Algebraic & Geometric Topology 16 (2016) 1343–1371

Explicit rank bounds for cyclic covers

JASON DEBLOIS

For a closed, orientable hyperbolic 3–manifold M and an onto homomorphism $\phi\colon \pi_1(M) \to \mathbb{Z}$ that is not induced by a fibration $M \to S^1$, we bound the ranks of the subgroups $\phi^{-1}(n\mathbb{Z})$ for $n \in \mathbb{N}$, below, linearly in n. The key new ingredient is the following result: if M is a closed, orientable hyperbolic 3–manifold and S is a connected, two-sided incompressible surface of genus g that is not a fiber or semifiber, then a reduced homotopy in (M, S) has length at most $14g-12$.

20F05, 57M10; 20E06

The *rank* of a group G, $\operatorname{rk} G$, is the minimal cardinality of a generating set. This paper gives lower bounds on the rank of π_1 among cyclic covers of certain 3–manifolds:

Theorem 0.1 *For a closed, orientable hyperbolic 3–manifold M, a homomorphism $\phi\colon \pi_1 M \twoheadrightarrow \mathbb{Z}$ and an integer $n \geq 2$, let $\Gamma_n = \phi^{-1}(n\mathbb{Z})$. Let $\|\phi\|$ denote the Thurston norm of the cohomology class of ϕ. If ϕ is not induced by a fibration $M \to S^1$, then*

$$\operatorname{rk} \Gamma_n \geq \frac{n-1}{7\|\phi\|+2}.$$

The *Thurston norm* of the cohomology class of ϕ is defined to be the minimum of $\sum_{i=1}^{k} \max\{-\chi(S_i), 0\}$, taken over all surfaces S embedded in M representing the Poincaré dual of ϕ, where the S_i are the components of S. See Thurston [17]. Theorem 0.1 immediately implies the following bound on the *rank gradient* of the pair $(\pi_1 M, \{\Gamma_n\})$, defined by Lackenby [12] as

$$\operatorname{rg}(\pi_1 M, \{\Gamma_n\}) = \liminf_{n\to\infty} (\operatorname{rk} \Gamma_n - 1)/n.$$

Corollary 0.2 *For a closed, orientable hyperbolic 3–manifold M, a homomorphism $\phi\colon \pi_1 M \twoheadrightarrow \mathbb{Z}$ and an integer $n \geq 2$, let $\Gamma_n = \phi^{-1}(n\mathbb{Z})$. Let $\|\phi\|$ denote the Thurston norm of the cohomology class of ϕ. If ϕ is not induced by a fibration $M \to S^1$, then*

$$\operatorname{rg}(\pi_1 M, \{\Gamma_n\}) \geq 1/(7\|\phi\|+2).$$

Published: 1 July 2016 DOI: 10.2140/agt.2016.16.1343

If ϕ is induced by a fibration, then $\operatorname{rk}\pi_1 M_n \leq 2g+1$ for every n, where g is the genus of a connected fiber. Hence $\operatorname{rg}(\pi_1 M, \{\Gamma_n\}) = 0$. In the earlier paper [3], joint with Stefan Friedl and Stefano Vidussi, we proved a weaker analog of Corollary 0.2 for a broader class of 3–manifolds: compact, orientable and connected with toroidal or empty boundary. For such M and $\phi\colon \pi_1 M \twoheadrightarrow \mathbb{Z}$, Theorem 1.1 in that paper implies that $\operatorname{rg}(\pi_1 M, \{\Gamma_n\}) > 0$ if ϕ is not induced by a fibration.

The strategy of the proof of Theorem 1.1 in DeBlois, Friedl and Vidussi [3] is to find a finite cover $p\colon M' \to M$ inheriting a map $\phi'\colon \pi_1 M' \twoheadrightarrow \mathbb{Z}$ so that, for homological reasons, $\operatorname{rg}(\pi_1 M', \{(\phi')^{-1}(n\mathbb{Z})\}) > 0$, whence $\operatorname{rg}(\pi_1 M, \{\Gamma_n\}) > 0$ as well. The "virtually special" machine (see Haglund and Wise [6] and Wise [21; 22]) produces p, and controlling its degree seems out of reach at present. Producing an explicit bound thus requires a different strategy. Our approach here, outlined in Section 1, instead follows that of [3, Section 3]. We use:

Corollary 2.2 *For a closed, orientable hyperbolic* 3*–manifold* M *and a connected, two-sided incompressible surface* $S \subset M$ *of genus* g *that is not a fiber or semifiber, the* $\pi_1 M$*–action on the tree determined by* S *is* $(14g-12)$*–acylindrical.*

Combining this with an acylindrical accessibility theorem of R Weidmann [20] immediately gives Theorem 0.1. The action at issue above is described by Bass–Serre theory, see, eg Scott and Wall [14]. A connected surface $S \subset M$ is a *semifiber* if it separates M into a disjoint union of twisted I–bundles over the nonorientable surface double covered by S. If S is a semifiber, then there is a twofold cover $\widetilde{M} \to M$ such that S lifts to a fiber of a fibration $\widetilde{M} \to S^1$. It is necessary in Corollary 2.2 that S not be a fiber or semifiber; otherwise each element of $\pi_1 S < \pi_1 M$ fixes the entire tree, so the action is k–cylindrical for all $k \geq 0$.

Corollary 2.2 in turn follows from Theorem 4.1 below, whose proof contains the main substantive work of the paper. It is an extension of the so-called "veg-o-matic" argument which has seen prior use in the works of Cooper and Long [2, Section 4], Li [13, Section 2], Walsh [19], and Boyer, Culler, Shalen, and Zhang [1, Theorem 5.4.1].

Theorem 4.1 *For a closed, orientable hyperbolic* 3*–manifold* M *and a connected, two-sided incompressible surface* $S \subset M$ *of genus* g *that is not a fiber or semifiber, a nondegenerate, reduced homotopy in* (M, S) *has length at most* $14g-12$.

Above, a homotopy in (M, S) is a map of pairs $H\colon (K \times I, K \times \partial I) \to (M, S)$, for a topological space K. It is *reduced of length* k if it is obtained by chaining together homotopies $H^1, \dots, H^k$ such that H^i is essential and $(H^i)^{-1}(S) = K \times \partial I$ for each i, and H^{i+1} starts on the opposite side of S from which H^i ends for $i < k$. (See also Definition 2.4.) An observation of Z Sela [15] draws the connection between homotopies through M of curves in S and cylinders of the $\pi_1 M$–action on the tree of S. In Section 2, we reproduce this observation as Lemma 2.1. With Theorem 4.1, it immediately implies Corollary 2.2.

Section 3 gives some results on intersections of surfaces that we use in Section 4 to prove Theorem 4.1. This argument has two main steps. The first step identifies a sequence $\Psi_1 \supset \Psi_2 \supset \cdots$ of subsurfaces of S, of minimal complexity, with the property that for each k, a reduced homotopy H with length k and target (M, S) has H_0 homotopic into Ψ_k in S. The primary technical tool in this step is the characteristic submanifold of the manifold obtained by cutting M along S.

The second step uses the fact that M is hyperbolic and S is not a fiber or semifiber to show that Ψ_k is not homotopic into Ψ_{k+2} in S as long as $\Psi_k \neq \varnothing$. Therefore eventually $\Psi_k = \varnothing$, and homotopies expire in finite time.

For various reasons, previous versions of this argument do not require accounting for solid torus components of the characteristic submanifold. However, homotopies through M of curves in S may indeed pass through such solid tori. The difficulty in extending the standard argument to accommodate this is that the time-0 map of a homotopy through such a component may not determine the time-1 map.

We sidestep this issue, producing the Ψ_k by adding judiciously chosen annuli to a sequence $\{\Phi_k\}$ of subsurfaces of S, identified in Boyer, Culler, Shalen, and Zhang [1], that carry time-0 maps of "large" homotopies (see Definition 3.1) with target (M, S). Indeed, many of the results of Sections 3 and 4 rely on and directly extend work in [1]. We indicate when this is so and cross-reference precisely.

Acknowledgements

Many thanks to Peter Shalen for explaining the argument of Boyer, Culler, Shalen, and Zhang [1] to me, and for helping me extend it to the present context. Thanks also to Peter Scott and the anonymous referee for suggestions that have improved the paper. This paper was originally written during a period of partial NSF support.

1 Proof of the main theorem

The proof of Theorem 0.1 closely follows the proof of [3, Theorem 3.4]. We will sketch it below, at times referring to [3] for details. But first we recall the definition of an acylindrical action and reproduce an "acylindrical accessibility" theorem of R Weidmann.

Definition 1.1 [15] An action $\Gamma \times T \to T$ is *k–acylindrical* if no $g \in \Gamma - \{1\}$ fixes a segment of length greater than k, and *k–cylindrical* otherwise.

Theorem (Weidmann [20]) *Let Γ be a noncyclic, freely indecomposable, finitely generated group and $\Gamma \times T \to T$ a minimal k–acylindrical action. Then $\Gamma\backslash T$ has at most $1 + 2k(\operatorname{rk}\Gamma - 1)$ vertices.*

Assuming Corollary 2.2, we now sketch the proof of Theorem 0.1.

Proof sketch, Theorem 0.1 For a closed hyperbolic 3–manifold M and an onto homomorphism $\phi\colon \pi_1 M \to \mathbb{Z}$, standard arguments produce a closed, oriented surface S embedded in M that is *dual* to ϕ in the sense that $\phi = p_*$ for the map

$$p\colon M \to S^1 = [-1,1]/(-1 \sim 1)$$

defined as follows: for a tubular neighborhood $\mathcal{N} = S \times [-1,1]$ of S in M and for $(x,t) \in S \times [-1,1]$, let $p(x,t) = t$, and let $p(x) = -1 \sim 1$ for each $x \in M - \mathcal{N}$.

There is a π_1–surjective map $q\colon M \to G_0$, where G_0 is a graph with one vertex for each component of $M - (S \times (-1,1))$ and one edge for each component of $S \times [-1,1]$ (with the obvious attaching maps), such that p factors through q. If $\chi(G_0) < 0$, then for each $n \geq 2$, $\operatorname{rk}\pi_1 M_n \geq -n\chi(G_0) + 1$. This follows from the fact that M_n π_1–surjects an n–fold cover of G_0, the motivating observation for [3, Lemma 3.3] (see also [3, Lemma 2.6]).

