^{\epsilon}_k$. Therefore H' maps into P.

Since $\mathfrak{c}$ is homotopic to H_0, it is homotopic in S to H'_0, and since it is not homotopic into ∂P, Proposition 2.4 implies it is homotopic in P to H'_0. Then $\tau_{\epsilon} \circ \mathfrak{c}$ is homotopic in P to $\tau_{\epsilon} \circ H'_0$. By Lemma 3.2, H'_1 is homotopic in P to $\tau_{\epsilon} \circ H'_0$, and the lemma follows for $k = 1$.

For $k > 1$, we write H as the composition of essential basic homotopies $H^1, \ldots, H^k$, and assume that the composition of $H^1, \ldots, H^{k-1}$ has time-1 map homotopic to $h^{\epsilon}_{k-1} \circ \mathfrak{c}$. Let $g^{\epsilon}_{k-1}\colon \Phi^{\epsilon}_k \to \Phi^{(-1)^{k-1}\epsilon}_1$ be the embedding supplied by [3, Proposition 5.3.4], so that h^{ϵ}_{k-1} is homotopic to g^{ϵ}_{k-1} and h^{ϵ}_k is homotopic to $\tau_{(-1)^{k-1}\epsilon} \circ g^{\epsilon}_{k-1}$. Let P be the I-bundle component of $\Omega^{(-1)^{k-1}\epsilon}$ such that $g^{\epsilon}_{k-1} \circ \mathfrak{c}$ maps into $\partial_h P$. Since $\mathfrak{c}$ is not homotopic into $\partial\Phi^{\epsilon}_k$, the same holds true for $g^{\epsilon}_{k-1} \circ \mathfrak{c}$ in $\partial_h P$.

Since $H^k_0 = H^{k-1}_1$ it is homotopic in S to $g^{\epsilon}_{k-1} \circ \mathfrak{c}$. It thus follows from the JSJ theorem as in the $k = 1$ case that H^k is homotopic as a map of $(I, \partial I)$-bundle pairs into P, and furthermore by Lemma 3.2 that H^k_1 is homotopic to $\tau_{(-1)^{k-1}\epsilon} \circ g^{\epsilon}_{k-1} \circ \mathfrak{c}$. Therefore $H^k_1 = H_1$ is homotopic to $h_k \circ \mathfrak{c}$, and the lemma follows by induction. □

Because solid torus components of Ω may have many components of intersection with ∂X, no homeomorphism analogous to h^{ϵ}_k is uniquely defined on Ψ^{ϵ}_k. We do have the much weaker fact below, however.

Lemma 3.4. *For $k \in \mathbb{N}$ and $\epsilon \in \{\pm 1\}$, let $H\colon S^1 \times I \to M$ be a reduced homotopy of length k that starts on the ϵ side. If H_0 is homotopic in S to $\mathfrak{c}^n$ for some curve $\mathfrak{c} \subset (\Omega \cap \partial X^{\epsilon})$ and $n \in \mathbb{Z} - \{0\}$, then there is a reduced homotopy $K\colon S^1 \times I \to M$ of length k that starts on the ϵ side, such that $K_0 = \mathfrak{c}$ and H_1 is homotopic to $(K_1)^n$. (Here by $\mathfrak{c}^n$ we refer to the composition of $\mathfrak{c}$ with the n-fold cover $S^1 \to S^1$.)*

Proof. For $\epsilon \in \{\pm 1\}$, let $H\colon (S^1 \times I, S^1 \times \partial I) \to (X^\epsilon, S^\epsilon)$ be an essential basic homotopy. By the JSJ theorem, H is homotopic through maps $(S^1 \times I, S^1 \times \partial I) \to (X^\epsilon, S^\epsilon)$ into some component P of the characteristic submanifold Ω^ϵ. We thus assume H maps into Ω^ϵ.

If P is an I-bundle component of Ω^ϵ, Lemma 3.2 implies that H_1 is homotopic to $\tau_\epsilon \circ H_0$. Thus if H_0 is homotopic to $\mathfrak{c}^n$, then H_1 is homotopic to $\tau_\epsilon \circ \mathfrak{c}^n = (\tau_\epsilon \circ \mathfrak{c})^n$. If P is a solid torus component of Ω^ϵ and H_0 is homotopic to $\mathfrak{c}^n$, let $\mathfrak{b}$ be a curve with the same degree as $\mathfrak{c}$ in the component of $V \cap S^\epsilon$ containing the image of H_1. Then by well-definedness of degree, $\mathfrak{c}$ is homotopic to $\mathfrak{b}$ in P and H_1 is homotopic to $\mathfrak{b}^n$ in $P \cap S^\epsilon$.

The lemma now follows from an induction argument. □

Proof of Proposition 3.1. We will prove the proposition by induction. Let $\Psi_1^{\pm 1}$ be obtained from $\Omega^{\pm 1} \cap \partial X^{\pm 1}$ by discarding redundant annuli, where $\Omega^{\pm 1}$ is the characteristic submanifold of $X^{\pm 1}$. Property (2) for $\Psi_1^{\pm 1}$ holds by the enclosing property of the JSJ theorem, and we note that $(\Psi_1^{\pm 1})_{\mathcal{L}} = \Phi_1^{\pm 1}$.

Now let $m \geq 2$ be given, and suppose that for each $\epsilon \in \{\pm 1\}$ we have identified a sequence of subsurfaces

$$\Psi_1^\epsilon \supset \Psi_2^\epsilon \ldots \supset \Psi_{m-1}^\epsilon,$$

such that for each $k < m$, Ψ_k satisfies (2) and $(\Psi_k^\epsilon)_{\mathcal{L}} = \Phi_k^\epsilon$. We will further assume (after discarding some annuli if necessary) that Ψ_k^ϵ is irredundant for $k < m$.

Before we define Ψ_m^ϵ, we let P_m^ϵ be a subsurface of $\Phi_{m-1}^{\epsilon\cdot(-1)^m}$ representing

$$\Phi_{m-1}^{\epsilon(-1)^m} \cap_{ess} \Psi_1^{\epsilon(-1)^{m+1}}.$$

By Proposition 2.2, $(P_m^\epsilon)_{\mathcal{L}}$ is maximal among large surfaces of $\Phi_{m-1}^{\epsilon(-1)^m}$ that admit a homotopy of length one starting on the $\epsilon(-1)^{m+1}$-side. If a large subsurface A of $\Phi_{m-1}^{\epsilon(-1)^m}$ admits an essential homotopy of length one starting on the $\epsilon(-1)^{m+1}$-side, then $(h_{m-1}^\epsilon)^{-1}(A)$ admits a homotopy of length m starting on the ϵ-side; thus [3, Proposition 5.2.8] implies that $h_{m-1}^\epsilon(\Phi_m^\epsilon)$ has the same maximality property as $(P_m^\epsilon)_{\mathcal{L}}$. Therefore by Proposition 2.1(3), these are isotopic subsurfaces of $\Phi_{m-1}^{\epsilon(-1)^m}$.

