THE KERVAIRE CONJECTURE AND THE MINIMAL COMPLEXITY OF SURFACES

LVZHOU CHEN

Abstract. The Kervaire conjecture asserts that adding a generator and then a relator to a nontrivial group always results in a nontrivial group. We introduce new methods from stable commutator length to study this type of problems about nontriviality of one-relator quotients. Roughly, we show that surfaces in certain HNN extensions bounding a given word have complexity no less than the complexity of its boundary. As a consequence, for any group G and the quotient Q of $G \star \mathbb{Z}$ by any proper power w^m with $w \in G \star \mathbb{Z}$ projecting to $1 \in \mathbb{Z}$, the natural map $G \to Q$ is injective. Another consequence is a new proof of Klaychko's theorem that confirms the Kervaire conjecture for torsion-free groups.

1. Introduction

Many problems in low-dimensional topology have group-theoretic formulations. Some of these topological problems remain unsolved due to our lack of understanding of groups obtained from some rather simple operations. One example is the following basic question about *one-relator products*, which are one-relator quotients of free products.

Question 1.1. *For a free product $H = \star_{\lambda \in \Lambda} G_\lambda$ of nontrivial groups $\{G_\lambda\}_{\lambda \in \Lambda}$ with $|\Lambda| \geq 2$, for which $w \in H$ is the quotient $H/\langle\langle w \rangle\rangle$ nontrivial, where $\langle\langle w \rangle\rangle$ is the normal subgroup generated by w?*

For $|\Lambda| \geq 3$, it is conjectured that $H/\langle\langle w \rangle\rangle$ is nontrivial for every $w \in H$; See for instance [Gor83, Conjecture 9.5]. This is a generalization[1] of the unsolved three summand conjecture in 3-manifold topology, which asserts that the Dehn surgery of S^3 along any knot cannot have three or more summands in the prime decomposition. The case $|\Lambda| \geq 3$ is known when all factors are cyclic groups by a theorem of Howie [How02].

The analogous statement fails when $|\Lambda| = 2$, for instance $H/\langle\langle w \rangle\rangle$ is trivial for $w = ab$ when $H = (\mathbb{Z}/m) \star (\mathbb{Z}/n)$ with natural generators a, b and m, n coprime. However, one still expects $H/\langle\langle w \rangle\rangle$ to be nontrivial for all w when the factors are torsion-free:

Conjecture 1.2. *If A and B are torsion-free, then $H/\langle\langle w \rangle\rangle$ is nontrivial for any $w \in H = A \star B$.*

Date: February 16, 2023.

[1]The knot group is normally generated by a single element (the meridian), so is the fundamental group of the Dehn surgery as it is a quotient.

A weaker statement contributed by Freedman appears on Kirby's (1970s) problem list [Kir78, Problem 66]. On the topological side, this is related to the cabling conjecture [GAnS86a] about irreducibility of Dehn surgeries on knots in S^3, which implies the three-summand conjecture; See e.g. [How02, Page 2]. The conjecture is known under the stronger assumption that A and B are locally indicable[2] by Brodskiĭ [Bro84] and independently Howie [How82].

The goal of this paper is to bring in new tools from the seemingly unrelated study of stable commutator length to tackle such problems; See [Cal09a] for a great reference to this topic. Roughly speaking, we show that the complexity (measured by the negative Euler characteristic) of certain surface maps to a $K(H,1)$ space for some HNN extension H is no less than the complexity of its boundary (measured by a geometric degree); See Theorems C and D in Section 1.1 for precise statements. This is a generalization of the so-called spectral gap phenomenon in stable commutator length.

Restricting to planar surfaces and $H = A \star \mathbb{Z}$ (as the free HNN extension of A) with A torsion-free, our result implies the Klyachko theorem [Kly93] and gives a new proof.

Theorem A (Klyachko, Theorem 6.7)**.** *For any torsion-free group A, the natural map $A \to H/\langle\langle w\rangle\rangle$ induced by the inclusion $A \to H = A \star \mathbb{Z}$ is injective for any $w \in H$ with $p(w) = 1$, where $p : H \to \mathbb{Z}$ is the standard projection to the $\mathbb{Z}$ factor.*

The interest in Klyachko's theorem and its (new) proofs is twofold. On the one hand, Theorem A implies the special case of Conjecture 1.2 where $B = \mathbb{Z}$, since one can easily reduce the problem to the case $p(w) = 1$ (by abelianization). On the other hand, it is one of the most important progress on the Kervaire–Laudenbach Conjecture 1.3 and the Kervaire Conjecture 1.4 below. These conjectures originate from Kervaire's classification of high-dimensional knot groups [Ker65] and remain open in general. Although Klyachko's theorem has been known for three decades, no significant breakthrough beyond this has been made towards the more general Conjectures 1.2, 1.3, and 1.4. It is our hope that new approaches can lead to further progress.

Conjecture 1.3 (Kervaire–Laudenbach)**.** *For any $H = A\star\mathbb{Z}$, the natural map $A \to H/\langle\langle w\rangle\rangle$ induced by the inclusion $A \to H$ is injective for any $w \in H$ with $p(w) \neq 0$.*

Conjecture 1.4 (Kervaire)**.** *Is $H/\langle\langle w\rangle\rangle$ nontrivial for all $w \in H = A\star\mathbb{Z}$ if A is nontrivial?*

Another influential progress on Conjectures 1.3 and 1.4 is the theorem of Gerstenhaber–Rothaus [GR62], proving the case for any A finite (and consequently any A residually finite). A more extensive summary of known results on these conjectures and their generalizations can be found in the survey [Rom12] from the view of equations over groups.

Our method also implies the following (to our best knowledge) new result, which works for an *arbitrary* factor group A but assumes the relator to be a proper power. Note that very few partial results about Conjecture 1.3 works for an arbitrary factor group A.

Theorem B (Theorem 6.8)**.** *For any group A, the natural map $A \to H/\langle\langle w^m\rangle\rangle$ induced by the inclusion $A \to H = A \star \mathbb{Z}$ is injective for any $w \in H$ with $p(w) = 1$ and $m \geq 2$.*

[2] A group is locally indicable if every finitely generated nontrivial subgroup surjects $\mathbb{Z}$.

This establishes a new partial result towards the following more general Conjecture 1.5 about one-relator quotients by proper powers. A theorem of Howie [How90] proves this conjecture when the exponent $m \geq 4$, generalizing an earlier result for $m \geq 6$ by González-Acuña and Short [GAnS86b, Theorem 4.3]. Theorem B seems to be the first result that brings the exponent down to 2 for a very general class of groups.

Conjecture 1.5. *For an arbitrary free product, the natural map $A_\lambda \to (\star_\lambda A_\lambda)/\langle\langle w^m \rangle\rangle$ induced by the inclusion $A_\lambda \to \star_\lambda A_\lambda$ is injective whenever $m \geq 2$ and w is not conjugate to an element of A_λ.*

1.1. **More detailed statements of results.** Now we give the more precise statements on our results about complexity of surfaces and their relation to the problems above.

Fix a $K(H, 1)$ space X for an HNN extension $H = A\star_C$ and an element $w \in H$ not conjugate into A. The objects of our study are *w-admissible surfaces*, each of which is a continuous map $f : S \to X$ from a compact oriented surface S so that the image of each boundary component B_i of S represents either the conjugacy class of w^{n_i} for some $n_i \neq 0 \in \mathbb{Z}$ or some conjugacy class in A (for which we take $n_i = 0$). The (geometric) *degree* of S, denoted $\deg(S)$, is defined as the sum of $|n_i|$ over all boundary components B_i of S. See Definition 2.1 for a more general definition.

We actually focus on *boundary-incompressible* w-admissible surfaces S, which essentially means that boundary components of S representing w^m and w^{-n} cannot cancel in a naive way for $m, n \in \mathbb{Z}_+$; See Definition 2.4 for the precise definition. Any w-admissible surface can be simplified to a boundary-incompressible one.

A less technical version of our main result is:

Theorem C (Corollary 6.6)**.** *For the free HNN extension $H = A \star \mathbb{Z}$ of A, where each nontrivial element of A has order at least n for some $2 \leq n \leq \infty$ (which is automatic if $n = 2$), for any $w \in H$ with $p(w) = 1$, every boundary-incompressible w-admissible surface S has*

$$-\chi(S) \geq (1 - \frac{1}{n}) \deg(S).$$

This implies Theorem A (and similarly Theorem B) for the following reason. If some $a \neq id \in A$ becomes trivial in the quotient $H/\langle\langle w \rangle\rangle$, then a can be written as a product of conjugates of w and w^{-1} in H (with $h_i \in H$):

$$a = (h_1 w^{\pm 1} h_1^{-1}) \cdots (h_k w^{\pm 1} h_k^{-1}).$$

Topologically this is equivalent to a w-admissible surface S with $\deg(S) = k$, which is a sphere with $k + 1$ disks removed and hence has $-\chi(S) = k - 1 < \deg(S)$. The inequality in Theorem C (with $n = \infty$ as A is torsion-free) rules out the existence of such surfaces provided that we simplify the equation above to ensure that the surface S is boundary-incompressible.