By the above, the desired bound on rank holds if $\chi(G_0) < 0$, so we may assume $\chi(G_0) = 0$. Assuming that S has minimal complexity among all surfaces dual to ϕ, it follows that G_0 has one vertex and one edge, ie S is connected and nonseparating. This assertion is proved in the final two paragraphs of the proof of [3, Theorem 3.4]. Here the *complexity* of $S = S_1 \sqcup \cdots \sqcup S_k$, where each S_i is connected, is defined as $\chi_- S = \sum_{i=1}^{k} \max\{-\chi(S_i), 0\}$. The Thurston norm $\|\phi\|$ of ϕ is by definition equal to $\chi_- S$ for S dual to ϕ with minimal complexity.

G_0 is the underlying graph of a *graph of spaces* decomposition of M in the sense of [14, page 155], with vertex space $X = \overline{M - (S \times [-1,1])}$ and edge space $S \times [-1,1]$. There is an associated action of $\pi_1 M$ on a tree T, without inversions, such that each vertex stabilizer is conjugate to $\pi_1(X)$ and each edge stabilizer to $\pi_1 S_0$ for some component S_0 of S. See [14, pages 166–167], also [16] and [18]. This is what we call *the action on the $\pi_1 M$-tree determined by S*.

We now apply the hypothesis that ϕ is not induced by a fibration $M \to S^1$. Then S is not the fiber of a fibration $M \to S^1$ (if it were, then p would be homotopic to a fibration), and since it is nonseparating, it is not a semifiber. Corollary 2.2 therefore asserts that the $\pi_1 M$–action on T is $(14g-12)$–acylindrical, where g is the genus of S. This property is inherited by each subgroup $\Gamma_n = \phi^{-1}(n\mathbb{Z}) < \pi_1 M$. By construction, the graph $\Gamma_n \backslash T$ has n vertices and edges, so the result follows directly from Weidmann's theorem upon noting that $\|\phi\| = 2g-2$. □

2 Cylinders and homotopies

We reproduce example (iv) of Z Sela's introduction to [15] below:

> Let S be an incompressible surface in a compact 3–manifold M. Let M' denote the 3–manifold obtained by cutting M along S. A homotopy H [in M] between two closed curves in S can be decomposed into essential homotopies in M'. The number of these essential subhomotopies is called the *length* of H. An incompressible surface is called k*–acylindrical* if no homotopy between closed curves in S has length bigger than k. To an incompressible surface S in M corresponds a splitting of $\pi_1 M$. The bound on the length of a homotopy between curves on S corresponds exactly to the dual splitting being $(k+1)$–acylindrical.

The purpose of this section is to expand on Sela's remarks, define his terms, and give a reasonably detailed sketch proof of the assertion of his final sentence in our case.

Lemma 2.1 *Let M be a closed, irreducible 3–manifold and $S \subset M$ a closed, connected, two-sided incompressible surface. For $k > 1$, the action $\pi_1 M \times T \to T$ on the tree T determined by S is k–cylindrical if and only if there is a nondegenerate reduced homotopy $(S^1 \times I, S^1 \times \partial I) \to (M, S)$ of length k.*

A surface S as above is *incompressible* if it is embedded in M with a π_1–injective inclusion map, and it is not a two-sphere that bounds a ball in M. See, eg [8, Chapter 6]. We prove Lemma 2.1 at the end of this section, but first note that combining it with Theorem 4.1 immediately yields:

Corollary 2.2 *For a closed, orientable hyperbolic 3–manifold M and a connected, two-sided incompressible surface $S \subset M$ of genus g that is not a fiber or semifiber, the $\pi_1 M$–action on the tree determined by S is $(14g-12)$–acylindrical.*

Definition 2.3 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* is a map $H_t\colon X \to Y$ 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 common domain X and common target Y. A homotopy H with domain X and target Y is the *composition of* $H^1, \dots, H^n$ if there exist numbers $0 = t_0 < t_1 < \cdots < t_n = 1$ and monotone increasing linear homeomorphisms $\alpha_i\colon [t_{i-1}, t_i] \to [0,1]$ such that $H(x,t) = H^i(x, \alpha_i(t))$ for all $x \in X$ and $t \in [t_{i-1}, t_i]$.

A path $\gamma\colon I \to Y$ may be regarded as a homotopy with domain $X = \varnothing$. We will denote the composition of paths $\gamma_1, \gamma_2, \dots, \gamma_n$, as defined above, by $\gamma_1.\gamma_2.\cdots.\gamma_n$.

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 $W \hookrightarrow X$. 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. We have borrowed their precise formulations from [1]. This is also our source for the definitions below that apply to 3–manifolds.

Definition 2.4 Let M be a closed 3–manifold, and let $S \subset M$ be an embedded, transversely oriented surface. A *homotopy in* (M,S) *with domain* K is a homotopy H with domain K and target M such that $H(K \times \partial I) \subset S$. It is *nondegenerate* if $H_*(\pi_1 K) \neq \{1\}$, and *basic* if $H^{-1}(S) = K \times \partial I$.

For $\epsilon \in \{+,-\}$, we say a basic homotopy *starts* (respectively, *ends*) *on the* ϵ*–side* if $H(K \times [0,\delta]) \subset \mathcal{N}_\epsilon$ (respectively, if $H(K \times [1-\delta, 1]) \subset \mathcal{N}_\epsilon$). Here $\mathcal{N} \cong S \times [-1,1] \subset M$ is a closed regular neighborhood of S, embedded so that $S = S \times \{0\}$ and the standard transverse orientation is preserved; $\mathcal{N}_+ = S \times [0,1]$; and $\mathcal{N}_- = S \times [-1,0]$.

We say that $X = M - (S \times (-1,1))$ is *obtained by cutting* M *along* S. If H is a basic homotopy in (M,S) with domain K, then after straightening in $\mathcal{N}$ 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 $(K \times I, K \times \partial I) \to (X, \partial X)$, ie π_1–injective and not properly homotopic into ∂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 $1 \leq i \leq k$ such that H is the composition of $H^1, \dots, H^k$, and for each $i < k$, H^i starts on the ϵ_i–side and ends on the $-\epsilon_{i+1}$–side, and H^k starts on the ϵ_k–side.

A connected, incompressible surface S in a closed 3–manifold M determines a *graph of spaces decomposition* of M whose underlying graph G has a single edge, corresponding to S, and (one or two) vertices corresponding to the components of the manifold X obtained by cutting M along S. By Bass–Serre theory, this determines an action of $\pi_1 M$ on a tree T, without inversions and with quotient graph G. We will use the following basic consequence of this set-up.

Lemma 2.5 *Suppose a group Π acts on a tree T, transitively on edges and without inversions. Let $\{e_0,\dots,e_k\}$ be a segment of T of length $k+1$, so $e_i \neq e_{i-1}$ but e_i and e_{i-1} share an endpoint v_i for each $i>0$, and let $\Lambda = \mathrm{Stab}_\Pi(e_0)$.*

There are two cases to consider:

Case S *If $G = \Pi\backslash T$ has two vertices, let $\Gamma_- = \mathrm{Stab}_\Pi(v_0)$ and $\Gamma_+ = \mathrm{Stab}_\Pi(v_1)$, where $v_0 \neq v_1$ is an endpoint of e_0 (recall from above that $v_1 = e_1 \cap e_0$). Then for each $i \in \{1,\dots,k\}$, there exists γ_i, in $\Gamma_- - \Lambda$ for i even and in $\Gamma_+ - \Lambda$ for i odd, such that for each $j \le k$, the element $\delta_j = \gamma_1\gamma_2\cdots\gamma_j$ takes e_0 to e_j.*

Case N *If G has a single vertex, let $\Gamma = \mathrm{Stab}_\Pi(v_0)$, let $\Lambda_+ < \Gamma$ stabilize an edge e' containing v_0 but not Γ–equivalent to e_0, and fix $\tau \in \Pi$ with $\tau(e') = e_0$. Orient the edge of G so that e_0 points toward v_1 in the inherited orientation on T, and for $0 < i \le k$, let $\epsilon_i = 1$ if e_i points from v_i to v_{i+1} and let $\epsilon_i = -1$ otherwise. For each $i \in \{1,\dots,k\}$, there exists $\gamma_i \in \Gamma$ so that for $1 \le j \le k$,*

$$\delta_j = \tau(\gamma_1\tau^{\epsilon_1})\cdots(\gamma_{j-1}\tau^{\epsilon_{j-1}})$$

has the property that

$$\delta_j(v_0) = v_j \quad \text{and} \quad e_j = \begin{cases} \delta_j\gamma_j(e_0) & \text{if } \epsilon_j = 1,\\ \delta_j\gamma_j(e') & \text{if } \epsilon_j = -1.\end{cases}$$

Let $\epsilon_0 = 1$. For $i \ge 1$, if $\epsilon_{i-1} \neq \epsilon_i$, then γ_{i+1} is not in an edge stabilizer.

Proof In case **S**, T has two Π–orbits of vertices, and the stabilizer of any vertex v acts transitively on the edges containing it. This is because on a small neighborhood U of v in T, the projection to $\Pi\backslash T$ factors through an embedding of $\mathrm{Stab}_\Pi(v)\backslash U$. This case, which we leave to the reader, is a straightforward induction argument.

With notation as described in case **N**, there are two Γ–orbits of edges of T containing v_0, one pointing toward v_0 and one away. In particular, e' points toward v_0, and $\tau(v_0) = v_1$. We therefore take $\delta_1 = \tau$. Then $\delta_1^{-1}(e_1)$ contains v_0, so depending on orientation it is Γ–equivalent to one of e' or e_0. If e_1 points toward v_1, then $\gamma_1^{-1}\delta_1^{-1}(e_1) = e'$ for some $\gamma_1 \in \Gamma$; otherwise there exists $\gamma_1 \in \Gamma$ with $\gamma_1^{-1}\delta_1^{-1}(e_1) = e_0$. This proves the base case of an induction argument.