We now define $\Psi_m^\epsilon = \Phi_m^\epsilon \cup (\bigcup A_i) \cup (\bigcup B_j) \cup (\bigcup C_k)$, where the A_i, B_j, and C_k are annuli defined as follows:

1. Let $\{A_i\}$ be the set of annular components of Ψ_{m-1}^ϵ that admit a reduced homotopy of length m.
2. Let $\{\mathfrak{b}_j\}$ be the set of components of the frontier in S of Φ_{m-1}^ϵ such that $\mathfrak{b}_j$ is not isotopic into Φ_m^ϵ but $\mathfrak{b}_j$ admits a reduced homotopy of length m, and for each j let B_j be a regular neighborhood of $\mathfrak{b}_j$ in $\Phi_{m-1}^\epsilon - int\, \Phi_m^\epsilon$.
3. Let $\{C_k'\}$ be the set of annular components of P_m^ϵ that are not boundary parallel in $\Phi_{m-1}^{(-1)^m\epsilon}$. For each k, let C_k be an annulus isotopic in Φ_{m-1}^ϵ to $(h_{m-1}^\epsilon)^{-1}(C_k')$ and disjoint from $\Phi_m^\epsilon \cup \bigcup B_j$.

Properties (1) and (3) are clear from this construction. Since Ψ_m^ϵ admits a reduced homotopy of length m by construction, it remains only to show that if $H\colon (K \times I, K \times \partial I) \to (M, S)$ is a reduced homotopy of length m, then H_0 is homotopic into Ψ_m^ϵ. If $H_0\colon K \to S$ is large, then this holds by [3, Proposition 5.2.8] and the fact that $\Phi_m^\epsilon \subset \Psi_m^\epsilon$, so using Lemma 2.3, we will assume $K = S^1$.

Write H as a composition of essential basic homotopies $H^1, \ldots, H^m$. Since the composition of $H^1, \ldots, H^{m-1}$ is a reduced homotopy of length $m-1$, by hypothesis

H_0 is homotopic into Ψ^ϵ_{m-1}. If H_0 is homotopic into an annular component of Ψ^ϵ_{m-1}, then by Lemma 3.4, this component admits a reduced homotopy of length k; hence it is of the form A_i for some i. If H_0 is homotopic into a boundary curve of Φ^ϵ_{m-1} that is not homotopic into Φ^ϵ_m, then by Lemma 3.4 again, the corresponding boundary component is of the form $\mathfrak{b}_j$ for some j.

If H_0 is homotopic into Φ^ϵ_m, we are done; therefore we assume that H_0 is homotopic to a curve $\mathfrak{c} \subset \Phi^\epsilon_{m-1}$ that is not boundary parallel or homotopic into Φ^ϵ_m. By Lemma 3.3, $h^\epsilon_{m-1} \circ \mathfrak{c} \subset \Phi^{\epsilon\cdot(-1)^m}_{m-1}$ is homotopic to H_1^{m-1} in S. It follows that $h^\epsilon_{m-1} \circ \mathfrak{c}$ admits an essential homotopy of length one, hence by Proposition 2.2 it is homotopic into P^ϵ_m.

Let C' be the component of P^ϵ_m into which $h^\epsilon_{m-1} \circ \mathfrak{c}$ is homotopic. If this homotopy cannot be taken to occur in $\Phi^{\epsilon(-1)^m}_{m-1}$, then Proposition 2.4 implies in particular that $h^\epsilon_{m-1} \circ \mathfrak{c}$ is homotopic in $\Phi^{\epsilon(-1)^m}_{m-1}$ into a boundary component. But then $\mathfrak{c}$ would be homotopic to a boundary component of Φ^ϵ_{m-1}, contradicting our assumption. Hence $h^\epsilon_{m-1} \circ \mathfrak{c}$ is homotopic into C' in $\Phi^{\epsilon(-1)^m}_{m-1}$.

If C' is large, then since P^ϵ_m is isotopic in $\Phi^{\epsilon(-1)^m}_{m-1}$ to $h^\epsilon_{m-1}(\Phi^\epsilon_m)$, $\mathfrak{c}$ is homotopic into Φ^ϵ_m. But this contradicts our assumption that it is not. Therefore C' is an annulus, of the form C'_k for some k, and we are in case (3) above. □

The second main result of this section asserts that the sequence $\{\Psi^\epsilon_k\}$ is shrinking. We cannot hope to establish that Ψ^ϵ_k is properly larger than Ψ^ϵ_{k+1} for each k. Indeed, in the case of interest to us — when S is the boundary of a regular neighborhood of a non-separating surface — Ψ^ϵ_k is identical to Ψ^ϵ_{k+1} for each odd or even k (depending on ϵ). Instead we obtain the following extension of [3, Proposition 5.3.9].

Proposition 3.5. *Let M be a closed, orientable hyperbolic 3-manifold and $S \subset M$ a two-sided incompressible surface that is not a fiber or semi-fiber such that $M \setminus\setminus S$ has two components $X^{\pm 1}$. For $\epsilon \in \{\pm 1\}$, let $\Psi^\epsilon_1 \supset \Psi^\epsilon_2 \supset \cdots$ be a sequence of irredundant surfaces that satisfy Theorem 3.1. Then for each k, Ψ^ϵ_k is not homotopic into Ψ^ϵ_{k+2}.*

Proof. Proposition 5.3.9 of [3] asserts that in this situation Φ^ϵ_k is not homotopic into Φ^ϵ_{k+2} for any $k \in \mathbb{N}$ or $\epsilon \in \{\pm 1\}$, so the result holds as long as Ψ^ϵ_k has nonempty large part. Suppose therefore that for some k, Ψ^ϵ_k is a disjoint union of annuli homotopic into Ψ^ϵ_{k+2}.

Let H be a reduced homotopy in (M, S) of length $k+2$ with domain Ψ_{k+2} that starts on the ϵ-side, and write H as the composition of H'', H', each starting on the ϵ-side, where H' has length 2 and H'' length k. Since $H'_1(\Psi^\epsilon_{k+2})$ admits a reduced homotopy of length k, Proposition 3.1 implies that H'_1 is homotopic to a map $f\colon \Psi^\epsilon_{k+2} \to \Psi^\epsilon_k$. After applying the homotopy that takes Ψ^ϵ_k into Ψ^ϵ_{k+2}, we may take f to map into Ψ^ϵ_{k+2}. It follows that there exists a homotopy of length 2 in (M, S) with domain and target Ψ^ϵ_{k+2}.

Associate a directed graph G to this homotopy as follows: G has a vertex v for each component of Ψ^ϵ_{k+2}, and a directed edge joining v to v' if and only if the component associated to v is taken to the component associated to v' by the time-1 map of the homotopy described above. Then every vertex has a unique edge that leaves it, and so G has a cycle.