The more general main result works for HNN extensions of a group A over isomorphic subgroups C_1, C_2, assuming that the group-subgroup pair (A, C_i) satisfies a *length-n relatively free (n-RF)* condition, which essentially assumes that there is no short relation (quantified by n) between any $a \in A \setminus C_i$ and C_i, $i = 1, 2$; See Definition 5.1 for the precise definition and Section 5.5 for a more detailed discussion on this condition.

Theorem D (Theorem 6.1). *Let $H = A\star_C$ be the HNN extension associated to inclusions $C \to A$ with images $C_1, C_2 \leq A$ such that (A, C_i) is n-RF for some $2 \leq n \leq \infty$ and $i = 1, 2$. Let $p : H \to \mathbb{Z}$ be the projection to $\mathbb{Z}$ that restricts trivially to A. Then for any $w \in H$ with $p(w) = \pm 1$ and not conjugate to $at^{\pm 1}$ for any $a \in A$, every boundary-incompressible w-admissible surface has*

$$-\chi(S) \geq (1 - \frac{1}{n}) \deg(S).$$

This is analogous to the spectral gap theorem proved by the author and Nicolaus Heuer in the context of (relative) stable commutator length (scl) in graphs of groups [CH19, Theorem A]. Our w-admissible surfaces are analogous to the admissible surfaces in graphs of groups relative to the vertex groups in the scl sense [CH19, Definition 2.8]. The key difference is that here we use the *geometric* degree instead of the algebraic degree in the scl context, and this crucial difference makes the problem harder in our context.

One of the key tool used to prove spectral gap properties of scl is to construct suitable quasimorphisms and apply Bavard's duality [Bav91]. Due to the key difference in the notion of degree, it is unclear if this approach is still applicable here.

However, we are still able to adapt in our context the LP-duality method that the author developed to prove sharp lower bounds and spectral gaps of scl; See [Che20, Section 6.3] and [CH19, Section 5.2] for an introduction of this method for scl. We focus on the adaptation of this method to w-admissible surfaces in HNN extensions, but the same method applies to graphs of groups and in particular free products of groups, which we leave to future work.

1.2. **Organization of the paper.** This paper is organized as follows: We give basic definitions about w-admissible surfaces in Section 2, and we introduce a normal form of such surfaces in Section 3. Then we explain the (adapted) LP-duality method in Section 4 and apply it to prove a main technical result (Theorem 5.3) in Section 5. A brief discussion on the main n-RF assumption can be found in Section 5.5. Finally in Section 6 we apply Theorem 5.3 to prove Theorem D (Theorem 6.1), and from its special case Theorem C (Corollary 6.6) we deduce Theorems A and B (Theorems 6.7 and 6.8).

Acknowledgment. The author deeply thanks Danny Calegari for suggesting the potential connection between stable commutator length and the Kervaire conjecture back in 2017. The author is very grateful to Cameron Gordon for re-stimulating the author's interest in such problems and for numerous discussions on this topic, from which a gap in an earlier proof was filled. The author also thanks Daniel Allcock, Jeff Danciger, Francesco Fournier-Fabio, Anton Klaychko, John Luecke, Bestvina Mladen, Yi Ni, Henry Wilton for helpful conversations and suggestions.

2. Admissible surfaces

Fix a group H with a proper subgroup A, and let X be a connected topological space with $\pi_1(X) = H$. In this section, we introduce *w-admissible surfaces* associated to an

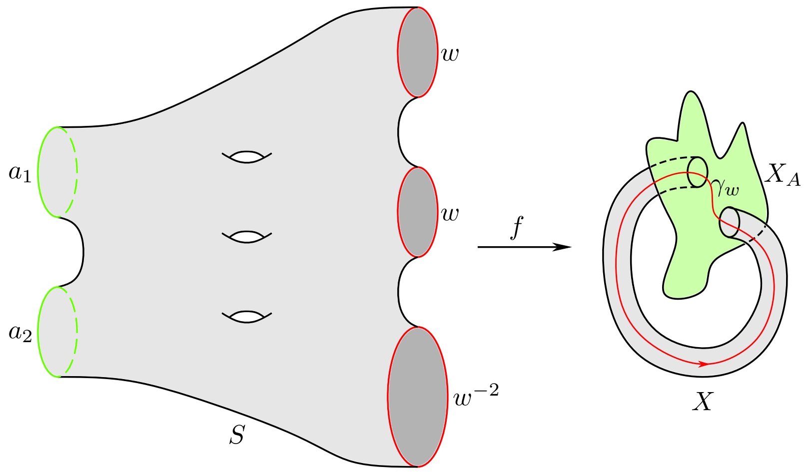

FIGURE 1. X has a subspace X_A representing the subgroup $A \leq H = \pi_1(X)$ and a loop γ_w representing some $w \in H$. S is a connected w-admissible surface of degree 4, where the two boundary components of S on the left are A-boundary, mapped to conjugacy classes of $a_1, a_2 \in A$, and the three on the right are w-boundary of S, representing powers of w.

element $w \in H$ not conjugate into A. We are mostly interested in the case of an HNN extension $H = A\star_C$, but the definitions make sense in general.

Definition 2.1 (w-admissible)**.** A map $f : S \to X$ from a compact oriented surface S is *w-admissible* if the image under f of each component of ∂S either

(1) represents a conjugacy class in A, or
(2) represents the conjugacy class of w^n for some $n \in \mathbb{Z} \setminus \{0\}$.

We refer to the union of boundary components of the first type as the *A-boundary* of S, and refer to the union of the second type as the *w-boundary* of S; See Figure 1. A w-boundary is *positive* (resp. *negative*) if the exponent $n > 0$ (resp. $n < 0$). We allow A-boundary to be empty but make the convention throughout this paper that w-boundary is nonempty for each component of S. In particular, each component of S has non-positive Euler characteristic as w is nontrivial.

Although the map f is part of the data, we often abbreviate and refer to S as a w-admissible surface by thinking of it as a (singular) subsurface in X. When w is understood, we simply call S an admissible surface.

The *degree* of a w-boundary component representing w^n is $|n|$. Define the *degree* $\deg(S)$ of a w-admissible surface S to be the sum of degrees of all w-boundary components. By our convention we have $\deg(S) \in \mathbb{Z}_+$. This is well defined when no w^n is conjugate to w^m whenever $m \neq n$. Similarly, we define $\deg_+(S)$ (resp. $\deg_-(S)$) to be the sum of degrees only over w-boundary components representing w^n for some $n > 0$ (resp. $n < 0$). We have $\deg(S) = \deg_+(S) + \deg_-(S)$.

Remark 2.2. One should really refer to these surfaces as w-admissible surfaces in X (or H) *relative to* A. However, in most part of this paper, we fix A and H as in an HNN extension $H = A\star_C$. The only exception is in the proof oF Theorem 6.1, where we pass to a different HNN extension structure on H that enlarges the subgroup A. Note that a w-admissible surface relative to A is also a w-admissible surface relative to A' if $A \le A'$ (and $w \notin A'$).

There is a similar notion of admissible surfaces in the topological definition of stable commutator length [Cal09b, Notation 2.5]; See also [Che20, Definition 2.8]. However, in that context a boundary component representing w^n for $n < 0$ is defined to have negative degree or is simply disallowed by considering the so-called monotone admissible surfaces [Cal09b, Lemma 2.7]. In our setting we consider the geometric degree instead of the algebraic degree.

For an HNN extension $H = A\star_C$, there is a surjective homomorphism $p : H \to \mathbb{Z}$ that vanishes on A, taking the standard new generator t to itself (see the presentation in (3.1) below).

Lemma 2.3. *If $H = A\star_C$ is an HNN extension and $p(w) \neq 0$, then for any w-admissible surface S, we have*

$$\deg_+(S) = \deg_-(S) = \frac{1}{2}\deg(S).$$

Proof. The homomorphism p factors through the abelianization $H_1(H;\mathbb{Z}) = H_1(X;\mathbb{Z})$. Note that $[\partial S]$ is a trivial first homology class, and p vanishes on A, so we must have

$$\deg_+(S)\cdot p(w) + \deg_-(S)\cdot p(w^{-1}) = 0.$$

As $p(w^{-1}) = -p(w) \neq 0$, it follows that $\deg_+(S) = \deg_-(S)$, which is half of the total degree $\deg(S)$. □

Definition 2.4 (Boundary incompressibility)**.** A w-admissible surface S is *boundary compressible* if there is a compact subsurface $\Sigma \subset S$ which is a pair of pants so that two components of $\partial\Sigma$ are w-boundary components of S representing the conjugacy classes w^n and w^{-m} for some $m, n \in \mathbb{Z}_+$ and the third component of $\partial\Sigma$ is a loop in S (with the orientation induced from Σ) representing the conjugacy class of w^{m-n} (under the map $f : S \to X$); See Figure 2.

A w-admissible surface S is *boundary incompressible* if it is not boundary compressible. In particular, given any base point p in such a surface, for any two w-boundary components, their images in $\pi_1(X, p)$ cannot be expressed as hw^nh^{-1} and $hw^{-m}h^{-1}$ for some $h \in \pi_1(X, p)$ and $m, n \in \mathbb{Z}_+$.