For the inductive step of the argument, we take $j > 1$ and suppose that we have identified δ_{j-1} and γ_{j-1} satisfying the required properties. If $\epsilon_{j-1} = 1$, then $\gamma_{j-1}^{-1}\delta_{j-1}^{-1}(v_j) = v_1$, so $\tau^{-1}\gamma_{j-1}^{-1}\delta_{j-1}^{-1}(v_j) = v_0$. If $\epsilon_{j-1} = -1$, then $\tau\gamma_{j-1}^{-1}\delta_{j-1}^{-1}(v_j) = v_0$. Therefore $\delta_j(v_0) = v_j$. Since $\delta_j^{-1}(e_j)$ thus contains v_0, arguing as in the base case we identify γ_j in Γ so that $\delta_j\gamma_j$ takes e_0 or e' (depending on ϵ_j) to e_j.

For the lemma's final assertion, we note that $\epsilon_{i-1} \neq \epsilon_i$ implies that e_{i-1} and e_i either both point toward or both point away from v_i, so $\delta_i^{-1}(e_{i-1})$ and $\delta_i^{-1}(e_i)$ are distinct Γ–translates. A definition-chase shows the translating element is γ_i. □

To the graph of spaces decomposition determined by an incompressible surface S, there corresponds a "graph of groups" decomposition of the fundamental group of M with underlying graph G. We record this in the standard lemma below, a paraphrase of [14, page 155].

Here for a closed path γ based at a point $x \in M$, we will also denote its based homotopy class in $\pi_1(M,x)$ by γ, letting context determine the proper interpretation, and we let $\alpha.\beta$ denote the composition of paths α and β, defined as in Definition 2.3.

Lemma 2.6 *For a closed 3–manifold M and a connected, transversely oriented incompressible surface S with closed regular neighborhood $\mathcal{N} \cong S\times[-1,1] \subset M$, let $S_\pm = S\times\{\pm 1\}$ and $X = M - (S\times(-1,1))$. Fix $x \in S$, take $x_\pm = (x,\pm 1) \in S_\pm$, let $\Lambda = \pi_1(S_-,x_-)$, and let $\alpha\colon t \mapsto (x, 2t-1)$ join x_- to x_+ in $\mathcal{N}$.*

There are two cases to consider:

Case S *If S is separating, then $\pi_1 M$ is a* free product with amalgamation*:*

$$\pi_1(M,x_-) \cong \Gamma_- *_\Lambda \Gamma_+ \doteq \langle \Gamma_-, \Gamma_+ \mid \lambda = \alpha.\phi_*(\lambda).\bar{\alpha},\ \lambda\in\Lambda\rangle.$$

Here $\Gamma_- = \pi_1(X_-,x_-)$, where X_- is the component of X with $S_- = \partial X_-$, and $\Gamma_+ = \{\alpha.\gamma.\bar{\alpha} \mid \gamma \in \pi_1(X_+,x_+)\}$ for X_+ with $S_+ = \partial X_+$.

Case N *If S is nonseparating, then $\pi_1 M$ is an* HNN extension *of $\Gamma = \pi_1(X,x_-)$:*

$$\pi_1(M,x_-) \cong \Gamma *_\Lambda \doteq \langle\, \Gamma, \tau \mid \tau^{-1}\lambda\tau = \bar{\beta}.\phi_*(\lambda).\beta,\ \lambda\in\Lambda\rangle. \tag{1}$$

Here $\tau \in \pi_1(M,x_-)$ is the pointed homotopy class of $\alpha.\beta$ for some fixed arc β in X joining x_+ to x_-.

In each case above, $\phi\colon S_- \to S_+$ takes $(x,-1)$ to $(x,1)$ for all $x\in S$, so ϕ_* is an isomorphism $\Lambda \to \pi_1(S_+,x_+)$.

Proof of Lemma 2.1 First suppose there is a nondegenerate, reduced homotopy $H\colon (S^1\times I, S^1\times\partial I) \to (M,S)$ of length k. Writing H as a composition of essential basic homotopies $H^1,\dots,H^k$, we may assume without loss of generality that $H^{-1}(S) = \bigsqcup_{i=0}^{k} S^{-1} \times \{i/k\}$ and for each $i > 0$, H^i linearly reparametrizes $H|_{S^1\times[(i-1)/k,i/k]}$. We may further assume that H is *vertical* with respect to a closed regular neighborhood $\mathcal{N} \cong S\times[-1,1]$ of S in M, by which we mean that each H^i is obtained from its restriction to $(H^i)^{-1}(X)$ by collaring, where $X = M - S\times(-1,1)$.

Let $\hat{H}^i\colon S^1 \times I \to X$ be obtained by reparametrizing the restriction of H^i to the preimage of X. Fix a base point $x \in S$ and for each i fix a path ρ^i in S from x to $H(1,i/k)$. Taking $S_\pm = S\times\{\pm 1\}$, let $x_\pm = (x,\pm 1) \in S_\pm$, and let $\rho^i_{\pm 1}$ be the path parallel to ρ^i in $S_\pm$.

Assume for now that S is separating. Then each H^i starts and ends on the same side of S, so since H is reduced, the H^i alternate sides. We will assume that H^i starts and ends on the $+$–side for odd i and the $-$–side for even i (the argument in the other case is completely analogous). Thus $\hat{H}^i$ maps into X_- for i odd and X_+ for i even, where $X_\pm$ is the component of X with $S_\pm = \partial X_\pm$. For $1 \le i \le k$, define

$$\gamma_i = \begin{cases} \rho_-^{i-1}.(t \mapsto \hat{H}^i(1,t)).\bar{\rho}^i_- & \text{for } i \text{ even,} \\ \alpha.\rho_+^{i-1}.(t \mapsto \hat{H}^i(1,t)).\bar{\rho}^i_+.\bar{\alpha} & \text{for } i \text{ odd.} \end{cases}$$

Here α is as described as in Lemma 2.6. For $\Gamma_\pm$ as described there, it follows by construction that $\gamma_i \in \Gamma_+$ if i is odd and $\gamma_i \in \Gamma_-$ otherwise.

We claim that $\gamma_i \notin \Lambda$, for all i. If $\gamma_i \in \Lambda$, then $H^i|_{\{1\}\times I} \subset S$ after a homotopy of H, so there is a map $(D,\partial D) \to (M,S)$ factoring through H for a disk D. Since S is incompressible, ∂D bounds a disk $D' \subset S$. The sphere theorem and the irreducibility of M imply that $\pi_2(M) = 0$, so $D \cup D' \to M$ extends over a ball. It follows that H is not essential, contradicting our hypotheses.

For each $i \in \{0,\dots,k\}$, one obtains a loop in M based at $H(1,0)$ by applying H to the concatenation of the straight-line path in $S^1 \times I$ joining $(1,0)$ to $(1,i/k)$ with the loop around $S^1 \times \{i/k\}$, followed by the straight-line path back to $(1,0)$. After connecting the base point x_- to $H(1,0)$ using ρ^0_- and a vertical arc, these loops all evidently represent the same element g of $\pi_1(M,x_-)$. A short induction argument shows that $\delta_i^{-1} g \delta_i \in \Lambda$ for all i, where $\delta_i = \gamma_1 \cdots \gamma_i$.

As we remarked directly before Lemma 2.5, by Bass–Serre theory, S determines an action $\pi_1 M \times T \to T$ on a tree T, without inversions and with quotient graph G. Under this action, the stabilizer of each edge is a conjugate of the edge group Λ of G, and the stabilizer of each vertex is conjugate to a vertex group of G, in this case, one of $\Gamma_\pm$. See [14, pages 166–167].

Let e_0 be the edge of T stabilized by Λ, and let v_0 and v_1 be the endpoints of e stabilized, respectively, by Γ_- and Γ_+. Then g is in Λ and by construction also in $\delta_i \Lambda \delta_i^{-1}$ for each i, stabilizing $e_i = \delta_i(e)$. These determine a path in T since $\gamma_i = \delta_{i-1}^{-1}\delta_i$ is in one of Γ_+ or Γ_- for each i. This path has length $k+1$ because $\gamma_i \notin \Lambda$, so $e_i \neq e_{i-1}$, for any i.

The separating case of the "if" direction of Lemma 2.1 is established. Note that the elements γ_i and δ_i above match the descriptions in case **S** of Lemma 2.5.

Suppose now that S is nonseparating, so that X is connected and has two boundary components $S_\pm$. Given a nondegenerate homotopy H of length k, decomposed into $H^1, \dots, H^k$ as previously, there are four possibilities for each of the H^i. If H^i starts and ends on the $-$–side, we define γ_i as for H^i in the separating case for i even, and if it starts and ends on the $+$–side, we define as for i odd. Otherwise

$$\gamma_i = \begin{cases} \rho_-^{i-1}.(t \mapsto \hat{H}^i(1,t)).\bar{\rho}_+^i.\beta & \text{if } H^i \text{ starts on the } +\text{, ends on the } - \text{ side,} \\ \alpha.\rho_+^{i-1}.(t \mapsto \hat{H}^i(1,t)).\bar{\rho}_-^i.\bar{\beta}.\bar{\alpha} & \text{if } H^i \text{ starts on the } -\text{, ends on the } + \text{ side.} \end{cases}$$

Here β is as described in case **N** of Lemma 2.6. If H^i starts and ends on the same side, then arguing as in the separating case shows γ_i is not in an edge stabilizer. We produce a path in T by a process similar to the separating case, using words δ_j which in this case match the description in case **N** of Lemma 2.5 (for τ as described in Lemma 2.6). The details of this case track those of the parallel case of the reverse implication, described below.

We now address the reverse implication of the lemma, proving that a nontrivial element g stabilizing a length-$(k+1)$ segment in T gives rise to a length-k reduced homotopy in (M,S). The idea of the proof is to use the description of Lemma 2.5 to reverse-engineer the construction above. We leave the separating case of this construction to the reader (it is simpler) and move directly to the case that S is nonseparating. The four different boundary behaviors of basic homotopies in this case correspond to the possible orientations on edges meeting at a vertex.

To make this precise let us fix some notation. For Γ, Λ and τ defined as in case **N** of Lemma 2.6, Γ stabilizes a vertex v_0 of T and $\Lambda < \Gamma$ stabilizes an edge e_0 containing v_0. It further follows from (1) above that $e' \doteq \tau^{-1}(e_0)$ contains v_0 since $\Lambda_+ \doteq \tau^{-1}\Lambda\tau < \Gamma$.