We associate to a cycle $v_0, \ldots, v_{m-1}$ a map of a torus into (M, S) as follows. For $0 \le i < m$, let $\mathfrak{a}_i$ be the core of the component of Ψ^ϵ_{k+2} corresponding to v_i, and let $F^i\colon (S^1\times I, S^1\times\partial I)\to(M,S)$ be a reduced homotopy of length 2 with $F^i_0=\mathfrak{a}_i$ and $F^i_1=\mathfrak{a}_{i+1}$ (where $i+1$ is taken modulo m). Dividing a torus T into m concentric essential annuli A_i, each homeomorphic to $S^1\times I$, we obtain a map $F\colon T\to M$ that restricts on A_i to F^i for each i. Since each F^i is essential, F is essential, contradicting hyperbolicity of M. □

We may now prove Theorem 1.4, which extends Theorem 5.4.1 of [3].

Proof of Theorem 1.4. If S is non-separating, we replace S by the boundary $\tilde{S}$ of a regular neighborhood, yielding a separating surface with two components of genus g. If S is separating we take $\tilde{S}=S$, and in either case let $X^{\pm1}$ be the components of $M\setminus\setminus\tilde{S}$. For $\epsilon\in\{\pm1\}$, let $\Psi^\epsilon_1\supset\Psi^\epsilon_2\supset\ldots$ be a sequence of irredundant surfaces that satisfies the conclusion of Proposition 3.1.

We now briefly review the proof of [3, Theorem 5.4.1]. Given a large surface A, a complexity of A is defined as $c(A)=g(A)-3\chi(A)/2-|A|$, where $\chi(A)$ is the Euler characteristic of A, $|A|$ is the number of its components, and $g(A)$ is the sum of their genera. It is easy to see that if A is nonempty and large, then $c(A)>0$. The key fact established in the proof of Theorem 5.4.1 is that if A and $B\subset A$ are large surfaces with even Euler characteristic, then $c(B)<c(A)$ unless A is a regular neighborhood of B.

Fixing $\epsilon\in\{\pm1\}$, consider the subsequence

$$\Phi^\epsilon_1\supset\Phi^\epsilon_3\supset\ldots$$

This is strictly decreasing by [3, Proposition 5.3.9], and consists of large surfaces with even Euler characteristic by [3, Corollary 5.3.8]. Thus for each $i\geq0$, $c(\Phi^\epsilon_{2i+1})>c(\Phi^\epsilon_{2i+3})$. If S is separating, then $c(\tilde{S})=c(S)=4g-4$, and otherwise $c(\tilde{S})=8g-8$. Taking $m_S=4g-4$ in the separating case and $m_S=8g-8$ in the non-separating case, it follows that for $i>m_S$, $\Phi_{2i+1}=\emptyset$.

The discussion above is enough to establish [3, Theorem 5.4.1]. In our situation of interest, it establishes that Ψ^ϵ_{2i+1} is a disjoint union of annuli for $i>m_S$. Since Ψ^ϵ_i is irredundant, Ψ_{2m_S+3} has at most $3g-3$ components in the separating case, and $6g-6$ otherwise. (This uses the standard fact that a collection of disjoint, non-parallel, essential simple closed curves on a closed surface of genus g has at most $3g-3$ members.) Since Proposition 3.5 implies Ψ^ϵ_{2i+1} is not homotopic into Ψ^ϵ_{2i+3}, if these are unions of irredundant collections of annuli then Ψ^ϵ_{2i+3} has fewer components than Ψ^ϵ_{2i+1}. Thus taking $n_S=3g-3$ in the separating case and $n_S=6g-6$ otherwise, we find that $\Psi_{2i+1}=\emptyset$ for $i>m_S+n_S$.

By Proposition 3.1, the time-0 map of a reduced homotopy in $(M,\tilde{S})$ with length k that starts on the ϵ-side is homotopic into Ψ^ϵ_k. Therefore $k\leq2(m_S+n_S)+2$. If S is separating, we therefore find that homotopies in $(M,S)=(M,\tilde{S})$ have length at most $14g-12$. If S is non-separating, a reduced homotopy of length k in (M,S) determines a reduced homotopy of length $2k-1$ in $(M,\tilde{S})$. Thus in this case we have for a homotopy of length k in (M,S) that $2k-1\leq2(14g-14)+2$, so $k\leq14g-13$. The theorem follows. □

4. Questions and further directions

Two directions immediately suggest themselves as avenues for further exploration of rank gradient questions. The most direct generalization seeks to expand the class of groups under consideration. Here is a sample conjecture.

Conjecture. The conclusion of Theorem 0.1 holds under the hypothesis that M is a compact, orientable, irreducible (not necessarily hyperbolic) 3-manifold.

This is easily seen to hold for manifolds with positive-genus boundary. Manifolds with boundary that is a disjoint union of tori or empty, and trivial JSJ decomposition, are covered by Theorem 0.1 and the following.

Proposition 4.1. *The conclusion of Theorem 0.1 holds under the hypothesis that M is a compact, orientable Seifert fibered space and $\phi\colon \pi_1 M \twoheadrightarrow \mathbb{Z}$.*

Proof. This uses the "horizontal-vertical" description of incompressible surfaces in Seifert fibered spaces (see eg. [8, Proposition 3.5]). Let S represent the Poincaré dual of ϕ; if S is a horizontal surface, then it is a fiber in a fibration $M \to S^1$ and the rank gradient of the corresponding family of covers is 0. We may thus assume S is a vertical torus $\pi^{-1}(\mathfrak{c})$, where $\pi\colon M \to B$ is projection to the base orbifold taking each Seifert fiber to a point and $\mathfrak{c}$ is an essential curve on B. S does not separate M only if $\mathfrak{c}$ does not separate B, hence $\phi\colon \pi_1(M) \to \mathbb{Z}$ factors through $\pi_1(B)$ (here "$\pi_1 B$" refers to the orbifold fundamental group of B).

If B is hyperbolic, then using Euler characteristic we find that the rank gradient of the corresponding family of covers $\{B_n\}$ is positive; since the $\pi_1 M_n$ map onto the $\pi_1 B_n$, the conclusion follows in this case. If B is spherical then it does not possess non-separating curves. If B is Euclidean and $\mathfrak{c}$ is non-separating, the classification of Seifert fibered spaces over Euclidean orbifolds implies that B is a torus or Klein bottle with empty singular locus (see eg. [18]). Thus $\mathfrak{c}$ is a fiber in a fibering of B by circles, hence S is a fiber in a fibering of M by vertical tori, and the rank gradient is 0. □

The main work in proving the conjecture above thus falls to understanding rank gradient of 3-manifolds with nontrivial JSJ decomposition. Along these lines, we note that work of Weidmann [28] bounds the rank of such a manifold below linearly in terms of the number of hyperbolic pieces.