One can keep simplifying a boundary compressible w-admissible surface until it either has no w-boundary left or becomes boundary incompressible. Indeed, for a pair of pants $\Sigma \subset S$ as in the definition, consider a new surface $S' = S \setminus \Sigma$, where we further cap off the new boundary representing w^{n-m} if $n = m$. The new surface S' has $-\chi(S') \le -\chi(S) - 1$ and $\deg(S') = \deg(S) - 2\min(m, n) \le \deg(S) - 2$; See Figure 2.

The following examples show how w-admissible surfaces naturally correspond to certain kinds of equations in the group H. Such equations arise naturally in our application to the Kervaire–Laudenbach conjecture.

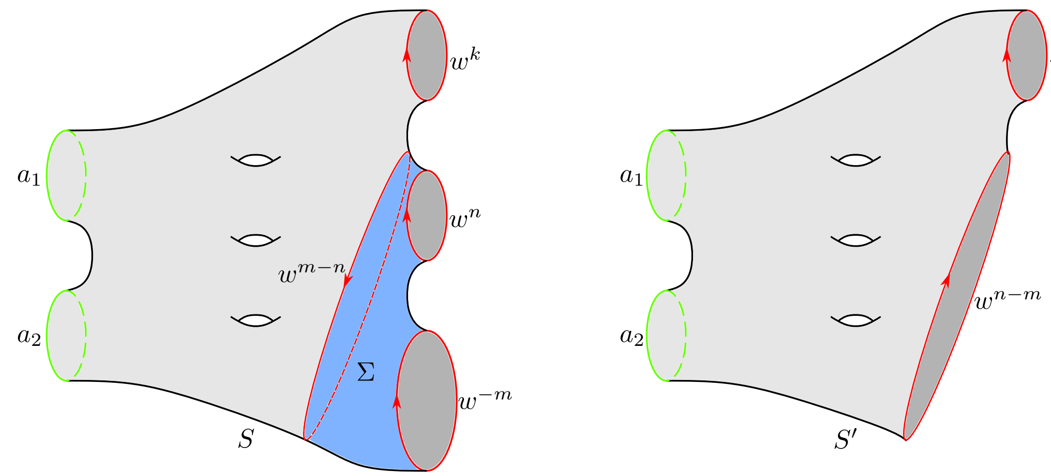

FIGURE 2. A boundary compressible w-admissible surface S with $k, m, n \in \mathbb{Z}_+$ and the simplified w-admissible surface $S' = S \setminus \Sigma$, whose boundary representing w^{n-m} (with the orientation induced from S') needs to be further capped off by a disk if $m = n$

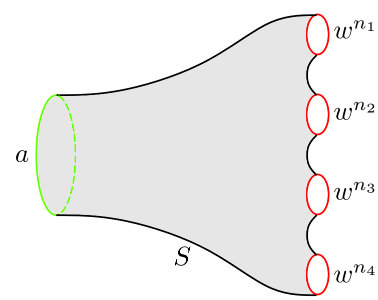

FIGURE 3. A w-admissible surface S corresponding to an equation of the form (2.1) in the case $k = 4$.

Example 2.5. *Suppose $a \in A \leq H$ lies in the normal closure $\langle\langle w \rangle\rangle$ of w, i.e. there are equations in H of the form*

$$a = (h_1 w^{n_1} h_1^{-1}) \cdots (h_k w^{n_k} h_k^{-1}) \tag{2.1}$$

for some $k \geq 1$, $n_i \in \mathbb{Z} \setminus \{0\}$, and $h_i \in H$. Such an expression provides a w-admissible S, which is a sphere with $k+1$ disks removed, where one boundary component represents g and the other k components represent the conjugacy classes of w^{n_i} for $i = 1, \ldots, k$.

If S is boundary compressible, then the simplified surface S' above gives another expression of a of the form (2.1) with a smaller k. Hence if k is minimal among all such expressions of g, then the corresponding w-admissible surface S is boundary incompressible.

Note that the surface in the example above is connected, planar, and has exactly one A-boundary component. None of these properties are required for a w-admissible surface in general.

3. A normal form

Starting from this section, we focus on an HNN extension $H = A\star_C$ given by two injections $i_P, i_N : C \hookrightarrow A$. Denote the two images as C_P and C_N respectively.

We develop a normal form for w-admissible surfaces, which is a decomposition of into disks and annuli with combinatorial boundary information. This is parallel to the normal form in [Che20] for admissible surfaces (in the context of stable commutator length) in a graph of spaces. The fact that here we have boundary components representing w^n with $n < 0$ does not affect the process of simplifying an admissible surface to put it in the normal form. We include some details for completeness. The discussion below works for any graph of groups, but we focus on the case of an HNN extension for concreteness.

3.1. **Basic setup.** Let (X_A, b_A) and (X_C, b_C) be based $K(A, 1)$ and $K(C, 1)$ spaces respectively. The two inclusions of C into A are represented by continuous maps i_P, i_N : $(X_C, b_C) \to (X_A, b_A)$ respectively. Thus we can build the space X as a graph of space, where the graph is just a loop with one vertex, X_A is the vertex space, and X_C is the edge space. Explicitly, X is a quotient of $X_A \sqcup (X_C \times [-1, 1])$, where any $(x, -1) \in X_C \times \{-1\}$ is glued to $i_N(x) \in X_A$ and $(x, 1) \in X_C \times \{1\}$ is glued to $i_P(x) \in X_A$. The space X built this way is a $K(H, 1)$ for $H = A\star_C$.

Note that X_A is naturally a subspace of X. We also identify X_C with the subspace $X_C \times \{0\}$ of X, which has a product neighborhood $X_C \times (-1, 1)$. Cutting X along X_C (and taking completion) yields a space V, which is the mapping cylinders associated to i_P and i_N with X_A identified. We call V the *thickened vertex space* and note that it deformation retracts to X_A; See the top-right of Figure 4. The image of $\{b_C\} \times [-1, 1]$ in X is a loop with the standard orientation on $[-1, 1]$, which we denote by t. We abuse the notation to also denote by t the corresponding element in $\pi_1(X, b_A) = H = A\star_C$. This way we obtain the standard presentation

$$H = \langle A, t \mid i_N(c) = t i_P(c) t^{-1} \text{ for all } c \in C\rangle. \tag{3.1}$$

Represent w as a loop $\gamma : S^1 \to X$, which we identify with its image in X. We choose in below a good representative of γ in its free homotopy class corresponding to a cyclically reduced expression $w = a_1 t^{e_1} \dots a_k t^{e_k}$ with $e_i = \pm 1$ and $a_i \in A$. We denote $|w| := k$. For each $a_i \in A$, represent it as a loop α_i in X_A based at b_A. In the cyclically reduced expression above, replace each a_i by α_i, interpret each t^{e_i} as the loop t with the appropriate orientation depending on e_i, and replace group operation by concatenation to obtain our representative of γ. By construction, γ intersects X_C transversely exactly k times. Since w does not conjugate into A, we have $k \geq 1$, and the intersections with X_C divide γ into k segments $\gamma_1, \dots, \gamma_k$, where each γ_i starts (resp. ends) with the second (resp. first) half of $t^{e_{i-1}}$ (resp. t^{e_i}) and follows α_i in the middle, indices taken mod k. We say such a representative γ is *tight* and fix it in the discussion below.

Using the product neighborhood of the edge space X_C homeomorphic to $X_C \times (-1, 1)$, each segment γ_i starts from (resp. ends at) either the negative or positive side of X_C. This divides the above segments into four types, PP, PN, NP, NN, where the first (resp. second) letter indicates which side the arc starts from (resp. ends at), with P and N standing for positive and negative respectively. Algebraically, the first (resp. second) letter for the type of γ_i is P if $e_{i-1} = 1$ (resp. $e_i = -1$). Similarly, this also defines the type of each segment γ_i^{-1} so that its type is the type of γ_i with the two letters swapped. We will use this in Section 3.3.

3.2. **Putting S in (simple) normal form.** Fix any w-admissible surface $f : S \to X$. Up to homotopy, we assume that f restricted to each w-boundary is a covering map to γ, and each A-boundary has image in X_A. Putting f in general position so that it is transverse to X_C, then $f^{-1}(X_C)$ is an embedded proper submanifold of codimension 1, i.e. a finite disjoint union of embedded loops and proper embedded arcs with endpoints on w-boundary components. $f^{-1}(X_C)$ divides each w-boundary of degree n into exactly $|n|k$ segments.

One can always homotop f and possibly simplify S so that $f^{-1}(X_C)$ has no embedded loops.