Suppose now that the $\pi_1 M$–action is k–cylindrical, so there exists $g \in \pi_1 M - \{1\}$ fixing a segment of length at least $k+1$. By transitivity, upon replacing g by a conjugate we may assume v_0 is the segment's initial vertex. Since X has two boundary components, each edge containing v_0 is a Γ–translate of exactly one of e_0 or e'. Thus conjugating g further in Γ, we may assume the segment's initial edge is either e' or e_0. If it is e_0, we apply Lemma 2.5; if it is e', we exchange Λ and Λ_+, replace τ by τ^{-1}, rename e' to e_0 and vice-versa, then apply case **N** of Lemma 2.5.

For each $j \geq 0$, since g stabilizes e_j, the lemma implies that $g = (\delta_j\gamma_j)\lambda_j(\delta_j\gamma_j)^{-1}$ for some λ_j, which is in Λ if $\epsilon_j = 1$ and Λ_+ otherwise. Since $\delta_j = \delta_{j-1}(\gamma_{j-1}\tau^{\epsilon_{j-1}})$, comparing the resulting descriptions of g at e_{j-1} and e_j, for $j > 0$, yields

$$\tau^{-\epsilon_{j-1}}\lambda_{j-1}\tau^{\epsilon_{j-1}} = \gamma_j\lambda_j\gamma_j^{-1}. \tag{2}$$

For each j such that $\epsilon_j = 1$, fix a closed curve $\mathfrak{c}_j$ on S_- through x_- that represents λ_j. If $\epsilon_j = -1$, then since $\Lambda_+ = \tau^{-1}\Lambda\tau$, we have that $\lambda_j \in \Lambda_+$ is, for some $\lambda_j^{(0)} \in \Lambda$, of the form $\tau^{-1}\lambda_j^{(0)}\tau = \overline{\beta}.\phi_*(\lambda_j^{(0)}).\beta$. In this case, let $\mathfrak{c}_j$ be a closed curve on S_+ that represents $\phi_*(\lambda_j^{(0)}) \in \pi_1(S_+, x_+)$.

For each $j > 0$, equation (2) above determines a homotopy in X either from $\phi(\mathfrak{c}_{j-1})$ to $\mathfrak{c}_j$ (if $\epsilon_{j-1} = 1$) or from $\phi^{-1}(\mathfrak{c}_{j-1})$ to $\mathfrak{c}_j$ (if $\epsilon_{j-1} = -1$). One produces from this a basic homotopy H^j in (M, S) by adjoining product collars in the obvious way. By construction, H^{j+1} starts on the opposite side of S from H^j for each $j < k$. To show that the composition of $H^1, \dots, H^k$ is reduced of length k, it remains only to show that each H^j is essential.

This is clear when H^j starts and ends on opposite components of ∂X, so let us consider a case where it does not. If $\epsilon_{j-1} = 1$ and $\epsilon_j = -1$, then $\phi(\mathfrak{c}_{j-1}) = H^j(S^1 \times \{0\})$ and $\mathfrak{c}_j = H^j(S^1 \times \{1\})$ each lie in S_+. Equation (2) becomes

$$\overline{\beta}.\phi_*(\lambda_{j-1}).\beta = \gamma_j(\overline{\beta}.\phi_*(\lambda_j^{(0)}).\beta)\gamma_j^{-1},$$

and H^j is a concatenation of four homotopies, the free homotopy from $\phi_*(\lambda_{j-1})$ to $\overline{\beta}.\phi_*(\lambda_{j-1}).\beta$, the pointed homotopy between left and right sides of (2), the free homotopy from $\gamma_j(\overline{\beta}.\phi_*(\lambda_j^{(0)}).\beta)\gamma_j$, and finally the free homotopy to $\phi_*(\lambda_j^{(0)})$.

If there were a proper homotopy of H^j into S_+, it would follow that $\gamma_j \in \Lambda_+$, contradicting the final assertion of Lemma 2.5. The case $\epsilon_{j-1} = -1$ and $\epsilon_j = 1$ is similar. □

3 Essential surfaces and essential intersections

Now we shift gears to extend the theory of "essential intersection" for subsurfaces of a 2–manifold that is introduced in [1, Section 4]. There it is remarked that this notion "has appeared implicitly in much of the literature on the characteristic submanifold of a Haken manifold". The results of [1] are proved for *large* subsurfaces (see below); we must allow annular components as well. Many results extend directly to this context using similar proof strategies, but 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. We also use "annulus" interchangeably with "cylinder" to refer to $S^1 \times I$.

Definition 3.1 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 $A \hookrightarrow S$ is π_1–injective. A component A_0 of an incompressible subsurface A is *redundant* if its inclusion map 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 core circles of the remaining annular components as the *small part* $A_{\mathcal{S}}$ of A. (Note that $A_{\mathcal{L}} \cup A_{\mathcal{S}}$ is properly contained in 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, then each redundant component of A is homeomorphic to an annulus.

Suppose A_0 is such a component, whose inclusion map is homotopic in S into another component A_1. We may assume A_1 lies in the interior $\operatorname{int} S$ of S, after pushing off the boundary. Choosing a basepoint in A_1, let $\widetilde{S} \to \operatorname{int} S$ be the cover corresponding to $\pi_1 A_1$. The inclusion $A_1 \hookrightarrow S$ lifts to an embedding to a subsurface $\widetilde{A}_1 \subset \widetilde{S}$ that carries $\pi_1 \widetilde{S}$. Therefore each component of $\widetilde{S} - \operatorname{int} \widetilde{A}_1$ is homeomorphic to a half-open annulus. Since the inclusion map of A_0 is homotopic into A_1, it lifts to an embedding in $\widetilde{S}$. The inclusion map $A_0 \hookrightarrow S$ is π_1–injective by hypothesis, so its lift is too, and the lift's image does not intersect $\widetilde{A}_1$. 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 [1, Lemma 4.1].

Lemma 3.2 *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 [1, 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. Analogous to [1, Lemma 4.1], we choose this object within its isotopy class so that $\partial A_{\mathcal{L}} \cup A_{\mathcal{S}}$ meets ∂B transversely in the minimal number of points possible, and, among all intersection-minimizing representatives, to minimize the number of components of $\partial A_{\mathcal{L}} \cup A_{\mathcal{S}}$ not contained in B.

Given a component A_0 of $A_{\mathcal{L}}$ that is homotopic into a component B_0 of B, the proof of [1, Lemma 4.1(1)] again shows here that, with our assumptions, $A_0 \subset B_0$. We must simply replace instances of ∂A by $\partial A_{\mathcal{L}} \cup A_{\mathcal{S}}$ in the paragraph spanning pages 2405–2406 there and its sequel. Pushing off ∂B_0 for each such component B_0, we will assume $A_{\mathcal{L}}$ is contained in the interior of B.

Now suppose $\mathfrak{a}_0$ is a component of $A_{\mathcal{S}}$ and let B_0 be the component of B into which it is homotopic. We again follow the proof of [1, Lemma 4.1]: Fixing a base point in B_0, let $p\colon \widetilde{S} \to \operatorname{int} S$ be the cover corresponding to $\pi_1 B_0$. The inclusion map $B_0 \hookrightarrow S$ lifts to an embedding to a component $\widetilde{B}_0$ of $p^{-1}(B_0) \subset \widetilde{S}$, and because $\mathfrak{a}_0$ is homotopic into B_0, it too lifts to a simple closed curve $\widetilde{\mathfrak{a}}_0$ in $\widetilde{S}$. Since B_0 is incompressible, the inclusion $\widetilde{B}_0 \hookrightarrow \widetilde{S}$ induces an isomorphism at the level of π_1, and so each component of $X = \widetilde{S} - \operatorname{int} \widetilde{B}_0$ is homeomorphic to a half-open annulus.

If $\widetilde{\mathfrak{a}}_0$ meets $\partial \widetilde{B}_0$, then the argument of the paragraph that spans pages 2405–2406 in [1] and its sequel again yields a contradiction to our minimality assumption (after the same adjustment as before). Therefore $\widetilde{\mathfrak{a}}_0$ is disjoint from $\partial \widetilde{B}_0$, and if $\widetilde{\mathfrak{a}}_0$ is not contained in $\widetilde{B}_0$, then it is contained in an annular component Z of X. (Unlike in the proof of [1, Lemma 4.1] this can occur, since $\pi_1 \mathfrak{a}_0 \cong \mathbb{Z}$.)

Since $\widetilde{\mathfrak{a}}_0$ is a homotopically nontrivial simple closed curve in Z, it cobounds an annulus with the component $\widetilde{\mathfrak{b}}_0$ of $\partial \widetilde{B}_0$ that bounds Z. This annulus projects to a free homotopy between $\mathfrak{a}_0$ and $\mathfrak{b}_0 = p(\widetilde{\mathfrak{b}}_0)$, a component of ∂B_0. Theorem 2.1 of [4] now implies that $\mathfrak{a}_0$ is isotopic to $\mathfrak{b}_0$ and hence, pushing a bit further, isotopic into the interior of B_0. This isotopy may be taken to be supported in a small enough neighborhood of the annulus bounded by $\mathfrak{a}_0$ and $\mathfrak{b}_0$ that it leaves invariant all components of $A_{\mathcal{L}} \cup A_{\mathcal{S}}$ inside B_0, and all components of $A_{\mathcal{S}}$ outside the annulus. After a finite sequence of such isotopies we have $A_{\mathcal{L}} \cup A_{\mathcal{S}} \subset B$.

To complete the proof of (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:

(a) The ϵ–neighborhood $\mathcal{N}_\epsilon(\mathfrak{a})$ is regular and contained in the component A_0 of A containing $\mathfrak{a}$.

(b) 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 $A_{\mathcal{L}} \cup A_{\mathcal{S}}$.

(c) 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 $A \subset \operatorname{int} B$. In particular, $A_{\mathcal{L}} \subset \operatorname{int} B_{\mathcal{L}}$. 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 [1, page 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 subannulus. 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). $\square$

The following proposition extends [1, Proposition 4.2]. Below we reference the "large intersection" $A \wedge_{\mathcal{L}} B$ of large surfaces A and B from [1, Definition 4.3].