Stallings' fibration theorem [24] suggests a form of Theorem 0.1 well-suited to generalization beyond fundamental groups of manifolds.

Theorem 0.1′. *Let G be the fundamental group of an orientable hyperbolic 3-manifold of finite volume, and let $\phi\colon G \twoheadrightarrow \mathbb{Z}$. Then $\mathrm{rg}\,\{\phi^{-1}(n\mathbb{Z})\} = 0$ if and only if $\ker\phi$ is finitely generated.*

One might ask whether the conclusions of Theorem 0.1′ apply to, say, hyperbolic groups. It is indeed even possible that our proof strategy applies in this more general context, since these also do not contain abelian groups of rank greater than 1. Furthermore, a JSJ decomposition that applies to these groups and has the enclosing property has been described by Scott–Swarup [19]. Extending our proof strategy to these groups thus only requires proving a version of "cylinders have bounded length" for the class of all hyperbolic groups.

Another direction for generalizing Theorem 0.1 lies in considering other families of covers. I am particularly interested in the following question.

Question. Under which circumstances does a *co-final* family of finite covers — that is, $\{M_n \to M\}$ such that $\bigcap \pi_1 M_n = \{1\} \subset \pi_1 M$ — have positive rank gradient?

This is motivated by work of Abert–Nikolov [1], that uses a theorem of Lackenby [15, Theorem 1.5] to relate "rank vs. Heegaard genus" to a question of Gaboriau

[6] concerning cost of group actions. Attacking this question would require new techniques.

Appendix: Tree actions and homotopies

This appendix is dedicated to proving Lemma 1.3, which draws the connection between cylindricity of the action of a 3-manifold group on a tree and cylinders (ie, annuli) immersed in the manifold. This connection was alluded to in Example (iv) of the introduction to [21], as a way of motivating the name "acylindrical accessibility."

Recall the set-up for Lemma 1.3. A connected, two-sided incompressible surface S in a closed oriented hyperbolic 3-manifold M determines a graph of spaces decomposition upon choosing a transverse orientation for S and a homeomorphism $f\colon S \times [-1,1] \to \mathcal{N}$, where $\mathcal{N}$ is a closed regular neighborhood of S in M, that takes $S \times \{0\}$ to S and preserves the orientation and transverse orientation. We will take $S_\pm = f(S \times \{\pm 1/2\})$, and let $i_\pm\colon S_\pm \to S$ be determined by the projection $(x, \pm 1/2) \mapsto (x, 0)$. Below we will frequently suppress the homeomorphism f and regard $S \times [-1/2, 1/2]$ as a regular neighborhood of S in M.

If $X \doteq M - S \times (-1/2, 1/2)$ is connected, then the associated graph of spaces G has a single vertex, corresponding to X, and edge, corresponding to S. In this case the associated graph of groups decomposition of $\pi_1 M$ describes it as the HNN-extension of $\pi_1 X$ over $\pi_1 S$. If S is separating, then X has two components $X^\pm$ such that $S_\pm = \partial X^\pm$, and G has vertices $v_\pm$ corresponding to $X^\pm$, respectively, and a unique edge corresponding to S. In this case, $\pi_1 M$ is the free product of $\pi_1 X^+$ with $\pi_1 X^-$, amalgamated over $\pi_1 S$. See [20, p. 155].

In either case above, there is an action of $\pi_1 M$ on a tree T, without inversions and with quotient graph G, such that the stabilizer of each edge is a conjugate of $\pi_1 S$ and that of each vertex is a conjugate of the fundamental group of a component of X. (See [20, pp. 166–167].) We will treat the cases individually, with the separating case first because it is less complicated.

Proof of Lemma 1.3, separating case. We first establish some notation. Fix $x \in S$, and let $\Gamma_- = \pi_1(X^-, (x, -1/2))$ and $\Lambda = \pi_1(S_-, f(x, -1/2))$. Let $\alpha\colon I \to M$ be the path $t \mapsto (x, t - 1/2)$ that joins $(x, -1/2) \in S_-$ to $(x, 1/2) \in S_+$. Take $\Gamma_+ = \{\alpha.\gamma.\bar{\alpha} \,|\, \gamma \in \pi_1(X^+, (x, 1/2))\}$. Then $\pi_1(M, (x, -1/2))$ is isomorphic to the free product with amalgamation $\Gamma_- *_\Lambda \Gamma_+$. We note that $\Lambda < \Gamma_+$, since each $\lambda \in \Lambda$ has a canonical pointed homotopy through $S \times [-1/2, 1/2]$ to $\alpha.\lambda'.\bar{\alpha}$, where $\lambda' = (i_+^{-1} \circ i_-)_*(\lambda) \in \pi_1(S_+, (x, 1/2))$.

Let v_+ be the vertex of G corresponding to X^+ and v_- the vertex corresponding to X^-, and orient the edge e to point from v_- to v_+. Orient the edges of T so that projection to G preserves orientation. Now fixing $k \geq 1$, suppose that the action of $\pi_1(M, f(x, 1/2))$ on T is k-cylindrical. Then some $g \in \pi_1(M, f(x, 1/2))$ fixes a segment in T of length $k+1$. Since the action is transitive on edges, we may assume that an initial edge e_0 in this segment is stabilized by Λ, thus its endpoints are stabilized by Γ_+ and Γ_-.

Let v_0 be the endpoint contained in e_0 of the segment stabilized by g, and for $1 \leq i \leq k$ let v_i be the vertex on this segment at distance i from v_0. We will assume that v_0 is stabilized by Γ_- (the other case is completely analogous); then v_1 is stabilized by Γ_+, and v_i is stabilized by a conjugate of Γ_- (or Γ_+) for each even (or, respectively, odd) i. For $1 \leq i \leq k$, let e_i be the edge joining v_i and v_{i+1}.

Claim. For $1 \leq i \leq k$, there exists γ_i, in $\Gamma_- - \Lambda$ for i even and $\Gamma_+ - \Lambda$ for i odd, so that taking

$$x_i = \gamma_1\gamma_2\cdots\gamma_i,$$

we have $x_i(e_0) = e_i$.

Proof. The Γ_+-action is transitive on the edges that abut v_1, so there exists $\gamma_1 \in \Gamma_+$ such that $\gamma_1(e_0) = e_1$. This element is not in Λ, since Λ stabilizes e_0. Now for $i > 1$, assume that the claim holds for $i-1$. Then there exist $\gamma_1, \ldots, \gamma_{i-1}$ such that $x_{i-1} = \gamma_1\cdots\gamma_{i-1}$ takes e_0 to e_{i-1}. Thus $x_{i-1}^{-1}(e_{i-1}) = e_0$.