Lemma 3.1. *For each w-admissible surface $f : S \to X$, there is another w-admissible surface $g : S' \to X$ with $\deg(S') = \deg(S)$ and $-\chi(S') \leq -\chi(S)$ such that $f^{-1}(X_C)$ has no embedded loops.*

Proof. For any embedded loop L in $f^{-1}(X_C)$, if its image is null homotopic in X then the restriction of f to L extends to a disk D. Moreover, since X_C is π_1-injective, we may assume that $f(D) \subset X_C$. In this case we can compress S along L to obtain S' and a map $g : S' \to X$ using the extension of $f|_L$ on D above. Homotop g to push $g(D)$ in the direction away from X_C, we see that $g^{-1}(S')$ has one less embedded loop. S' has all the required properties and $\chi(S') = \chi(S) + 2$, except that S' might have a component Σ that has no w-boundary. To make sure S' meet our convention, in this situation we simply remove this component from S' and $\chi(S' \setminus \Sigma) = \chi(S') - \chi(\Sigma) \geq \chi(S)$ as desired.

If the image of L represents a nontrivial conjugacy class, then we cut S along L to obtain a new surface S' with $\chi(S') = \chi(S)$ and a map $g : S' \to X$ induced by f. Pushing the two new boundary components corresponding to L away from X_C and into $X_A \subset X$, this reduces the number of embedded loops in the preimage of X_C and makes S' into a w-admissible surface, which simply has two more A-boundary components compared to S. In particular $\deg(S') = \deg(S)$. In the special case where L cuts out a component Σ that has no w-boundary, since $f(L)$ represents a nontrivial class in X, we see that $\chi(S) \leq 0$ and hence removing it from S' gives the desired inequality $-\chi(S') \leq -\chi(S)$.

Repeating the two procedures above on each embedded loop in $f^{-1}(X_C)$ completes the proof. □

Now suppose $F := f^{-1}(X_C)$ is a finite disjoint union of embedded proper arcs with endpoints on w-boundary components; See Figure 4. It follows that $S \setminus F$ has two types of boundary components:

(1) A-boundary components, exactly corresponding to those on S;

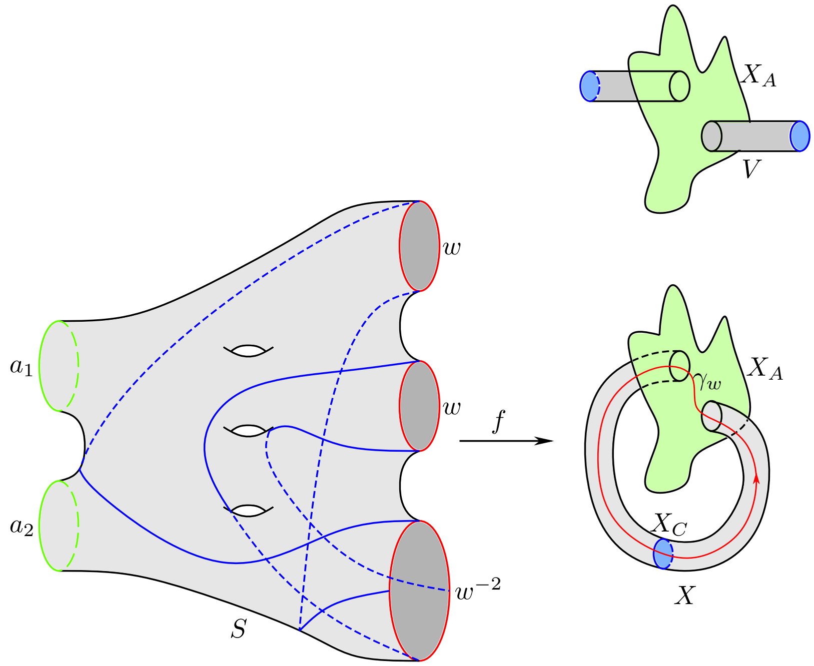

FIGURE 4. $F = f^{-1}(X_C)$ is a set of embedded disjoint proper arcs in the w-admissible surface S after applying Lemma 3.1. After cutting, $S \setminus F$ maps into the thickened vertex space V, which deformation retracts to X_A.

(2) *polygonal boundary components*, each of which is divided into an even number of sides that alternate between arcs in F and segments on some w-boundary of S. Its structure will be discussed in more detail in Section 3.3.

Now $S \setminus F$ maps into the thickened vertex space V and hence each polygonal boundary represents a conjugacy class in A, referred to as its *winding class*.

Definition 3.2 ((simple) normal form, disk-pieces, and annulus-pieces)**.** We refer to each component of $S \setminus F$ as a *piece*. Such a decomposition of S into pieces is called a *normal form* of S. A normal form is *simple* if each piece has exactly one polygonal boundary and is either a disk or an annulus, depending on whether the winding class of the unique polygonal boundary is trivial.

We refer to the two kinds of pieces in a simple normal form as *disk-pieces* and *annulus-pieces* based on their topological type; See Figure 5 an illustration of such pieces.

We can always simplify S so that it admits a simple normal form.

Lemma 3.3. *For any w-admissible surface S, there is a w-admissible surface S' with* $\deg(S') = \deg(S)$ *and* $-\chi(S') \leq -\chi(S)$ *so that S' admits a simple normal form.*

Proof. By the discussion above, we may simplify S so that it admits a normal form, which may not be simple in general. Note that each piece of S has at least one polygonal boundary since each component of S has nonempty w-boundary and F contains no embedded loop. Suppose S has a piece P with at least two polygonal boundary components. Then $\chi(P) \leq 0$ Cut out a collar neighborhood of each polygonal boundary and remove the remaining part of P to obtain a new surface S'. The part ignored has the same homotopy type as P and hence has non-positive Euler characteristic. Hence $-\chi(S') \leq -\chi(S)$. Up to homotopy we may assume the non-polygonal boundary of each collar neighborhood is mapped to the vertex space X_A. This makes S' a w-admissible surface with $\deg(S') = \deg(S)$. The same procedure can be done if P has exactly one polygonal boundary with $\chi(P) < 0$. If there is an annulus piece P where the polygonal boundary has trivial winding class, then the other boundary is a null homotopic loop in X_A, which we cap it off and decreases the negative Euler characteristic. Repeating the procedures above we arrive at a desired w-admissible surface S' in simple normal form. □

3.3. **The structure of a polygonal boundary.** Suppose as above that w is written as a cyclically reduced word $w = a_1 t^{e_1} \cdots a_k t^{e_k}$ with $e_i = \pm 1$ and $a_i \in A$, represented by a tight loop γ in X corresponding to this expression. Recall from Section 3.1 that the edge space X_C cuts γ into k segments $\gamma_1, \ldots, \gamma_k$, equipped with the orientation induced from γ. The segments with the reversed orientation are denoted as $\gamma_1^{-1}, \ldots, \gamma_k^{-1}$. Also recall that segments fall into four types, PP, PN, NP, NN, depending one which side of X_C the segment starts and ends at.

Fix a (disk- or annulus-)piece. Its unique polygonal boundary has an induced orientation. By definition, half of the sides on the polygonal boundary each is a copy of some γ_i or γ_i^{-1} (depending on whether the w-boundary it lies on is positive or negative). We refer to these sides as *arcs*. The other half of the sides are proper arcs in $F = f^{-1}(X_C)$, which we call *turns*, each starting from an arc $\alpha = \gamma_i^{\pm 1}$ to another arc $\alpha' = \gamma_j^{\pm 1}$ for some i, j. By our choice of γ and t, such a turn as a path in X_C starts and ends at the base point b_C and hence is a based loop representing some element $c \in C$, referred to as the *winding number*. We encode the type of each turn as an ordered triple (α, c, α').

Recall that each piece is mapped to the thickened vertex space V, and hence turn is either on the positive side or the negative side of X_C. If a turn has type (α, c, α') and lies on the positive side, then α must end on the positive side and α' must start from the positive side. Similarly if such a turn lies on the negative side. In particular, not every ordered pairs of arcs (α, α') can appear in the triple describing the type of a turn.

There is a pairing of turns in the normal form of a w-admissible surface as pieces are glued together along turns. Two paired turns are on the opposite sides of X_C. The type of a turn determines the type of its paired turn. For instance, a turn of type (γ_i, c, γ_j) must be paired with a turn of type $(\gamma_{j-1}, c^{-1}, \gamma_{i+1})$, indices taken mod k, where the winding number becomes its inverse due to the opposite orientation induced from the two pieces; See Figure 5. We say two such turn types are *paired*.

In below are some basic observations in relation to some crucial assumptions we made. The first is related to the tightness of γ.

Lemma 3.4. *The polygonal boundary of any disk-piece in a simple normal form of a w-admissible surface has at least two turns.*

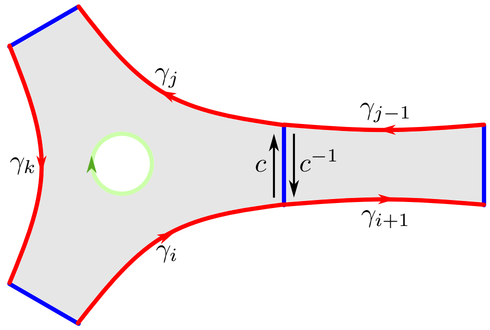

FIGURE 5. An annulus-piece (left) and a disk-piece (right) glued along paired turns that are of types (γ_i, c, γ_j) and $(\gamma_{j-1}, c^{-1}, \gamma_{i+1})$ for some $c \in C$.