Proposition 3.3 *Suppose A and B are irredundant subsurfaces of an orientable compact surface S with no 2–sphere or torus components. Then up to nonambient 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 for a polyhedron K and a map $f\colon K \to S$ such that $f_*(\pi_1 K_0) \neq 1$ for each component K_0 of K, f is homotopic into each of A and B if and only if f is homotopic into C.*

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

Definition 3.4 If A and B are irredundant subsurfaces of an orientable compact surface S, we say an irredundant surface C that satisfies condition ($*$) of Proposition 3.3 *represents the essential intersection $A \cap_{\mathrm{ess}} B$ of A and B.*

Proposition 3.3 implies in particular that each of A and B contains a subsurface that represents $A \cap_{\mathrm{ess}} B$, and that these subsurfaces are isotopic in S.

Proof of Proposition 3.3 We assume without loss of generality that $A, B \subset \operatorname{int} S$. If C and C' are surfaces with property ($*$), then C is homotopic into C' and vice-versa. Hence Lemma 3.2(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 [1], $C_0 = \mathcal{L}(A \cap B_0)$.) The proof of [1, Proposition 4.2] implies that for any polyhedron K, every large map $f\colon K \to S$ that is homotopic into A and B is also homotopic into C_0. (Recall from Definition 3.1 that $f\colon K \to S$ is *large* if $f_*(\pi_1 K_0)$ is nonabelian for each component K_0 of K.) We will construct C by adding annular components to C_0.

Suppose K is a connected polyhedron and $f\colon K \to S$ is a map with $f_*(\pi_1 K) \neq \{1\}$, homotopic into A and B but not C_0. Let A_1 be a component of A such that f is homotopic into A_1, let $p\colon \widetilde{S} \to \operatorname{int} S$ be the covering space corresponding to $\pi_1 A_1$, and let $\widetilde{A} \subset \widetilde{S}$ be a component of $p^{-1}(A)$ mapping homeomorphically under p. Since A_1 is π_1–injective in S, the inclusion-induced homomorphism $\widetilde{A} \to \widetilde{S}$ is an isomorphism and hence every component of $X = \widetilde{S} - \operatorname{int} \widetilde{A}$ 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 $\widetilde{f}$ to $\widetilde{S}$; furthermore, the homotopy into B_0 lifts to a homotopy of $\widetilde{f}$ to a map g with image in $p^{-1}(B_0)$. Let $\widetilde{B}_0$ be the component of $p^{-1}(B_0)$

containing $g(K)$. Unlike in the proof of [1, Proposition 4.1], it is not necessarily true that $\widetilde{B}_0$ intersects $\widetilde{A}_1$. We will treat the two cases separately.

Suppose first that $\widetilde{B}_0 \cap \widetilde{A}_1 \neq \varnothing$. Then the argument that begins in the paragraph of [1] spanning pages 2407–2408 establishes that $\widetilde{B}_0$, hence also $\widetilde{f}$, deforms in $\widetilde{S}$ into $\widetilde{B}_0 \cap \widetilde{A}_1$. Projecting this homotopy of $\widetilde{f}$ to S yields a homotopy of f into a component of $B_0 \cap A_1$. Since $f_*(\pi_1 K) \neq \{1\}$, this component is not a disk. Since f is not homotopic into C_0, this component is not large, so it is an annulus Z_1 which moreover is not parallel to any component of C_0.

In this case, let $C_1 = C_0 \cup Z_1$, and let $A_1' = A$ and $B_1' = B_0$. These are subsurfaces of S respectively isotopic to A and B, such that C_1 is a union of components of their intersection.

Suppose now that $\widetilde{B}_0 \cap \widetilde{A}_1 = \varnothing$, and let Z be the component of X containing $\widetilde{B}_0$. Since $\pi_1 \widetilde{B}_0$ contains $g_*(\pi_1 K)$, which is nontrivial, $\widetilde{B}_0$ has a boundary component $\mathfrak{b}_0$ that is a homotopically nontrivial simple closed curve in Z. Hence $\mathfrak{b}_0$ cobounds an annulus $Z_0 \subset Z$ together with $\mathfrak{a}_0 = \partial Z$. If any component of the frontier in $\widetilde{S}$ 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 [1] spanning pages 2405–2406. Thus $\mathfrak{a}_0 \cap p^{-1}(B_0) = \varnothing$.

Since p projects A_1 homeomorphically, it sends $\mathfrak{a}_0$ homeomorphically to a component of ∂A_1 in S. Since $\widetilde{B}_0$ is a component of $p^{-1}(B_0)$, the covering map 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 k^{th} power of $\mathfrak{b}$. Since S is an orientable surface, by [4, Lemma 2.4], $k = 1$ and $\mathfrak{a}$ and $\mathfrak{b}$ bound an annulus Z_1 in S. It still holds in this case that g, and hence also f, is homotopic into Z_1 since the annulus $Z \subset \widetilde{S}$ containing $\widetilde{B}_0$ deformation retracts to $\mathfrak{a}_0$.

We may assume that $Z_1 \cap A = \mathfrak{a}$. If this is not so, then since A is essential and irredundant, a component A_2 of A intersects $\mathfrak{b}$. If a component of ∂A_2 intersected $\mathfrak{b}$, then by an innermost disk argument there would be an isotopy of B_0 reducing the number of intersections of ∂B_0 with ∂A, a contradiction. Therefore $\mathfrak{b} \subset A_2$, so f is homotopic into A_2, and putting A_2 in the role of A_1 in the argument above we find that $\widetilde{B}_0 \cap \widetilde{A}_2 \neq \varnothing$, ie we are in the first case. So assuming that none of the possible choices of A_1 yields the first case, we have $Z_1 \cap A = \mathfrak{a}$.

A similar argument shows that $B_0 \cap A_1 = \mathfrak{b}$, and it follows in this case that $A_1' = A \cup Z_1$ and $B_1' = B_0 \cup Z_1$ are respectively isotopic to A and B and that Z_1 is a component of $A_1' \cap B_1'$. Again in this case, let $C_1 = C_0 \cup Z_1$.

We now repeat the argument above but with C_1 in the role of C_0. If there is a polyhedron K and a map $f\colon K \to S$ homotopic into A and B but not into C_1, with $f_*(\pi_1 K) \neq \{1\}$, then this argument produces an essential annulus $Z_2 \subset S$ with f homotopic into Z_2. We may take Z_2 either to be a component of $A \cap B$ or to intersect each in a distinct component of its frontier. In either case Z_2 is disjoint from C_1, and there exist surfaces A_2' and B_2' respectively isotopic to A and B such that $C_2 = C_0 \cup Z_1 \cup Z_2$ is a union of components of $A_2' \cap B_2'$.

Iterating this process produces a sequence $\{C_n\}$ of subsurfaces of S with the property that $C_n = C_{n-1} \cup Z_n$ for an essential annulus Z_n disjoint from and not isotopic into C_{n-1}, which is either a component of $A \cap B$ or intersects A and B in distinct components of its frontier. The process terminates at some finite n, since $A \cap B$ has only finitely many components, and each of A and B have only finitely many boundary components. It then follows from the construction above that $C \doteq C_n$ has property $(*)$. □

The result below, which we will use in the proof of Proposition 4.3, extends [1, Proposition 4.4]. Its statement and proof follow those of its predecessor, but an additional case must be considered.

Below, for a subset S of a topological space X, we refer to the *frontier* of S in X as $\operatorname{fr} S \doteq \overline{S} \cap \overline{X - S}$.

Proposition 3.5 *Suppose B is an irredundant subsurface of a compact, orientable surface S with no 2–sphere components, and for a connected polyhedron K, let $f\colon K \to B$ satisfy $f_*(\pi_1 K) \neq \{1\}$. If $g\colon K \to B$ is homotopic to f in S, then*

(1) *either f and g are homotopic in B, or*
(2) *for distinct components $\mathfrak{a}$ and $\mathfrak{b}$ of the frontier of B that are parallel in S but not B, f is homotopic into $\mathfrak{a}$, and g into $\mathfrak{b}$, in B.*

Remark To directly extend [1, Proposition 4.4] we must allow K to be disconnected. Such a result is obtained by applying Proposition 3.5 component-by-component.

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

Since f maps K into B_0, composing with the lift of the inclusion map gives a lift $\tilde{f}\colon K \to \widetilde{S}$ with $\tilde{f}(K) \subset \widetilde{B}_0$; 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 $\tilde{B}_0$, then $H_1 = p \circ r \circ H$ is a homotopy between f and g with image in B_0.

If $\tilde{g}$ does not map into $\tilde{B}_0$, then the component of $p^{-1}(B)$ containing its image lies in a component Z of $\tilde{S} - \operatorname{int} B_0$, a half-open annulus. In this case, the time-1 map of $r \circ H$ has its image in the frontier $\tilde{\mathfrak{a}} = Z \cap \tilde{B}_0$ of Z. So $p \circ r \circ H$ is a homotopy of f in B, into a component $\mathfrak{a} = p(\tilde{\mathfrak{a}})$ of the frontier of B.

Switching the roles of f and g in the argument above, we find that if f and g are not homotopic in B, then g is homotopic in B into a component $\mathfrak{b}$ of the frontier of B. This is distinct from $\mathfrak{a}$ and not parallel to it in B, since it follows from algebraic topology that two maps from a polyhedron (or more generally, a CW-complex) K to S^1 that induce the same map on $\pi_1 K$ are homotopic. See, eg [7, Section 4.A, Exercise 2]. Let us now choose arcs from $\mathfrak{a}$ and $\mathfrak{b}$ to the basepoint of $\pi_1 S$ and again denote by $\mathfrak{a}$ and $\mathfrak{b}$ the elements of $\pi_1 S$ thus determined. The nontrivial subgroup $f_*(\pi_1 K) = g_*(\pi_1 K)$ of $\pi_1 S$ is contained in both a conjugate of the subgroup $\langle \mathfrak{a} \rangle$ generated by $\mathfrak{a}$ and a conjugate of $\langle \mathfrak{b} \rangle$.