If i is even, then $x_{i-1}^{-1}(v_i) = v_0$ since e_{i-1} points away from v_i. Let $\gamma_i \in \Gamma_- - \Lambda$ be the element that takes e_0 to $x_{i-1}^{-1}(e_i)$. If i is odd, then $x_{i-1}^{-1}(v_i) = v_1$, and there is an element $\gamma_i \in \Gamma_+ - \Lambda$ taking e_0 to $x_{i-1}^{-1}(e_i)$.

Now taking $x_i = x_{i-1}\gamma_i$, we have

$$x_i(e_0) = x_{i-1}\gamma_i(e_0) = x_{i-1}(x_{i-1}^{-1}(e_i)) = e_i,$$

and the claim follows by induction. □

Since $\Lambda = \mathrm{Stab}_{\pi_1 M}(e_0)$, the claim and the fact that g stabilizes the edges e_i, $0 \leq i \leq k$ together imply that for each such i, there exists $\lambda_i \in \Lambda$ such that $g = x_i\lambda_i x_i^{-1}$. We will use this description to obtain essential basic homotopies $H^1, \ldots, H^k\colon (S^1 \times I, S^1 \times \partial I) \to (M, S)$, whose composition will establish the lemma.

Fix i between 1 and k, and suppose first that i is even. Since e_{i-1} and e_i each abut v_i, we have

$$x_{i-1}\lambda_{i-1}x_{i-1}^{-1} = g = x_i\lambda_i x_i^{-1} \in x_{i-1}\Gamma_+ x_{i-1}^{-1}.$$

Using the descriptions of the x_i, after cancelling x_{i-1} we find $\lambda_{i-1} = \gamma_i\lambda_i\gamma_i^{-1}$.

We now define $H^i\colon (S^1\times I, S^1\times\partial I) \to (X^-, S_-)$ as follows. For each i, divide the annulus $S^1\times I$ into concentric essential sub-annuli $A^{(1)}$ and $A^{(2)}$, so that $S^1\times\{0\} \subset A^{(1)}$ and $S^1 \times \{1\} \subset A^{(2)}$. Let the restriction of H^i to $A^{(1)}$ be determined by the pointed homotopy between λ_{i-1} and $\gamma_i\lambda_i\gamma_i^{-1}$, so that in particular $H_0^i = \lambda_{i-1}$. Then let $H^i|_{A^{(2)}}$ be determined by the free homotopy between $\gamma_i\lambda_i\gamma_i^{-1}$ and λ_i, so that $H_1^i = \lambda_i$.

Suppose now that i is odd. In this case we find as above that $\lambda_{i-1} = \gamma_i\lambda_i\gamma_i^{-1}$, but this time all elements are in Γ_+. Let λ'_{i-1} and λ'_i be the images in $\pi_1(S_+, (x, 1/2))$ of λ_{i-1} and λ_i, respectively, under $(i_+^{-1} \circ i_-)_*$, and let $\gamma'_i \in \pi_1(X^+, (x, 1/2))$ satisfy $\gamma_i = \alpha.\gamma'_i.\bar{\alpha}$. Since $\lambda_i = \alpha.\lambda'_i.\bar{\alpha}$ and the same holds for λ_{i-1}, the relation above implies that $\lambda'_{i-1} = \gamma'_i\lambda'_i(\gamma'_i)^{-1}$.

We now obtain a homotopy $H^i\colon (S^1 \times I, S^1 \times \partial I) \to (X^+, S_+)$ as in the even case, with the property that $H_0^i = \lambda'_{i-1}$ and $H_1^i = \lambda'_i$. In either the odd or the even case, a homotopy of H^i into $S_\pm$ through maps $(S^1\times I, S^1\times\partial I) \to (X^\pm, S_\pm)$ would determine a pointed homotopy of γ_i into Λ_i, contradicting the claim. Hence there is no such homotopy, and H^i is essential.

There is a standard way to extend each $H^i\colon (S^1\times I, S^1\times\partial I) \to (X^\pm, S_\pm)$ to an essential basic homotopy in (M, S), as follows. Supposing i is even, we replace H^i by the composition of $\bar{K}^{i-1}, H^i, K^i$, where $K^i(s,t) = (\lambda_i(s), \frac{t-1}{2})$ and $\bar{K}^{i-1}$ is the reverse of K^{i-1}. If i is odd, then we replace H^i by the composition of $\bar{L}^{i-1}, H^i, L^i$, where $L^i(s,t) = (\lambda'_i(s), \frac{1-t}{2})$.

Referring to the extended map again as H^i, we obtain a basic essential homotopy with target (M,S), starting on the $+$ side for i odd and the $-$ side for i even. Furthermore, for i odd we find that

$$H^i_0 = (i_+)_*(\lambda'_{i-1}) = (i_-)_*(\lambda_{i-1}) \qquad\qquad H^i_1 = (i_+)_*(\lambda'_i) = (i_-)_*(\lambda_i).$$

For i even, the extended H^i similarly has time-0 map $(i_-)_*(\lambda_{i-1})$ and time-1 map $(i_-)_*(\lambda_i)$. It thus follows that there is a homotopy of length k in (M,S) obtained as the composition of $H^1,\ldots,H^k$. □

The proof when S is non-separating follows the same outline but is somewhat more involved in execution. Recall that in this case $X = M - S\times(-1/2,1/2)$ is connected, with two boundary components $S_\pm = S\times\{\pm 1/2\}$.

Proof of Lemma 1.3, non-separating case. We first establish notation. Fixing $x\in S$, define $\Gamma = \pi_1(X,(x,-1/2))$ and $\Lambda_- = \pi_1(S_-,(x,-1/2))$. Let $\alpha\colon t\mapsto (x,t-1/2)$ join $(x,-1/2)$ to $(x,1/2)$ in $\mathcal{N}$, let β be a path in X joining $(x,1/2)$ to $(x,-1/2)$, and let τ be the pointed homotopy class of $\alpha.\beta$ in $\pi_1(M,(x,0))$. Define $\Lambda_+ = \{\bar\beta.\lambda.\beta \,|\, \lambda\in\pi_1(S_+,(x,1/2))\}$, and note that for $\lambda_1 = \bar\beta.\lambda.\beta\in\Lambda_1$, where $\lambda\in\pi_1(S_+,(x,1/2))$, its τ-conjugate $\alpha.\lambda.\bar\alpha$ has a canonical pointed homotopy through $S\times[-1/2,1/2]$ to $(i_-^{-1}i_+)_*(\lambda)\in\pi_1(S_-,(x,-1/2))$.

Let v and e be the unique vertex and edge, respectively, of the associated graph G. Let v_0 be the vertex of T such that $\Gamma = \mathrm{Stab}(v_0)$, and let edges e_0 and e_1 be the edges of T abutting v_0 such that $\Lambda_- = \mathrm{Stab}(e_0)$ and $\Lambda_+ = \mathrm{Stab}(e_1)$. Since conjugation by τ takes Λ_+ to Λ_-, we find that $\tau(e_1) = e_0$. Orient e_0 pointing away from v_0 and e_1 pointing toward it. Then τ preserves orientation, so this determines an orientation on e and hence on all edges of T. Since $\partial X = S_-\sqcup S_+$ and G has a single edge, each edge of T abutting v_0 is Γ-conjugate to either e_0 or e_1.