Proof. It the polygonal boundary has only one turn, then it has only one arc as well. Suppose the arc is a copy of γ_i.Then the disk-piece provides a homotopy between α_i and the turn relative to the endpoints. Since the turn is a loop in the edge space X_C, this contradicts with our assumption that γ is tight. □

The second is an interpretation of the boundary incompressibility of admissible surfaces.

Lemma 3.5. *In the normal form of a w-admissible boundary incompressible surface, there is no turn of type $(\gamma_i, id, \gamma_i^{-1})$ or $(\gamma_i^{-1}, id, \gamma_i)$ for any i.*

Proof. If there were such a turn of type $(\gamma_i, id, \gamma_i^{-1})$, then it is a proper arc going from a positive w-boundary of S representing w^n to a negative one representing w^{-m} for some $m, n \in \mathbb{Z}_+$. These two boundary components union this proper arc has a collar neighborhood $\Sigma \subset S$ that is a pair of pants. The fact that the winding number of the turn is $id \in C$ implies that the third boundary of Σ represents w^{m-n} (with the orientation induced from Σ). This contradicts the boundary incompressibility of S. □

3.4. **Possible pieces.** For a w-admissible surface in simple normal form, we know by definition it consists of disk-pieces and annulus-pieces. We define explicitly a set $\mathcal{P}$ of disk-pieces and annulus-pieces, which include all pieces that may appear in a simple normal form of some boundary incompressible w-admissible surface. The pieces in $\mathcal{P}$ a priori may not come from a w-admissible surface.

To describe a piece in $\mathcal{P}$, we start by constructing a map into X from an oriented circle divided into $2n$ sides for some $n \geq 1$, which will be the polygonal boundary. Label the sides in a cyclic sequence $(s_1, \ldots, s_{2n})$. For any $1 \leq j \leq n$, let s_{2j-1} be a copy of $\gamma_{i_j}^{e_j}$ with $e_j = \pm 1$ so that it serves as an arc, and let s_{2j} be mapped to a loop in X_C based at b_C representing some $c_j \in C$ to serve as a turn of type $(\gamma_{i_j}^{e_j}, c_j, \gamma_{i_{j+1}}^{e_{j+1}})$.

There are two requirements. Firstly, each turn s_{2j} is on one side of X_C: Either $\gamma_{i_j}^{e_j}$ ends on the positive side and $\gamma_{i_{j+1}}^{e_{j+1}}$ starts from the positive side so that s_{2j} is on the positive side, or s_{2j} is on the negative side defined in a similar manner. Secondly, if $i_j = i_{j+1}$ and

$e_j = -e_{j+1}$, then we require $c_j \neq id_C$ as they are ruled out by boundary incompressibility as in Lemma 3.5.

We say a turn type is *admissible* if it satisfies both requirements. Denote by $\mathcal{T}$ the set of admissible turn types.

We say such a circle with the map described above satisfying both requirements is *an abstract polygonal boundary*, which defines a loop in X. As a consequence of the first requirement, this loop naturally shrinks to a loop in the thickened vertex space V and further to a loop in X_A. Hence each abstract polygonal boundary represents a conjugacy class in A, which we refer to as *the winding class*.

The assumption that γ is tight implies that any abstract polygonal boundary with only one turn (and one arc) has nontrivial winding class, similar to Lemma 3.5.

Now we construct an abstract piece in $\mathcal{P}$ for any given abstract polygonal boundary. If the winding class is trivial, think of the underlying circle as the boundary of a disk, then the map extends to the interior of the disk. This disk with the map into X is an abstract disk-piece in $\mathcal{P}$. If the winding class is nontrivial, consider an annulus where one of the boundary circle is the abstract polygonal boundary and the other is a loop in X_A whose inverse represents the winding class. The map on the annulus is a homotopy, which defines an annulus-piece in $\mathcal{P}$.

The set $\mathcal{P}$ is the set of all abstract disk- or annulus-pieces. Clearly by Lemma 3.5, the polygonal boundary of a genuine disk-piece or annulus-piece has the structure of an abstract polygonal boundary, and the notion of the winding class agrees. Thus each piece that appear in a simple normal form of some boundary incompressible w-admissible surface lies in $\mathcal{P}$. It is not important to us but actually all pieces in $\mathcal{P}$ appear this way.

3.5. **The gluing graph and Euler characteristic.** For any admissible surface S in normal form, there is a *gluing graph* Γ_S that encodes how the surface decomposes into pieces. Each vertex of Γ represents a piece in the normal form and each edge represents a gluing along paired turns of two pieces. By Mayer–Vietoris, we have

$$\chi(S) = \sum_v \chi(v) - \#e,$$

where the summation is taken over all vertices v of Γ_S, $\chi(v)$ is the Euler characteristic of the piece corresponding to v, and $\#e$ is the number of edges in Γ_S.

When S is decomposed in simple normal form, each piece is either a disk or annulus. Hence $v_d := \sum_v \chi(v)$ is the number of disk pieces.

Note that each edge e glues two turns together, so $2\#e$ is the total number of turns. Since on each polygonal boundary, half of the sides are turns and the other half are arcs, the total number of turns is also the total number of arcs. Recall that each copy of the tight loop γ representing w is cut into $|w|$ arcs, so the total number of arcs is $\deg(S)|w|$. Hence

$$2\#e = \#\text{turns} = \deg(S) \cdot |w|.$$

The following lemma summarizes the calculations above.

Lemma 3.6. *For any w-admissible surface S in simple normal form, we have*

$$-\chi(S) = \frac{1}{2}\deg(S)\cdot|w| - v_d, \tag{3.2}$$

where v_d is the total number of disk pieces in S.

4. The LP-duality method

In all theorems, the goal is to establish a lower bound of $-\chi(S)$ by a multiple of $\deg(S)$ for all boundary-incompressible w-admissible surfaces S, which we may assume to be in a simple normal form by Lemma 3.3. In view of formula (3.2), this is equivalent to proving an upper bound of v_d by a multiple of $\deg(S)$.

We prove such inequalities using a method analogous to the weak duality of linear programming.The method was originally developed by the author to prove uniform lower bounds (called spectral gaps) of stable commutator lengths; See [Che20, Section 6.3] and [CH19, Section 5.2]. We explain the approach in our setting in this section, which comes down to the construction of a cost function meeting certain requirements. The theorems of interest are proved using different cost functions that we construct in the remaining sections.

Given a boundary-incompressible w-admissible surfaces S in simple normal form, we can count the total number $t_T \in \mathbb{Z}_{\geq 0}$ of turns that have a given type $T \in \mathcal{T}$, which is nonzero for finitely many turn types by compactness. The collection of numbers $(t_T)_{T\in\mathcal{T}}$ satisfies a *gluing condition*, namely, $t_T = t_{T'}$ if T and T' are paired turn types, since each turn of type T is glued to a turn of type T' in S when pieces are glued together.

A cost function on turns is a map $c : \mathcal{T} \to \mathbb{R}$ that assigns a value to each admissible turn type in $\mathcal{T}$. This naturally induces a cost function on the set $\mathcal{P}$ of possible pieces. Namely, for each $P \in \mathcal{P}$, the value $c(P)$ is the sum of $c(T)$ over all turns T on the polygonal boundary of P and $c(T)$ is determined by the type of the turn T.

We are interested in cost functions meeting two requirements, one relating the cost to v_d, the total number of disk pieces, and the other relating the cost to the degree $\deg(S)$.

Lemma 4.1. *For a cost function $c : \mathcal{T} \to \mathbb{R}$, if the induced cost function on possible pieces satisfies $c(P) \geq \chi(P)$ for any $P \in \mathcal{P}$, then using the notation above we have*

$$\sum_{T\in\mathcal{T}} c_T t_T \geq v_d$$

for any boundary-incompressible w-admissible surfaces S in simple normal form.

Proof. Note that $\chi(P)$ is either one or zero, depending on whether P is a disk-piece or annulus-piece. Hence its sum over all pieces P in the simple normal form is exactly v_d, the number of disk pieces. Hence by assumption we have

$$\sum_P c(P) \geq v_d,$$

where the sum is taken over all pieces in the simple normal form.

On the other hand, by definition $c(P)$ is itself the sum of $c(T)$ over all turns T that appear in the piece P. By changing the order of summation, we see that

$$\sum_P c(P) = \sum_{T\in\mathfrak{T}} c_T t_T.$$

Hence the desired inequality follows. $\square$

Proposition 4.2. *If a cost function $c : \mathfrak{T} \to \mathbb{R}$ satisfies the requirement in Lemma 4.1 and $\sum_{T\in\mathfrak{T}} c_T t_T = \lambda \deg(S)$ for any boundary-incompressible w-admissible surfaces S in simple normal form, where λ is a constant independent of S (but possibly depending on w or the underlying group), then*

$$\lambda \deg(S) \geq v_d.$$

As a consequence, we have

$$-\chi(S) \geq \left(\frac{|w|}{2} - \lambda\right) \deg(S)$$

for all boundary-incompressible w-admissible surfaces S.