Our hypotheses ensure that $\pi_1 S$ is isomorphic either to $\mathbb{Z} \oplus \mathbb{Z}$ or a Fuchsian group. In either case, standard results ensure that any two cyclic subgroups with nontrivial intersection both lie in a single cyclic group. (In the Fuchsian case, see, eg [11, Theorems 2.3.3 and 2.3.5].) Thus there exists $\gamma \in \pi_1 S$ such that certain conjugates of $\mathfrak{a}$ and $\mathfrak{b}$ are powers of γ. But since these conjugates represent the *simple* closed curves $\mathfrak{a}$ and $\mathfrak{b}$, they are primitive elements of $\pi_1 S$ (see, eg [5, Proposition 1.4]), and it follows that $\mathfrak{a}$ is conjugate to $\mathfrak{b}^{\pm 1}$ in $\pi_1 S$. Lemma 2.4 of [4] now implies that $\mathfrak{a}$ and $\mathfrak{b}$ are parallel. □

The lemma below distills a fact from the proof of Proposition 3.5 that we will use in the following section.

Lemma 3.6 *Let B be a compact, connected incompressible subsurface of a surface S, and for a polyhedron K suppose $f: K \to B$ is homotopic into $S - B$. Then f is homotopic in B into* $\operatorname{fr} B$.

Proof After pushing off boundaries, we will assume that $B \subset \operatorname{int} S$ and f maps into $\operatorname{int} B$. Choose a base point for $\pi_1 S$ in B, and let $p: \tilde{S} \to \operatorname{int} S$ be the cover corresponding to $\pi_1 B$. If $\tilde{B}$ is the image in $\tilde{S}$ of the lift of the inclusion map $B \hookrightarrow S$, then every component of $\tilde{S} - \operatorname{int} \tilde{B}$ is homeomorphic to a half-open annulus, and there is a retraction $r: \tilde{S} \to \tilde{B}$ that takes each such component to a component of $\partial \tilde{B}$. The homotopy of f out of B lifts to a homotopy $\tilde{H}$ whose time-0 map has its image in $\tilde{B}$. The time-1 map $\tilde{H}_1$ has its image in $\tilde{S} - \operatorname{int} \tilde{B}$, so $p \circ r \circ \tilde{H}$ is a homotopy of f in B to a map with its image in ∂B. □

4 Cylinders have bounded length

This section is dedicated to proving Theorem 4.1:

Theorem 4.1 *For a closed, orientable hyperbolic 3–manifold M and a connected, two-sided incompressible surface $S \subset M$ of genus g that is not a fiber or semifiber, a nondegenerate, reduced homotopy in (M, S) has length at most $14g - 12$.*

The proof uses the *characteristic submanifold* of the manifold X obtained by cutting M along S, which has the key property that it captures all nontrivial homotopies in X. We recall its definition below.

We say a 3–manifold X with boundary is *simple* if

- X is compact, connected, orientable, irreducible and boundary-irreducible;
- 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.

For a closed, orientable hyperbolic 3–manifold M containing an incompressible surface S, each component of the manifold obtained by cutting M along S is simple.

Below, an *essential annulus* in a 3–manifold X with boundary is the image of an essential, nondegenerate homotopy $(S^1 \times I, S^1 \times \partial I) \to (M, S)$ (recall Definition 2.3) that is an embedding. 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.

Theorem (Jaco and Shalen [9], Johansson [10]) *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 solid torus S such that $S \cap \partial X$ is a collection of disjoint, embedded annuli in ∂S that are homotopically nontrivial in S.*

(2) *Every component of the frontier of Ω is an essential annulus 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 $(\Omega, \Omega \cap \partial X)$.

Let the *characteristic set* of X be $\Omega \cap \partial X$. If X is a component of the manifold obtained by cutting M along S, then by the JSJ theorem its characteristic set carries a homotopic image of the time-0 map of any essential basic homotopy in (M, S) (recall Definition 2.4) that intersects X. The first main result of this section identifies a sequence of subsurfaces that play a role analogous to the characteristic set for homotopies with length $k \geq 1$. This extends [1, Proposition 5.2.8].

Before we state the result, we translate [1, Definition 5.1.1] into our context.

Definition 4.2 A *splitting surface* in a closed, orientable hyperbolic 3–manifold M is a transversely oriented, incompressible surface $S \subset M$ such that the manifold obtained by cutting M along S is a disjoint union of submanifolds $X^{\pm 1}$ with the property that $\mathcal{N}_\epsilon \subset X^\epsilon$ for each $\epsilon \in \{\pm 1\}$, for $\mathcal{N}_\epsilon$ as in Definition 2.4.

Separating, connected, two-sided incompressible surfaces are splitting surfaces, but note that S is not required above to be connected. In fact, given a nonseparating connected, two-sided incompressible surface S_0 in M, the boundary S of a regular neighborhood $\mathcal{N}_0$ of S_0 becomes a splitting surface upon taking $X^{-1} = \mathcal{N}_0$ and $X^{+1} = \overline{M - \mathcal{N}_0}$ and giving each component of S the transverse orientation pointing out of X^{-1}.

Proposition 4.3 *Let M be a closed, orientable hyperbolic 3–manifold and $S \subset M$ a splitting surface, and decompose the manifold obtained by cutting M along S into submanifolds $X^{\pm 1}$ as in Definition 4.2. 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. $\Psi_k^\epsilon \supset \Psi_{k+1}^\epsilon$.
2. *If K is a polyhedron with $\pi_1 K \neq \{1\}$ and $H\colon K \times I \to M$ is a reduced homotopy in (M, S) of length k, starting on the ϵ–side, then H_0 is homotopic in S to a map with image in Ψ_k^ϵ. Conversely, for such a polyhedron K if $f\colon K \to S$ is π_1–injective and homotopic into Ψ_k^ϵ, then there exists such a homotopy H with $H_0 = f$.*
3. $(\Psi_k^\epsilon)_{\mathcal{L}} = \Phi_k^\epsilon$*, where Φ_k^ϵ is the surface identified in [1, Proposition 5.2.8].*

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

Below we will briefly review some definitions and results proved in [1, Section 5]. These were proven there under the hypothesis that M is a *knot manifold*, with a single torus boundary component, whereas we take M closed. However, they depend only on the results on large intersection developed in [1, Section 4] and basic facts about I–bundles and so carry over to our context without alteration. The blanket hypotheses below are those of Proposition 4.3; in each case we paraphrase the result or definition from [1] that is referenced.

5.2.1 Let $(\Sigma^\epsilon, \Phi^\epsilon)$ 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 *There is a sequence* $\{\Phi_1^\epsilon \supset \Phi_2^\epsilon \supset \cdots\}$ *of large subsurfaces of* Φ^ϵ*, with* $\Phi_1^\epsilon = (\Phi^\epsilon)_{\mathcal{L}}$*, that satisfies property* (2) *of Proposition 4.3 with the hypothesis that* $\pi_1 K \neq \{1\}$ *replaced by the assertion that* H_0 *is large. The* Φ_i^ϵ *are determined up to isotopy by this property.*

Proposition 5.3.1 *There is a homeomorphism* $h_k^\epsilon \colon \Phi_k^\epsilon \to \Phi_k^{(-1)^{k+1}\epsilon}$*, for each* $k \in \mathbb{N}$*, 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* $h_k^\epsilon \circ f$*.*

To motivate the existence of h_k^ϵ, we note that the analog of Proposition 4.3(2) implies the inclusion of Φ_k^ϵ is the time-0 map of a length-k homotopy H with target (M, S). Since H is length-k, the image of H_1 lies in $\partial X^{(-1)^{k+1}\epsilon}$, and since H can be run backwards, this image is homotopic into $\Phi_k^{(-1)^{k+1}\epsilon}$.

The precise definition of the h_k^ϵ is as follows. Let τ_ϵ be the fixed-point free involution 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. 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 [1] the following is proven:

Proposition 5.3.4 *For* $k > 1$*, the restriction of* h_{k-1}^ϵ *to* Φ_k^ϵ *is homotopic in* S *to an embedding* $g_{k-1}^\epsilon \colon \Phi_k^\epsilon \to \Phi_1^{(-1)^{k-1}\epsilon}$ *with the property that* h_k^ϵ *is homotopic in* S *to* $\tau_{(-1)^{k-1}\epsilon} \circ g_{k-1}^\epsilon$*.*

Proposition 5.3.5 $h_{k-1}^\epsilon(\Phi_k^\epsilon)$ *is isotopic in* S *to* $(\Phi_{k-1}^{(-1)^k\epsilon} \cap_{\mathrm{ess}} \Phi_1^{(-1)^{k-1}\epsilon})_{\mathcal{L}}$*.*

The statements above are special cases of the results cited. Our phrasing of the latter implicitly uses our Proposition 3.3 (also see above it, and Definition 3.4).

The lemma below is a version of [1, Lemma 5.2.4], where the original hypothesis that the homotopy H in question has large time-0 map has been replaced here by the assertion that H maps into Σ^ϵ. In the original version this follows from the largeness hypothesis; the remainder of its proof carries through without revision.

The *standard* essential basic homotopy referenced below is from [1, Definition 5.2.3]. That definition in turn refers to the *fundamental* homotopy defined in 5.2.1 there.

For a component P of Σ^ϵ, which is an I–bundle in X^ϵ such that $P \cap \partial X^\epsilon$ is the associated ∂I–bundle (see above), the fundamental homotopy has domain $P \cap \partial X^\epsilon$ and takes I–fibers to I–fibers.

Lemma 4.4 *For $\epsilon \in \{\pm 1\}$ and a polyhedron K, let $H\colon (K \times I, K \times \partial I) \to (\Sigma^\epsilon, \Phi^\epsilon)$ be an essential basic homotopy. Then H is homotopic as a map of pairs to a standard essential basic homotopy. In particular, H_1 is homotopic in P to $\tau_\epsilon \circ H_0$.*

The lemma below extends the conclusion of [1, Proposition 5.3.1] to certain reduced homotopies whose time-0 maps are not necessarily large.

Lemma 4.5 *For $\epsilon \in \{\pm 1\}$ and $k \in \mathbb{N}$, suppose H is a reduced homotopy in (M, S) of length k that starts on the ϵ–side, with domain a polyhedron K, such that H_0 is homotopic in S into Φ_k^ϵ but not into $\partial\Phi_k^\epsilon$. Then H_1 is homotopic in S to $h_k^\epsilon \circ f$, where H_0 is homotopic to $f\colon K \to \Phi_k^\epsilon$.*

Proof We will assume without loss of generality that K is connected, since the desired homotopy can be constructed component-by-component. We prove the result first for $k = 1$; thus assume that $H\colon (K \times I, K \times \partial I) \to (X^\epsilon, \partial X^\epsilon)$ is an essential basic homotopy. Applying the JSJ theorem, after homotoping H through maps of pairs to $(X^\epsilon, \partial X^\epsilon)$, we will assume that it maps into Ω^ϵ.