Fixing $k\geq 1$, we suppose that the action of $\pi_1(M,(x,-1/2))$ on T is k-cylindrical. Thus some $g\in\pi_1(M,(x,-1/2))$ fixes a segment in T of length $k+1$. Since the action of $\pi_1 M$ is transitive on vertices of T, after conjugating g we may assume that v_0 is an endpoint of this segment. For $1\leq i\leq k+1$, let v_i be the vertex on this segment at distance i from v_0, and let $\epsilon_i\in\{\pm 1\}$ be -1 if the edge joining v_{i-1} and v_i points from v_{i-1} to v_i, and 1 otherwise.

Claim. For $1\leq i\leq k+1$ there exist $\gamma_i\in\Gamma$ so that taking

$$x_i = (\gamma_1\,\tau^{\epsilon_1})\ldots(\gamma_i\,\tau^{\epsilon_i}),$$

we have $x_i(v_0) = v_i$. Defining $x_0 = id$, if e is the edge joining v_{i-1} to v_i, then $e = \begin{cases} x_{i-1}\gamma_i(e_0) & \text{if } \epsilon_i = -1 \\ x_{i-1}\gamma_i(e_1) & \text{if } \epsilon_i = 1.\end{cases}$ Also, for $i\leq k$ if $\epsilon_{i+1}\neq\epsilon_i$, then $\gamma_{i+1}\notin\Lambda_{\epsilon_{i+1}}$.

Proof. We will prove the claim by induction. If the edge e joining v_0 to v_1 points away from v_0, then there exists $\gamma_0\in\Gamma$ such that $e = \gamma_0(e_0)$. Since the other endpoint of e_0 is $\tau(v_0)$, the other endpoint of e is $\gamma_0\tau(v_0)$. If e points toward v_0, then $e = \gamma_0(e_1)$ for some $\gamma_0\in\Gamma$; hence $v_1 = \gamma_0\tau^{-1}(v_0)$ in this case. This establishes the base case.

For $i>1$, suppose that we have established the claim for $i-1$. If the edge e joining v_{i-1} and v_i points away from v_{i-1}, then $x_{i-1}^{-1}(e) = \gamma_i(e_0)$ for some $\gamma_i\in\Gamma$, so the other endpoint of $x_{i-1}^{-1}(e)$ is $\gamma_i\tau(v_0)$ and $v_i = x_{i-1}\gamma_i\tau(v_0)$. If e points toward v_{i-1}, then $x_{i-1}^{-1}(e) = \gamma_i(e_1)$ for some $\gamma_i\in\Gamma$, and $v_i = x_{i-1}\gamma_i\tau^{-1}(v_0)$. It now follows by induction that for all i, $v_i = x_i(v_0)$ and the edge e joining v_{i-1} to v_i is $x_{i-1}\gamma_i(e_{\epsilon_i})$.

In addressing the final assertion of the claim, let us suppose that for some i, $\epsilon_i = -1$ and $\epsilon_{i+1} = 1$. Using the description above, we have $v_i = x_i(v_0)$ and, taking e to be the edge joining v_{i-1} to v_i,

$$e = x_{i-1}\gamma_i(e_0) = x_i\tau^{-1}(e_0) = x_i(e_1).$$

Taking e' to be the edge joining v_i and v_{i+1}, we have $e' = x_i\gamma_{i+1}(e_1)$. Since Λ_1 stabilizes e_1 and $e' \neq e$, it follows that $\gamma_{i+1} \notin \Lambda_1$. The case $\epsilon_i = 1$, $\epsilon_{i+1} = -1$ follows similarly, establishing the claim. □

Since $\Lambda_- = \mathrm{Stab}_{\pi_1 M}(e_0)$ and $\Lambda_+ = \mathrm{Stab}_{\pi_1 M}(e_1)$, the claim's description of the edge joining v_{i-1} to v_i implies that for $1 \leq i \leq k+1$, there exists $\lambda_i \in \Lambda_{\epsilon_i}$ such that $g = (x_{i-1}\gamma_i)\lambda_i(x_{i-1}\gamma_i)^{-1}$. We will obtain homotopies H^i, $1 \leq i \leq k$, by comparing the descriptions of g obtained from adjacent edges. There are four cases.

Case 1: $\epsilon_{\mathbf{i}} = \epsilon_{\mathbf{i+1}} = -\mathbf{1}$

In this case we have $\lambda_i, \lambda_{i+1} \in \Lambda_-$, and using the claim, that $x_i = x_{i-1}\gamma_i\tau$. Thus:

$$(1) \qquad (x_{i-1}\gamma_i)\lambda_i(x_{i-1}\gamma_i)^{-1} = (x_{i-1}\gamma_i\tau\gamma_{i+1})\lambda_{i+1}(x_{i-1}\gamma_i\tau\gamma_{i+1})^{-1}$$

We note that $\tau^{-1}\lambda_i\tau \in \Lambda_+$; in fact, we have $\tau^{-1}\lambda_i\tau = \bar{\beta}.(i_+^{-1} \circ i_-)_*(\lambda_i).\beta$, where $(i_+^{-1} \circ i_-)_*(\lambda_i) \in \pi_1(S_+, (x, 1/2))$. Rearranging the terms of (1), we find

$$\bar{\beta}.(i_+^{-1} \circ i_-)_*(\lambda_i).\beta = \gamma_{i+1}\lambda_{i+1}\gamma_{i+1}^{-1}.$$

Divide the annulus $S^1 \times I$ into concentric essential sub-annuli $A^{(1)}$, $A^{(2)}$, and $A^{(3)}$ such that $S^1 \times \{0\} \subset A^{(1)}$ and $S^1 \times \{1\} \subset A^{(3)}$. We define $H^i\colon (S^1 \times I, S^1 \times \partial I) \to (X, \partial X)$ as follows:

- H^i is determined on $A^{(1)}$ by the free homotopy between $H^i_0 \doteq (i_+ \circ i_-^{-1})_*(\lambda_i)$ and $H^i|_{A^{(1)} \cap A^{(2)}} \doteq \bar{\beta}.(i_+^{-1} \circ i_-)_*(\lambda_i).\beta$.
- $H^i|_{A^{(2)}}$ is determined by the pointed homotopy between $\bar{\beta}.(i_+^{-1} \circ i_-)_*(\lambda_i).\beta$ and $\gamma_{i+1}\lambda_{i+1}\gamma_{i+1}^{-1}$.
- $H^i|_{A^{(3)}}$ is determined by the free homotopy between $\gamma_{i+1}\lambda_{i+1}\gamma_{i+1}^{-1}$ and λ_{i+1}; in particular, $H^i_1 = \lambda_{i+1} \in \Lambda_-$.