Proof. The first inequality is evident by the assumption and Lemma 4.1. Combining this with formula (3.2) we obtain the second inequality for any boundary-incompressible w-admissible surfaces S in simple normal form. For a general boundary-incompressible w-admissible surfaces S, we can put it into simple normal form by Lemma 3.3. $\square$

Remark 4.3. It might look surprising at first that there exists cost functions so that $\sum_{T\in\mathfrak{T}} c_T t_T = \lambda \deg(S)$ for all S. Actually, there are many such functions. Note that both sides are linear functions in variables $(t_T)_{T\in\mathfrak{T}}$ since the degree is a multiple of the total number of turns. There is a huge space of linear functionals on $\mathbb{R}^{\mathfrak{T}}$ (or $\ell^1(\mathfrak{T})$) that vanish on the subspace defined by the gluing conditions, hence adding such a functional to the linear function $\lambda \deg(S)$ yields a cost function satisfying this property.

5. A lower bound of the minimal complexity

As in Section 3, let $H = A\star_C$ be the HNN extension associated to injections $i_P, i_N : C \hookrightarrow A$. In this section, we focus on a cyclically reduced word w taking the special form $w = a_1t^{-1}b_1ta_2t^{-1}b_2\cdots a_mt^{-1}b_mtxt \in H$, where $m \geq 1$, $x \in A$, $a_i \in A \setminus i_P(C)$ and $b_i \in A \setminus i_N(C)$ under the standard presentation (3.1). The goal is to prove the following Theorem 5.3, establishing a lower bound of the minimal complexity of w-admissible boundary-incompressible surfaces. It is a somewhat standard trick to reduced the case of a general word with t-exponent sum ± 1 to this special case; See Lemma 6.2.

The assumptions of Theorem 5.3 involve two conditions, which we now introduce.

Definition 5.1. Given a subgroup $C \leq A$, for some $2 \leq n \leq \infty$, an element $a \in A \setminus C$ is *length-n relatively free to C (n-RF)* if $a^{e_1}c_1 \cdots, a^{e_k}c_k \neq id$ in A for any $k \in \mathbb{Z}_+$, $e_i = \pm 1$, and $c_i \in C$, provided that

(1) $c_i \neq id$ for any i with $e_i = -e_{i+1}$ (indices taken mod k), and
(2) there are no n e_i's of the same sign.

We say the pair (A, C) is n-RF if a is n-RF rel C for all $a \in A \setminus C$.

Roughly speaking, the n-RF condition requires that there is no short relation (measured by the quantifier n) among a and C. In particular, if the subgroup generated by a and C is isomorphic to $\langle a\rangle \star C$, where $\langle a\rangle$ is the cyclic subgroup of A generated by a, then a is n-RF, where $2 \le n \le \infty$ is the order of a.

A weaker condition only puts restrictions on relations where all exponents of a have the same sign.

Definition 5.2. Given a subgroup $C \le A$ and $2 \le n \le \infty$, we say $a \in A \setminus C$ is *n-relatively torsion-free (n-RTF)* in the group-subgroup pair (A, C) if $ac_1 \cdots ac_k \neq id$ for any $c_i \in C$ and any $1 \le k < n$. Note that this is automatically true if $n = 2$ as $a \notin C$.

We say the pair (A, C) is n-RTF if a is n-RTF for all $a \in A \setminus C$.

Clearly if $a \in A \setminus C$ is n-RF rel C then it is also n-RTF.

The n-RTF condition holds in many examples (even for pairs (A, C)), and it has nice inheritance properties in the context of graphs of groups and graph products; see [CH19, Example 5.5, Section 5.4, and Lemma 7.3] for more examples and details on this condition.

Theorem 5.3. *With the notation above, for $w = a_1t^{-1}b_1ta_2t^{-1}b_2 \cdots a_mt^{-1}b_mtxt \in H = A\star_C$, suppose for some $2 \le n \le \infty$*

(1) a_1 is n-RF rel $i_P(C)$ and b_m is n-RF $i_N(C)$, and

(2) each a_i is n-RTF in $(A, i_P(C))$ and each b_i is n-RTF in $(A, i_N(C))$ for all $1 \le i \le m$.

Then

$$-\chi(S) \ge \left(1 - \frac{1}{n}\right) \deg(S)$$

for any w-admissible boundary-incompressible surface S.

It is worth noting that only the n-RTF assumption is needed to prove the analogous estimate ([CH19, Theorem 5.8]) in the context of stable commutator length in a graph of groups.

The following corollary explains how estimates of the complexity of w-admissible surfaces can be applied to obtain injectivity of subgroup under the quotient map. This is almost identical to a result of Fenn–Rourke [FR96, Theorem 4.1] carefully explaining and generalizing Klyachko's method [Kly93], except that in their statement each a_i (resp. b_i) is assumed to be free relative to $i_P(C)$ (resp. $i_N(C)$)[3], while we only need the weaker condition ∞-RF on a_1 and b_m and much weaker conditions on the other a_i's and b_i's. Klyachko's Theorem 6.7 and other Freiheitssatz theorems quickly follow from this result after applying a standard algebraic trick (Lemma 6.2), which we will explain in Section 6.

Corollary 5.4. *For the HNN extension $H = A\star_C$ and the word w satisfying the assumptions in Theorem 5.3 with $n = \infty$, the natural map $A \to H\langle\langle w\rangle\rangle$ induced by the inclusion $A \hookrightarrow H$ is injective.*

[3]An element $a \in A \setminus C$ is free relative to C if the subgroup generated by a and C is $\mathbb{Z} \star C$, which implies that a is ∞-RF

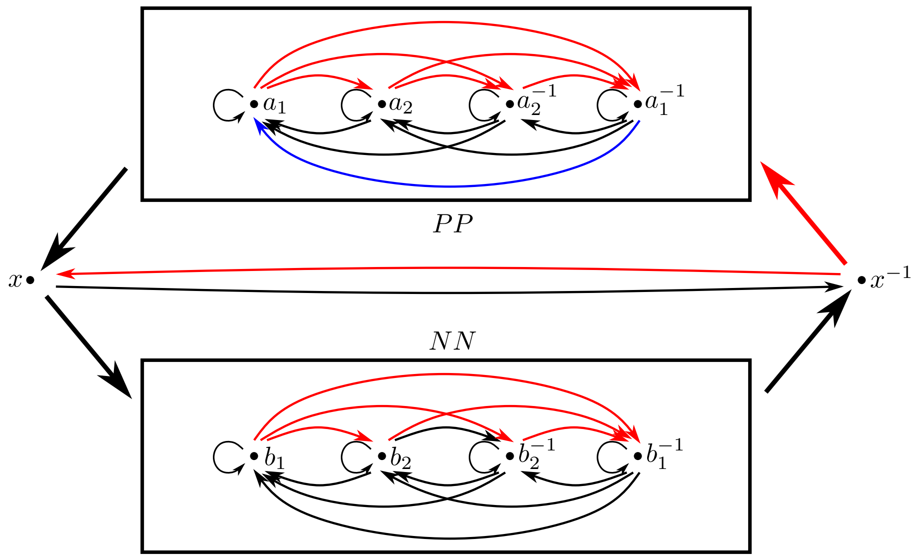

FIGURE 6. The directed graph encoding admissible turns between arcs, where the two rectangular boxes enclose all arcs of type PP and NN respectively. Each of the four thick big arrow represents a collection of edges connecting the vertex represented by x or x^{-1} with all vertices in a rectangular box. Under the cost function c, red edges have cost 1 when $n = \infty$, and the blue edge is the only one with negative cost.

Proof. Suppose the natural map is not injective, that is, there is some $a \neq id \in A$ that lies in $\langle\langle w \rangle\rangle$. As seen in Example 2.5 (with $H = A \star \mathbb{Z}$), this gives rise to an equation (2.1), which provides a w-admissible surface S of degree $\deg(S) = k$ with $-\chi(S) = k - 1$ (as it is a sphere with $k+1$ disks removed) for some $k \in \mathbb{Z}_+$. Moreover, as explained in Example 2.5, when k is minimal among all equations of this form, S is boundary-incompressible. Hence by Theorem 5.3 (with $n = \infty$), we have $k - 1 = -\chi(S) \geq \deg(S) = k$, which leads to a contradiction. Thus the natural map must be injective. $\square$

In the rest of this section, we prove Theorem 5.3 using the LP-duality method introduced in Section 4. We first define a cost function $c : \mathcal{T} \to \mathbb{R}$ and then verify the desired properties.

Note that in this case the tight loop γ corresponding to w is decomposed into $|w| = 2m+1$ arcs, which we denote suggestively by $\gamma_1 = a_1, \gamma_2 = b_1, \cdots, \gamma_{2m-1} = a_m, \gamma_{2m} = b_m, \gamma_{2m+1} = x$. Denote the arcs on $\bar{\gamma}$ by $\gamma_{2m+1}^{-1} = x^{-1}, \gamma_{2m}^{-1} = b_m^{-1}, \gamma_{2m-1}^{-1} = a_m^{-1}, \cdots, \gamma_2^{-1} = b_1^{-1}, \gamma_1^{-1} = a_1^{-1}$. The arcs $a_i^{\pm 1}$ (resp. $b_i^{\pm 1}$) are of type PP (resp. NN), and the arc x (resp. x^{-1}) is the only arc of type PN (resp. NP).