Let P be an I–bundle component of Ω^ϵ such that $P \cap \partial X^\epsilon \subset \Phi_1^\epsilon$ contains the image of a map f homotopic to H_0. Since f is not homotopic into $\partial\Phi_1^\epsilon$, Lemma 3.6 implies that f is not homotopic out of $P \cap \partial X^\epsilon$, so H_0 and hence all of H maps into P. Lemma 4.4 now yields the conclusion in this case, since $h_1^\epsilon = \tau_\epsilon|_{\Phi_1^\epsilon}$.

Now take $k > 1$ and assume that the lemma holds for all reduced homotopies of length $k-1$. Given a reduced homotopy H of length k that satisfies the hypotheses, writing H as the composition of essential basic homotopies $H^1, \dots, H^k$, we have that the composition of $H^1, \dots, H^{k-1}$ has time-1 map homotopic to $h_{k-1}^\epsilon \circ f$, where H_0 is homotopic to $f\colon K \to \Phi_k^\epsilon \subset \Phi_{k-1}^\epsilon$.

Let $g_{k-1}^\epsilon\colon \Phi_k^\epsilon \to \Phi_1^{(-1)^{k-1}\epsilon}$ be the embedding supplied by [1, Proposition 5.3.4], homotopic to the restriction of h_{k-1}^ϵ and so that h_k^ϵ is homotopic to $\tau_{(-1)^{k-1}\epsilon} \circ g_{k-1}^\epsilon$. Let P be the I–bundle component of $\Omega^{(-1)^{k-1}\epsilon}$ such that $g_{k-1}^\epsilon \circ f$ maps into $\partial_h P$. Since f is not homotopic into $\partial\Phi_k^\epsilon$, the same holds true for $g_{k-1}^\epsilon \circ f$ in $\partial_h P$.

Since $H_0^k = H_1^{k-1}$, it is homotopic in S to $g_{k-1}^\epsilon \circ f$. 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 4.4 that H_1^k is homotopic to $\tau_{(-1)^{k-1}\epsilon} \circ g_{k-1}^\epsilon \circ f$. Therefore $H_1^k = H_1$ is homotopic to $h_k \circ f$, 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_k^ϵ is uniquely defined on Ψ_k^ϵ. But it is still true that every reduced homotopy is tracked by a homotopy of a surface containing the image of its time-0 map.

Lemma 4.6 *For $\epsilon \in \{\pm 1\}$ and $k \in \mathbb{N}$, suppose H is a reduced homotopy in (M, S) of length k that starts on the ϵ–side, with domain a connected, non-simply connected polyhedron K, such that H_0 is homotopic in S to a map f with image in an annulus A in $\Omega^\epsilon \cap \partial X^\epsilon$. There is a reduced homotopy J in (M, S) of length k that starts on the ϵ–side, with domain A, such that H_1 is homotopic to $J_1 \circ f$.*

Proof Consider the case in which $H\colon (K \times I, K \times \partial I) \to (X^\epsilon, \partial X^\epsilon)$ is an essential basic homotopy, for $\epsilon \in \{\pm 1\}$. By the JSJ theorem, after a homotopy through maps $(K \times I, K \times \partial I) \to (X^\epsilon, \partial X^\epsilon)$, we may assume H maps into some component P of the characteristic submanifold Ω^ϵ.

If the annulus A supplied by the hypotheses does not lie in P, then Lemma 3.6 implies that H_0 is homotopic into a component $\mathfrak{b}$ of $\partial(P \cap \partial X^\epsilon)$. The subgroups of $\pi_1 S$ respectively generated by $\mathfrak{b}$ and the core circle $\mathfrak{a}$ of A thus share the nontrivial subgroup $(H_0)_*(\pi_1 K)$. Since S is orientable and $\mathfrak{a}$ and $\mathfrak{b}$ are simple, this implies they generate identical subgroups, so $\mathfrak{a}$ and $\mathfrak{b}$ are parallel in S by [4, Lemma 1.4]. It follows that even if A does not lie in $P \cap \partial X^\epsilon$ it is still isotopic into it in ∂X^ϵ.

If P is an I–bundle component of Ω^ϵ, then applying Lemma 4.4, after a further homotopy we may assume that H is standard. If A lies outside of P, then by the paragraph above we may homotope H_0 so that its image lies in an annular neighborhood $B \subset P \cap \partial X^\epsilon$ of $\mathfrak{b}$, isotopic to A, with the property that B is a component of $\pi^{-1}(\pi(B)) \cap \partial X^\epsilon$. (Here π is the bundle projection of P.) Since H is standard, this determines a homotopy of H to a standard homotopy in the restriction of π to $\pi^{-1}(\pi(B))$.

A homotopy of A through X^ϵ is now determined by composing the isotopy J^0 from A to B with the restriction J^1 to B of the fundamental homotopy of $P \cap \partial X^\epsilon$. This becomes a basic essential homotopy J upon pushing $J^0.J^1\colon A \times I \to X^\epsilon$ off ∂X^ϵ on $\operatorname{int} I$. Since f is homotopic to H_0 in ∂X^ϵ, Proposition 3.5 now implies that $(J^0)_1 \circ f$ is homotopic to H_0 in B; possibility (2) there does not occur since the components of ∂B are not parallel in S, which has genus at least two. Since H and J^1 are standard, it now follows that $J_1 \circ f$ is homotopic to H_1.

Now suppose P is a solid torus component of Ω^ϵ, and let B and C be the components of $P \cap \partial X^\epsilon$ containing the images of H_0 and H_1, respectively. As in the previous case, if $A \neq B$, then there is an isotopy J^0 from A to B, and H_0 and $(J^0)_1 \circ f$ are homotopic in B. We now require a homotopy J_1 from B to C to replace the fundamental homotopy of the previous case. We construct this below.

Fix a homeomorphic lift $\widetilde{B}$ of B to the cover $p\colon \widetilde{P} \to P$ corresponding to $\pi_1 B$, let $\widetilde{H}$ be the lift of H with $\widetilde{H}_0(K) \subset \widetilde{B}$, and let $\widetilde{C}$ be the component of $p^{-1}(C)$ containing the image of $\widetilde{H}_1$. Note that since C is parallel to B on ∂P, the component $\widetilde{C}$ is also a homeomorphic lift of C. Moreover, since $\widetilde{B}$ and $\widetilde{C}$ carry $\pi_1 \widetilde{P}$, there is a product structure on $\widetilde{P}$, namely, $\widetilde{P} \cong X \times I$ for an annulus X, with $\widetilde{B} \cong X \times \{0\}$ and $\widetilde{C} \cong X \times \{1\}$. Restricting the fundamental homotopy of this product structure to $\widetilde{B}$ yields a homotopy $\widetilde{J}^1\colon \widetilde{B} \times I \to \widetilde{P}$ such that $(\widetilde{J}^1)_1$ is a homeomorphism to $\widetilde{C}$.

Let J^1 be $p \circ \widetilde{J}^1$ following the lift $B \to \widetilde{B}$ of the inclusion map $B \hookrightarrow P$. Now define a homotopy J through X^ϵ with domain A by pushing the composition $J^0.J^1$ off ∂X^ϵ on $\operatorname{int} I$. Lemma 4.4 implies that $\widetilde{H}$ is homotopic to a standard homotopy with respect to the product structure on $\widetilde{P}$, so since $(J^0)_1 \circ f$ is homotopic in B to H_0, it follows that $J_1 \circ f$ is homotopic in C to H_1.

This completes the proof of the essential basic case. The lemma now follows from this case and induction on the length k of the reduced homotopy. $\square$

Proof of Proposition 4.3 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

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

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 $\Phi_{m-1}^{(-1)^m \epsilon}$ representing

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

By Proposition 3.3, $(P_m^\epsilon)_{\mathcal{L}}$ is maximal among large surfaces of $\Phi_{m-1}^{(-1)^m \epsilon}$ that admit a homotopy of length one starting on the $(-1)^{m+1}\epsilon$–side. If a large subsurface A of $\Phi_{m-1}^{(-1)^m \epsilon}$ admits an essential homotopy of length one starting on the $(-1)^{m+1}\epsilon$–side, then $(h_{m-1}^\epsilon)^{-1}(A)$ admits a homotopy of length m starting on the ϵ–side; thus [1, 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 Lemma 3.2(3), these are isotopic subsurfaces of $\Phi_{m-1}^{(-1)^m \epsilon}$.

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

(a) Let $\{A_i\}$ be the set of annular components of Ψ_{m-1}^ϵ that admit a reduced homotopy of length m.

(b) 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 $\Phi_{m-1}^\epsilon - \operatorname{int}\Phi_m^\epsilon$.

(c) 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 for a reduced homotopy $H\colon (K\times I, K\times\partial I)\to(M,S)$ of length m that H_0 is homotopic into Ψ_m^ϵ.

Write H as a composition of essential basic homotopies $H^1,\dots,H^m$. Since the composition $H^1.H^2\dots H^{m-1}$ is a reduced homotopy of length $m-1$, by hypothesis $H_0=(H^1)_0$ is homotopic into Ψ_{m-1}^ϵ. If H_0 is homotopic into an annular component of Ψ_{m-1}^ϵ, then by Lemma 4.6, this component admits a reduced homotopy of length k; hence it is of the form A_i for some i. We thus assume below that this does not hold, hence that H_0 is homotopic into Φ_{m-1}^ϵ.

If H_0 is homotopic into Φ_m^ϵ, then we are done, so let us assume this is not the case. In particular, by [1, Proposition 5.2.8], H_0 is not large. If H_0 is homotopic into a boundary curve of Φ_{m-1}^ϵ that is not homotopic into Φ_m^ϵ, then by Lemma 4.6 again, the corresponding boundary component is of the form $\mathfrak{b}_j$ for some j.