In this case, H^i maps $S^1 \times \{0\}$ and $S^1 \times \{1\}$ into different components of ∂X; hence it is essential; ie, π_1-injective (which is evident by construction) and not homotopic into ∂X through maps $(S^1 \times I, S^1 \times \partial I) \to (X, \partial X)$.

Case 2: $\epsilon_{\mathbf{i}} = -\mathbf{1},\ \epsilon_{\mathbf{i+1}} = \mathbf{1}$

In this case we have $\lambda_i \in \Lambda_-$, $\lambda_{i+1} \in \Lambda_+$, and again that $x_i = x_{i-1}\gamma_i\tau$. We thus obtain the same relation as in (1) above, and after rearranging we again have $\bar{\beta}.(i_+^{-1} \circ i_-)_*(\lambda_i).\beta = \gamma_{i+1}\lambda_{i+1}\gamma_{i+1}^{-1}$, where $(i_+^{-1} \circ i_-)_*(\lambda_i) \in \pi_1(S_+, (x, 1/2))$.

In this case divide $S^1 \times I$ into four concentric essential sub-annuli $A^{(1)}, \ldots, A^{(4)}$ so that $S^1 \times \{0\} \subset A^{(1)}$ and $S^1 \times \{1\} \subset A^{(4)}$, and for each $j > 1$, $A^{(j)}$ intersects $A^{(j-1)}$. We define $H^i\colon (S^1 \times I, S^1 \times \partial I) \to (X, \partial X)$ as follows: for $j = 1, 2, 3$, the restriction of H^i to $A^{(j)}$ is specified exactly as in the previous case, and $H^i|_{A^{(4)}}$ is determined by the free homotopy between λ_{i+1} and the element $\lambda'_{i+1} \in \pi_1(S_+, (x, 1/2))$ such that $\lambda_{i+1} = \bar{\beta}.\lambda'_{i+1}.\beta$. In particular, we require $H^i_1 = \lambda'_{i+1}$.

In this case H^i maps each of $S^1 \times \{0\}$ and $S^1 \times \{1\}$ into S_+. A homotopy of H^i, through maps $(S^1 \times I, S^1 \times \partial I) \to (X, \partial X)$, to a map with image in S_+ would determine a pointed homotopy of γ_{i+1} to an element of Λ_1, but this would contradict the final assertion of the claim. Therefore H^i is essential.

Case 3: $\epsilon_{\mathbf{i}} = \mathbf{1}$, $\epsilon_{\mathbf{i+1}} = \mathbf{-1}$

Now we have $\lambda_i \in \Lambda_+$, $\lambda_{i+1} \in \Lambda_-$, and $x_i = x_{i-1}\gamma_i\tau^{-1}$. Therefore:

$$(2) \qquad (x_{i-1}\gamma_i)\lambda_i(x_{i-1}\gamma_i)^{-1} = (x_{i-1}\gamma_i\tau^{-1}\gamma_{i+1})\lambda_{i+1}(x_{i-1}\gamma_i\tau^{-1}\gamma_{i+1})^{-1}$$

Since $\lambda_i \in \Lambda_+$, there is an element $\lambda'_i \in \pi_1(S_+, (x, 1/2))$ such that $\lambda_i = \bar{\beta}.\lambda'_i\beta$, and $\tau\lambda_i\tau^{-1}$ has a canonical pointed homotopy through $S \times [-1/2, 1/2]$ to $(i_-^{-1} \circ i_+)_*(\lambda'_i) \in \Lambda_-$. Rearranging terms of (2) yields:

$$(i_-^{-1} \circ i_+)_*(\lambda'_i) = \gamma_{i+1}\lambda_{i+1}\gamma_{i+1}^{-1}.$$

Dividing $S^1 \times I$ into concentric essential sub-annuli $A^{(1)} \supset S^1 \times \{0\}$ and $A^{(2)} \supset S^1 \times \{1\}$, we define $H^i\colon (S^1 \times I, S^1 \times \partial I) \to (X, \partial X)$ as follows: let H^i be determined on $A^{(1)}$ by the pointed homotopy between $(i_-^{-1} \circ i_+)_*(\lambda'_i)$ and $\gamma_{i+1}\lambda_{i+1}\gamma_{i+1}^{-1}$, with $H^i_0 = (i_-^{-1} \circ i_+)_*(\lambda'_i)$, and let it be determined on $A^{(2)}$ by the free homotopy between $\gamma_{i+1}\lambda_{i+1}\gamma_{i+1}^{-1}$ and λ_{i+1}, with $H^i_1 = \lambda_{i+1}$.

In this case H^i takes each of $S^1 \times \{0\}$ and $S^1 \times \{1\}$ into S_-. A homotopy of H^i, through maps $(S^1 \times I, S^1 \times \partial I) \to (X, \partial X)$, to a map with image in S_- would determine a pointed homotopy of γ_{i+1} to an element of Λ_0, but this would contradict the final assertion of the claim. Therefore H^i is essential.

Case 4: $\epsilon_{\mathbf{i}} = \epsilon_{\mathbf{i+1}} = \mathbf{1}$

We have $\lambda_i, \lambda_{i+1} \in \Lambda_+$, and $x_i = x_{i-1}\gamma_i\tau^{-1}$; therefore the equation (2) applies in this case as well. Again there is an element $\lambda'_i \in \pi_1(S_+, (x, 1/2))$ such that $\lambda_i = \bar{\beta}.\lambda'_i.\beta$, and in this case there is also $\lambda'_{i+1} \in \pi_1(S_+, (x, 1/2))$ such that $\lambda_{i+1} = \bar{\beta}.\lambda'_{i+1}.\beta$. Rearranging terms of (2) again produces the relation $(i_-^{-1} \circ i_+)_*(\lambda'_i) = \gamma_{i+1}\lambda_{i+1}\gamma_{i+1}^{-1}$.

Divide $S^1 \times I$ into three concentric essential sub-annuli $A^{(1)}$, $A^{(2)}$, and $A^{(3)}$ as in Case 1, and define $H^i\colon (S^1 \times I, S^1 \times \partial I) \to (X, \partial X)$ as follows:

- H^i is determined on $A^{(1)}$ by the pointed homotopy between $H^i_0 \doteq (i_-^{-1} \circ i_+)_*(\lambda'_i)$ and $H^i|_{A^{(1)} \cap A^{(2)}} \doteq \gamma_{i+1}\lambda_{i+1}\gamma_{i+1}^{-1}$.
- $H^i|_{A^{(2)}}$ is determined by the free homotopy between $\gamma_{i+1}\lambda_{i+1}\gamma_{i+1}^{-1}$ and λ_{i+1}.
- $H^i|_{A^{(3)}}$ is determined by the free homotopy between λ_{i+1} and λ'_{i+1}, with $H^i_1 = \lambda'_{i+1}$ in particular.