A key observation here is that there is no admissible turn going from any $a_i^{\pm 1}$ to $b_j^{\pm 1}$ or vice versa. This is indicated in the directed graph in Figure 6 for $m = 2$, where any admissible turn type $(\gamma_i^{\pm 1}, \kappa, \gamma_j^{\pm 1})$ for some $\kappa \in C$ has the ordered pair $(\gamma_i^{\pm 1}, \gamma_j^{\pm 1})$ represented as an oriented edge.

5.1. **The cost function.** We define the cost function in a way so that the cost of an admissible turn $(\gamma_i^{\pm 1}, \kappa, \gamma_j^{\pm 1}) \in \mathcal{T}$ only depends on the ordered pair $(\gamma_i^{\pm 1}, \gamma_j^{\pm 1})$. Hence we will simply define the cost $c(\gamma_i^{\pm 1}, \gamma_j^{\pm 1})$ below for all $1 \le i, j \le 2m+1$. To simplify the notation, we write $c_{i,j} = c(\gamma_i, \gamma_j)$, $c_{i,-j} = c(\gamma_i, \gamma_j^{-1})$, and similarly for $c_{-i,j}$ and $c_{-i,-j}$. The cost for some pairs (say (a_1, b_1)) is irrelevant if the pair does not appear in any admissible turn type.

For $1 \le i, j \le 2m+1$, we define

$$c_{i,j} = \begin{cases} 1 - \frac{1}{n}, & i < j; \\ \frac{1}{n}, & j \le i < 2m+1; \\ 0, & i = 2m+1. \end{cases}$$

$$c_{i,-j} = \begin{cases} 0, & i = 2m+1; \\ 0, & j = 2m+1; \\ 0, & i = j = 2m; \\ 1 - \frac{1}{n}, & \text{otherwise.} \end{cases}$$

$$c_{-i,j} = \begin{cases} 1, & i = j = 2m+1; \\ \frac{2}{n} - 1, & i = j = 1; \\ \frac{1}{n}, & \text{otherwise.} \end{cases}$$

$$c_{-i,-j} = \begin{cases} 1 - \frac{1}{n}, & i > j; \\ \frac{1}{n}, & i \le j < 2m+1; \\ 0, & j = 2m+1. \end{cases}$$

In the case $n = \infty$, turns with cost 1 are represented by red edges in Figure 6 illustrating the case of $m = 2$, where one can observe that most oriented loops contain at least one such edges. The reason to choose this cost function would become clearer if one restricts attention to the cost of turns among b_i's (resp. a_i's); See the red edges in the rectangular boxes in Figure 6 as well as Sections 5.2.1 and 5.2.2 below.

5.2. **Comparison with χ.** In this section, we prove the following lemma to verify one of the conditions in Proposition 4.2.

Lemma 5.5. *For any piece $P \in \mathcal{P}$ we have $c(P) \ge \chi(P)$.*

Recall that the polygonal boundary of any piece P is a cyclic sequence of arcs connected by admissible turns. Thus we can view it as an oriented loop in the graph Γ, where vertices are arcs and oriented edges are admissible turns, shown in Figure 6. Each oriented edge of the loop has a cost according to the definition of c above. This allows us to define the cost of any oriented path (and loop) as the sum of the cost of its edges.

The goal is to show that the total cost of the loop corresponding to the polygonal boundary of P is non-negative, and moreover no less than 1 if P is a disk-piece (i.e. the polygonal boundary has trivial winding class).

We have the following basic observation.

Lemma 5.6. *Any oriented loop in Γ falls into one of the following three types:*

(1) It is a loop supported on $b_i^{\pm 1}$'s.
(2) It is a loop supported on $a_i^{\pm 1}$'s.
(3) It passes through both x and x^{-1}, and it contains at least one path from x to x^{-1} through $a_i^{\pm 1}$'s.

Proof. As we observed earlier, x^{-1} is the only arc of type NP and x is the only arc of type PN, and there is no oriented edge from $a_i^{\pm 1}$ to $b_j^{\pm 1}$ and vice versa. Thus either the loop passes through both x and x^{-1} or it is disjoint from both.

In the latter case, the loop is supported either only on $a_i^{\pm 1}$'s or only on $b_i^{\pm 1}$'s. These are the first two cases in the lemma.

In the other case, there are finitely many x's and x^{-1}'s on the loop, and they must alternate as there is no path from any $a_i^{\pm 1}$ to x or from any $b_i^{\pm 1}$ to x^{-1}. Hence there must be a path from x to x^{-1} through a bunch of $a_i^{\pm 1}$'s, which is the last case of the lemma. □

We prove Lemma 5.5 by examining these three cases respectively. In the process, we will use the following basic observation repeatedly as our (only) way of using the n-RF condition.

Lemma 5.7. *Consider a graph with two vertices u and v and all four possible distinct oriented edges. Suppose for some $2 \leq n \leq \infty$,*

(1) both edges (u, u) and (v, v) have cost at least $1/n$, and
(2) the sum of the cost of (u, v) and (v, u) is at least $1/n$.

Then any oriented loop visiting u (resp. v) at least n times has total cost at least 1.

Proof. By symmetry, it suffices to consider a loop visiting u at least n times. Such a loop decomposes into sub-loops each visiting u exactly once. There are exactly two types of such sub-loops:

(1) Either it has exactly one edge (u, u),
(2) or it starts with (u, v), ends with (v, u), and has s copies of (v, v) in the middle for some $s \geq 0$.

In the first case, such a sub-loop has cost $1/n$, and in the second case, such it has cost at least $1/n + s/n \geq 1/n$. Hence each sub-loop has cost at least $1/n$ no matter the type. The number of such sub-loops in the decomposition is the number of times that the given loop visits u, which is at least n by assumption. Hence the total cost of the loop is at least 1. □

5.2.1. *Case I: Only involving $b_i^{\pm 1}$'s.*

Lemma 5.8. *Under the assumptions of Theorem 5.3, suppose the loop in Γ corresponding to the polygonal boundary of a piece $P \in \mathcal{P}$ is supported on $b_i^{\pm 1}$'s. Then $c(P) \geq \chi(P)$.*

Proof. Define a linear order on the $b_i^{\pm 1}$'s as follows:

$$b_1 \prec b_2 \prec \cdots \prec b_m \prec b_m^{-1} \prec \cdots \prec b_2^{-1} \prec b_1^{-1}.$$

Then the cost c we defined has the property that

$$c(b_i^{e_i}, b_j^{e_j}) = \begin{cases} 1 - \frac{1}{n}, & \text{if } b_i^{e_i} \prec b_j^{e_j} \text{ and } (b_i^{e_1}, b_j^{e_j}) \neq (b_m, b_m^{-1}) \\ \frac{1}{n}, & \text{if } b_i^{e_i} \succeq b_j^{e_j}, \\ 0, & \text{if } (b_i^{e_i}, b_j^{e_j}) = (b_m, b_m^{-1}), \end{cases}$$

for any $1 \leq i, j \leq m$ and $e_i, e_j = \pm 1$.

If the loop passes through at least two distinct vertices, then it contains one oriented edge that is ascending in the order $\prec$ and another that is descending.

(1) If an ascending edge is not (b_m, b_m^{-1}), then these two edges contribute $(1 - 1/n) + 1/n = 1$ to $c(P)$ and all the other edges have non-negative cost, so $c(P) \geq 1 \geq \chi(P)$.
(2) If all ascending edges in the loop are (b_m, b_m^{-1}), then all arcs on the polygonal boundary are b_m and b_m^{-1}. In this case, the winding class takes the form

$$b_m^{e_1}\kappa_1 b_m^{e_2}\kappa_2 \cdots b_m^{e_s}\kappa_s$$

for some $s \geq 2$, $e_i = \pm 1$, and $\kappa_i \in i_N(C)$. As turns are admissible, we have $\kappa_i \neq id_A$ whenever $e_i = -e_{i+1}$, indices taken mod s. By assumption, b_m is n-RF rel $i_N(C)$, so the above word cannot be the identity unless it contains at least n copies of b_m (resp. b_m^{-1}). In the former case, the winding class is nontrivial, we have $\chi(P) = 0$, so the desired inequality clearly holds; In the exceptional case, the total cost is at least 1 by Lemma 5.7 and hence no less than $\chi(P)$.

The remaining case is when the loop visits the same vertex, say $b_i^{e_i}$ throughout. Then the winding class is

$$b_i\kappa_1 \cdots b_i\kappa_s$$

up to conjugation and possibly taking the inverse (when $e_i = -1$), where s is the length of the loop. The winding class is nontrivial unless $s \geq n$ by our n-RTF assumption. Hence either $\chi(P) = 0$ and the inequality holds trivially or $c(P) = s/n \geq 1 \geq \chi(P)$.