By the preceding paragraph, we may assume that H_0 is homotopic into Φ_{m-1}^ϵ but not into $\partial\Phi_{m-1}^\epsilon$. Lemma 4.5 therefore implies that $(H^{m-1})_1$ is homotopic in S to $h_{m-1}^\epsilon\circ f\subset\Phi_{m-1}^{(-1)^m\epsilon}$, where $f\colon K\to\Phi_{m-1}^\epsilon$ is homotopic to H_0. It follows that $h_{m-1}^\epsilon\circ f$ admits an essential homotopy of length one, hence by Proposition 3.3 it is homotopic into a component C' of $P_m^\epsilon\subset\Phi_{m-1}^\epsilon$.

If $h_{m-1}^\epsilon\circ f$ were not homotopic into C' in $\Phi_{m-1}^{(-1)^m\epsilon}$, then Proposition 3.5 would imply in particular that it is homotopic in $\Phi_{m-1}^{(-1)^m\epsilon}$ into a boundary component. But then f, and hence H_0, would be homotopic to a boundary component of Φ_{m-1}^ϵ, contradicting our assumption. Hence $h_{m-1}^\epsilon\circ f$ is homotopic into C' in $\Phi_{m-1}^{(-1)^m\epsilon}$.

Recalling from above that $(P_m^\epsilon)_{\mathcal{L}}$ is isotopic in $\Phi_{m-1}^{(-1)^m\epsilon}$ to $h_{m-1}^\epsilon(\Phi_m^\epsilon)$, we find that C' is an annulus since we have assumed H_0 is not homotopic into Φ_m^ϵ. Therefore C' is of the form C_k' for some k, and we are in case (c) above. $\square$

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

Proposition 4.7 *Let M be a closed, orientable hyperbolic 3–manifold and $S \subset M$ a splitting surface that is not a fiber or a semifiber, and decompose the manifold obtained by cutting M along S into submanifolds $X^{\pm 1}$ as in Definition 4.2. For $\epsilon \in \{\pm 1\}$, let $\Psi_1^\epsilon \supset \Psi_2^\epsilon \supset \cdots$ be a sequence of irredundant surfaces that satisfy Proposition 4.3. Then for each k, we have that Ψ_k^ϵ is not homotopic into Ψ_{k+2}^ϵ.*

Proof Proposition 5.3.9 of [1] asserts that in this situation Φ_k^ϵ is not homotopic into Φ_{k+2}^ϵ for any $k \in \mathbb{N}$ or $\epsilon \in \{\pm 1\}$, so the result holds as long as Ψ_k^ϵ has nonempty large part. Therefore suppose for some k that Ψ_k^ϵ is a disjoint union of annuli homotopic into Ψ_{k+2}^ϵ.

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'' and H', each starting on the ϵ–side, where H' has length 2 and H'' length k. Since $H_1'(\Psi_{k+2}^\epsilon)$ admits a reduced homotopy of length k, Proposition 4.3 implies that H_1' is homotopic to a map $f\colon \Psi_{k+2}^\epsilon \to \Psi_k^\epsilon$. After applying the homotopy that takes Ψ_k^ϵ into Ψ_{k+2}^ϵ, we may take f to map into Ψ_{k+2}^ϵ. It follows that there exists a homotopy of length 2 in (M, S) with domain and target Ψ_{k+2}^ϵ.

Associate a directed graph G to this homotopy as follows: G has a vertex v for each component of Ψ_{k+2}^ϵ, 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 $v_0, \dots, v_{m-1}$ 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_{k+2}^ϵ 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 $F_0^i = \mathfrak{a}_i$ and $F_1^i = \mathfrak{a}_{i+1}$ (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 the hyperbolicity of M. $\square$

We may now prove Theorem 4.1, which extends [1, Theorem 5.4.1].

Proof of Theorem 4.1 If S is nonseparating, we replace S by the boundary $\widetilde{S}$ of a regular neighborhood, yielding a separating surface with two components of genus g. If S is separating we take $\widetilde{S} = S$, and in either case let $X^{\pm 1}$ be the components of the manifold obtained by cutting M along S. For $\epsilon \in \{\pm 1\}$, let $\Psi_1^\epsilon \supset \Psi_2^\epsilon \supset \cdots$ be a sequence of irredundant surfaces that satisfies the conclusion of Proposition 4.3.

We now briefly review the proof of [1, Theorem 5.4.1]. Given a large surface A, the complexity of A is defined as $c(A) = g(A) - 3\chi(A)/2 - |A|$, where $\chi(A)$ is the Euler characteristic of A, the number of its components is $|A|$, 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 $\epsilon \in \{\pm 1\}$, consider the subsequence

$$\Phi_1^\epsilon \supset \Phi_3^\epsilon \supset \cdots .$$

This is strictly decreasing by [1, Proposition 5.3.9] and consists of large surfaces with even Euler characteristic by [1, Corollary 5.3.8]. Thus for each $i \geq 0$, we have $c(\Phi_{2i+1}^\epsilon) > c(\Phi_{2i+3}^\epsilon)$. If S is separating, then $c(\widetilde{S}) = c(S) = 4g - 4$, and otherwise $c(\widetilde{S}) = 8g - 8$. Taking $m_S = 4g - 4$ in the separating case and $m_S = 8g - 8$ in the nonseparating case, it follows that $\Phi_{2i+1} = \varnothing$ for $i > m_S$.

The discussion above is enough to establish [1, Theorem 5.4.1]. In our situation of interest, it establishes that Ψ_{2i+1}^ϵ is a disjoint union of annuli for $i > m_S$. Since Ψ_i^ϵ 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, nonparallel, essential simple closed curves on a closed surface of genus g has at most $3g - 3$ members.) Since Proposition 4.7 implies Ψ_{2i+1}^ϵ is not homotopic into Ψ_{2i+3}^ϵ, if these are unions of irredundant collections of annuli, then Ψ_{2i+3}^ϵ has fewer components than Ψ_{2i+1}^ϵ. Thus taking $n_S = 3g - 3$ in the separating case and $n_S = 6g - 6$ otherwise, we find that $\Psi_{2i+1} = \varnothing$ for $i > m_S + n_S$.

By Proposition 4.3, the time-0 map of a reduced homotopy in $(M, \widetilde{S})$ with length k that starts on the ϵ–side is homotopic into Ψ_k^ϵ. Therefore $k \leq 2(m_S + n_S) + 2$. If S is separating, we therefore find that homotopies in $(M, S) = (M, \widetilde{S})$ have length at most $14g - 12$. If S is nonseparating, a reduced homotopy of length k in (M, S) determines a reduced homotopy of length $2k - 1$ in $(M, \widetilde{S})$. Thus in this case we have for a homotopy of length k in (M, S) that $2k - 1 \leq 2(14g - 14) + 2$, so $k \leq 14g - 13$. The theorem follows. □

References

[1] **S Boyer**, **M Culler**, **P B Shalen**, **X Zhang**, *Characteristic subsurfaces and Dehn filling*, Trans. Amer. Math. Soc. 357 (2005) 2389–2444 MR

[2] **D Cooper**, **D D Long**, *Virtually Haken Dehn-filling*, J. Differential Geom. 52 (1999) 173–187 MR

[3] **J DeBlois**, **S Friedl**, **S Vidussi**, *Rank gradients of infinite cyclic covers of 3–manifolds*, Michigan Math. J. 63 (2014) 65–81 MR

[4] **D B A Epstein**, *Curves on 2–manifolds and isotopies*, Acta Math. 115 (1966) 83–107 MR

[5] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Mathematical Series 49, Princeton University Press (2012) MR

[6] **F Haglund**, **D T Wise**, *Special cube complexes*, Geom. Funct. Anal. 17 (2008) 1551–1620 MR

[7] **A Hatcher**, *Algebraic topology*, Cambridge University Press (2002) MR

[8] **J Hempel**, *3–Manifolds*, Princeton University Press; University of Tokyo Press (1976) MR

[9] **W H Jaco**, **P B Shalen**, *Seifert fibered spaces in 3–manifolds*, Mem. Amer. Math. Soc. 220, Amer. Math. Soc., Providence, RI (1979) MR

[10] **K Johannson**, *Homotopy equivalences of 3–manifolds with boundaries*, Lecture Notes in Mathematics 761, Springer, Berlin (1979) MR

[11] **S Katok**, *Fuchsian groups*, University of Chicago Press (1992) MR

[12] **M Lackenby**, *Expanders, rank and graphs of groups*, Israel J. Math. 146 (2005) 357–370 MR

[13] **T Li**, *Immersed essential surfaces in hyperbolic 3–manifolds*, Comm. Anal. Geom. 10 (2002) 275–290 MR

[14] **P Scott**, **T Wall**, *Topological methods in group theory*, from: "Homological group theory", (C T C Wall, editor), London Math. Soc. Lecture Note Ser. 36, Cambridge University Press (1979) 137–203 MR

[15] **Z Sela**, *Acylindrical accessibility for groups*, Invent. Math. 129 (1997) 527–565

[16] **J-P Serre**, *Trees*, corrected reprint of 1st edition, Springer, Berlin (1980) MR

[17] **W P Thurston**, *A norm for the homology of 3–manifolds*, Mem. Amer. Math. Soc. 339, Amer. Math. Soc., Providence, RI (1986) i–vi and 99–130 MR

[18] **M D Tretkoff**, *A topological approach to the theory of groups acting on trees*, J. Pure Appl. Algebra 16 (1980) 323–333 MR

[19] **G S Walsh**, *Incompressible surfaces and spunnormal form*, Geom. Dedicata 151 (2011) 221–231 MR

[20] **R Weidmann**, *The Nielsen method for groups acting on trees*, Proc. London Math. Soc. 85 (2002) 93–118 MR

[21] **D T Wise**, *From riches to raags: 3–manifolds, right-angled Artin groups, and cubical geometry*, CBMS Regional Conference Series in Mathematics 117, Amer. Math. Soc., Providence, RI (2012) MR

[22] **D T Wise**, *The structure of groups with a quasiconvex hierarchy*, preprint (2012) Available at `http://www.math.mcgill.ca/wise/papers.html`

Department of Mathematics, University of Pittsburgh
301 Thackeray Hall, Pittsburgh, PA 15260, United States

`jdeblois@pitt.edu`

`http://www.pitt.edu/~jdeblois`

Received: 4 November 2013 Revised: 19 October 2015