Since H^i takes opposite boundary components of $S^1 \times I$ to opposite components of ∂X, it is not homotopic into ∂X through maps $(S^1 \times I, S^1 \times \partial I) \to (X, \partial X)$. Hence H^i is essential.

Each homotopy H^i defined above, for $1 \le i \le k$, has property that H^i_0 is either $(i_+^{-1} \circ i_-)_*(\lambda_i)$, when $\lambda_i \in \Gamma_-$ (as in Cases 1 and 2 above), or $(i_-^{-1} \circ i_+)_*(\lambda'_i)$, for $\lambda'_i \in \pi_1(S_+, (x, 1/2))$ such that $\lambda_i = \beta.\lambda'_i.\bar{\beta} \in \Lambda_+$ (as in Cases 3 and 4). Similarly, for each i, H^i_1 is either $\lambda_{i+1} \in \Lambda_-$ (in Cases 1 and 3) or λ'_{i+1}, where $\lambda_{i+1} = \beta.\lambda'_{i+1}.\bar{\beta} \in \Lambda_+$ (in Cases 2 and 4).

We may extend the H^i to essential basic homotopies in (M, S) in the standard way, after which it is clear from the description above that H^i starts on the $-\epsilon_i$- and ends on the ϵ_{i+1}-side. Since H^i_1 agrees with H^{i+1}_0 for each $i < k$, there is a reduced homotopy of length k obtained as the composition of $H^1, \dots, H^k$. □

References

[1] M. Abert and N. Nikolov. Rank gradient, cost of groups, and the rank versus Heegaard genus problem. Preprint. arXiv:math/0701361, May 2008.

[2] M. Boileau and H. Zieschang. Heegaard genus of closed orientable Seifert 3-manifolds. *Invent. Math.*, 76(3):455–468, 1984.

[3] Steve Boyer, Marc Culler, Peter B. Shalen, and Xingru Zhang. Characteristic subsurfaces and Dehn filling. *Trans. Amer. Math. Soc.*, 357(6):2389–2444 (electronic), 2005.

[4] D. Cooper and D. D. Long. Virtually Haken Dehn-filling. *J. Differential Geom.*, 52(1):173–187, 1999.

[5] D. B. A. Epstein. Curves on 2-manifolds and isotopies. *Acta Math.*, 115:83–107, 1966.

[6] Damien Gaboriau. Invariants l^2 de relations d'équivalence et de groupes. *Publ. Math. Inst. Hautes Études Sci.*, (95):93–150, 2002.

[7] Wolfgang Haken. Various aspects of the three-dimensional Poincaré problem. In *Topology of Manifolds (Proc. Inst., Univ. of Georgia, Athens, Ga., 1969)*, pages 140–152. Markham, Chicago, Ill., 1970.

[8] Allen Hatcher. Notes on basic 3-manifold topology. Available for download at: www.math.cornell.edu/∼hatcher/.

[9] Allen Hatcher. *Algebraic topology.* Cambridge University Press, Cambridge, 2002.

[10] John Hempel. *3-manifolds.* AMS Chelsea Publishing, Providence, RI, 2004. Reprint of the 1976 original.

[11] Craig D. Hodgson and Steven P. Kerckhoff. Universal bounds for hyperbolic Dehn surgery. *Ann. of Math. (2)*, 162(1):367–421, 2005.

[12] William H. Jaco and Peter B. Shalen. Seifert fibered spaces in 3-manifolds. *Mem. Amer. Math. Soc.*, 21(220):viii+192, 1979.

[13] Klaus Johannson. *Homotopy equivalences of* 3*-manifolds with boundaries*, volume 761 of *Lecture Notes in Mathematics.* Springer, Berlin, 1979.

[14] Marc Lackenby. Expanders, rank and graphs of groups. *Israel J. Math.*, 146:357–370, 2005.

[15] Marc Lackenby. Heegaard splittings, the virtually Haken conjecture and property (τ). *Invent. Math.*, 164(2):317–359, 2006.

[16] Tao Li. Immersed essential surfaces in hyperbolic 3-manifolds. *Comm. Anal. Geom.*, 10(2):275–290, 2002.

[17] Jennifer Schultens and Richard Weidmann. On the geometric and the algebraic rank of graph manifolds. *Pacific J. Math.*, 231(2):481–510, 2007.

[18] Peter Scott. The geometries of 3-manifolds. *Bull. London Math. Soc.*, 15(5):401–487, 1983.

[19] Peter Scott and Gadde A. Swarup. Regular neighbourhoods and canonical decompositions for groups. *Astérisque*, (289):vi+233, 2003.

[20] Peter Scott and Terry Wall. Topological methods in group theory. In *Homological group theory (Proc. Sympos., Durham, 1977)*, volume 36 of *London Math. Soc. Lecture Note Ser.*, pages 137–203. Cambridge Univ. Press, Cambridge, 1979.

[21] Z. Sela. Endomorphisms of hyperbolic groups. I. The Hopf property. *Topology*, 38(2):301–321, 1999.

[22] Jean-Pierre Serre. *Trees.* Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2003. Translated from the French original by John Stillwell, Corrected 2nd printing of the 1980 English translation.

[23] Juan Souto. Geometry, Heegaard splittings and rank of the fundamental group of hyperbolic 3-manifolds. In *Workshop on Heegaard Splittings*, volume 12 of *Geom. Topol. Monogr.*, pages 351–399. Geom. Topol. Publ., Coventry, 2007.

[24] John Stallings. On fibering certain 3-manifolds. In *Topology of 3-manifolds and related topics (Proc. The Univ. of Georgia Institute, 1961)*, pages 95–100. Prentice-Hall, Englewood Cliffs, N.J., 1962.

[25] William P. Thurston. A norm for the homology of 3-manifolds. *Mem. Amer. Math. Soc.*, 59(339):i–vi and 99–130, 1986.

[26] Marvin D. Tretkoff. A topological approach to the theory of groups acting on trees. *J. Pure Appl. Algebra*, 16(3):323–333, 1980.
[27] Friedhelm Waldhausen. Some problems on 3-manifolds. In *Algebraic and geometric topology (Proc. Sympos. Pure Math., Stanford Univ., Stanford, Calif., 1976), Part 2*, Proc. Sympos. Pure Math., XXXII, pages 313–322. Amer. Math. Soc., Providence, R.I., 1978.
[28] Richard Weidmann. The Nielsen method for groups acting on trees. *Proc. London Math. Soc. (3)*, 85(1):93–118, 2002.

Department of Mathematics, Statistics, and Computer Science (M/C 249), University of Illinois at Chicago, 851 S. Morgan St., Chicago, IL 60607-7045
E-mail address: jdeblois@math.uic.edu