Thus we have $c(P) \geq \chi(P)$ as desired for all such P. □

5.2.2. *Case II: Only involving $a_i^{\pm 1}$'s.* Next we show

Lemma 5.9. *Under the assumptions of Theorem 5.3, suppose the loop in Γ corresponding to the polygonal boundary of a piece $P \in \mathcal{P}$ is supported on $a_i^{\pm 1}$'s. Then $c(P) \geq \chi(P)$.*

Similar to the proof of the previous case, we introduce a linear order on $a_i^{\pm 1}$'s as follows:

$$a_1 \prec a_2 \prec \cdots \prec a_m \prec a_m^{-1} \prec \cdots \prec a_2^{-1} \prec a_1^{-1}.$$

Then the cost c we defined has the property that

$$c(a_i^{e_i}, a_j^{e_j}) = \begin{cases} 1 - \frac{1}{n}, & \text{if } a_i^{e_i} \prec a_j^{e_j}, \\ \frac{1}{n}, & \text{if } a_i^{e_i} \succeq a_j^{e_j} \text{ and } (a_i^{e_1}, a_j^{e_j}) \neq (a_1^{-1}, a_1) \\ \frac{2}{n} - 1, & \text{if } (a_i^{e_i}, a_j^{e_j}) = (a_1^{-1}, a_1), \end{cases}$$

for any $1 \leq i, j \leq m$ and $e_i, e_j = \pm 1$.

This case is slightly more complicated than the previous one since $c(a_1^{-1}, a_1) = \frac{2}{n} - 1$ could be negative.

Suppose the loop corresponding to the polygonal boundary of a piece P as in Lemma 5.9 contains s copies of the edge (a_1^{-1}, a_1), where $s \in \mathbb{Z}_{\geq 0}$. Then the complement of these s edges in the loop consists of s oriented paths from a_1 to a_1^{-1}, where each path does not contain the edge (a_1^{-1}, a_1).

Lemma 5.10. *Let ρ be an oriented path from a_1 to a_1^{-1} supported on $a_i^{\pm 1}$'s so that ρ does not contain the edge (a_1^{-1}, a_1). Then the cost $c(\rho) \geq 1 - 1/n$. Moreover, we have $c(\rho) \geq 2(1 - 1/n)$ if ρ visits any vertex other than a_1 and a_1^{-1}.*

Proof. By our assumption, each edge in ρ has non-negative cost. As $a_1 \prec a_1^{-1}$, the path ρ contains at least one ascending edge, which contributes $1 - 1/n$ to the cost, so $c(\rho) \geq 1 - 1/n$. If ρ visits any vertex $a_i^{\pm 1}$ other than a_1 and a_1^{-1}, then ρ contains at least two ascending edges since $a_1 \prec a_i^{\pm 1} \prec a_1^{-1}$. Thus in this case we have $c(\rho) \geq 2(1 - 1/n)$. □

Now we are ready to prove Lemma 5.9.

Proof. By the discussion above, suppose the loop corresponding to the polygonal boundary of P contains s copies of the edge (a_1^{-1}, a_1).

If $s = 0$, then each edge in the loop has non-negative cost. If the loop visits at least two distinct vertices, then there is an ascending edge and a descending edge with respect to the order $\prec$, so $c(P) \geq (1 - 1/n) + 1/n = 1$. If the loop keeps visiting the same vertex, say $a_i^{e_i}$, then similar to the last part of the proof of Lemma 5.8, either $\chi(P) = 0 \leq c(P)$ or $c(P) \geq 1 \geq \chi(P)$ by our assumption.

If $s \geq 1$, consider the s paths from a_1 to a_1^{-1} obtained by removing the s copies of (a_1^{-1}, a_1) from the loop. By Lemma 5.10, we have

$$c(P) \geq s\left(\frac{2}{n} - 1\right) + s(1 - \frac{1}{n}) = \frac{s}{n} \geq 0.$$

Moreover, if at least one of the paths visits some vertex other than a_1 or a_1^{-1}, then

$$c(P) \geq \frac{s}{n} + 1 - \frac{1}{n} = \frac{s-1}{n} + 1 \geq 1 \geq \chi(P).$$

So the remaining case is where the entire loop only visits a_1 and a_1^{-1}. Then we are in a situation to apply Lemma 5.7, noting that the cost of (a_1, a_1^{-1}) and (a_1^{-1}, a_1) sums to $1/n$.

In this case, the winding class of the polygonal boundary is

$$a_1^{e_1}\kappa_1 a_1^{e_2}\kappa_2\cdots a_1^{e_{s'}}\kappa_{s'}$$

where $e_i = \pm 1$, $s' \geq 2s$, and $\kappa_i \in i_P(C)$. Since the turns are admissible, $\kappa_i \neq id_A$ if $e_i = -e_{i+1}$. Thus it is nontrivial by the n-RF condition on a_1 unless it contains at least n copies of a_1 (resp. a_1^{-1}). If it is nontrivial, then $\chi(P) = 0 \leq c(P)$. In the exceptional case, we have the total cost $c(P)$ is at least 1 by Lemma 5.7 and hence no less than $\chi(P)$. Hence in any case we have $c(P) \geq \chi(P)$. □

5.2.3. *Case III: Involving both x and x^{-1}.* Now we prove

Lemma 5.11. *Under the assumptions of Theorem 5.3, suppose the loop in Γ corresponding to the polygonal boundary of a piece $P \in \mathcal{P}$ passes through both x and x^{-1}. Then $c(P) \geq \chi(P)$.*

As shown in Lemma 5.6, the loop corresponding to such P must contains a path from x^{-1} to x through $a_i^{\pm 1}$'s. The key is to show

Lemma 5.12. *Any path from x^{-1} to x through $a_i^{\pm 1}$'s has cost at least 1.*

To prove this, we need the following observation:

Lemma 5.13. *Any path from x^{-1} to a_1^{-1} through $a_i^{\pm 1}$'s has cost at least $1-1/n$. The same holds for any path from a_1 to x through $a_i^{\pm 1}$'s.*

Proof. Note that by the definition of the cost c, we have $c(x^{-1}, a_i) = 1/n$ and $c(x^{-1}, a_i^{-1}) = 1 - 1/n$. The cost among $a_i^{\pm 1}$'s is described in Section 5.2.2 using the linear order $\prec$.

Consider a path from x^{-1} to a_1^{-1}; See Figure 6 for an illustration. We may assume that it only visits a_1^{-1} once, since otherwise it is such a path concatenated with several loops supported on $a_i^{\pm 1}$'s and each such loop has non-negative cost by Lemma 5.9. Then the path does not contain the edge (a_1^{-1}, a_1) and thus all edges involved have non-negative cost. Now the first edge of the path is either of the form (x^{-1}, a_i^{-1}) for some i, or (x^{-1}, a_i) for some i. In the former case, the first edge already has cost $1 - 1/n$ so the cost of the entire path is no smaller. In the latter case, there must be an ascending edge among $a_i^{\pm 1}$'s with respect to the order $\prec$, which has cost $1 - 1/n$. Hence the cost of such a path is at least $1 - 1/n$ in any case.

The case for a path from a_1 to x is similar (actually symmetric by reversing the orientation of the path), noting that $c(a_i, x) = 1 - 1/n$ and $c(a_i^{-1}, x) = 1 - 1/n$. Thus we omit the detailed proof. □

Proof of lemma 5.12. Consider a path from x^{-1} to x through $a_i^{\pm 1}$'s, that is, x and x^{-1} only appear at the two ends and all other vertices on the path are $a_i^{\pm 1}$'s. If the path is simply the edge (x^{-1}, x), then its cost is 1 by definition.

Now we assume the path passes through some a_i's. Suppose the path contains s copies of the edge (a_1^{-1}, a_1), where $s \in \mathbb{Z}_{\geq 0}$.

If $s = 0$, then all edges in the path has non-negative cost. The two edges at the two ends of the path has total cost at least 1 unless they are of the form (x^{-1}, a_i) and (a_j^{-1}, x)

respectively for some i, j. In this case, there is a subpath from a_i to a_j^{-1}. Then there must be an ascending edge as $a_i \prec a_j^{-1}$, which has cost $1 - 1/n$. Then the total cost of the entire path is at least $2/n + (1 - 1/n) \geq 1$.

If $s \geq 1$, then the complement of these s edges in the path consists of a subpath from x^{-1} to a_1^{-1}, $(s-1)$ subpaths from a_1 to a_1^{-1}, and a subpath from a_1^{-1} to x. By Lemmas 5.10 and 5.13, each of these subpaths has cost at least $1 - 1/n$. Therefore, the cost of the entire path is at least

$$s\left(\frac{2}{n} - 1\right) + (s+1)\left(1 - \frac{1}{n}\right) = \frac{s-1}{n} + 1 \geq 1.$$

□

Now we are ready to prove Lemma 5.11.

Proof. By the structure revealed in Lemma 5.6, the loop corresponding to P alternates between paths from x to x^{-1} through $b_i^{\pm 1}$'s and paths from x^{-1} to x through $a_i^{\pm 1}$'s. Note that the only edge that possibly has negative cost is (a_1^{-1}, a_1), the cost of the each path from x to x^{-1} through $b_i^{\pm 1}$'s is non-negative. On the other hand, any path from x^{-1} to x through $a_i^{\pm 1}$'s has cost at least 1 by Lemma 5.12. Hence the total cost is no less than 1 as desired. □

5.2.4. *Proof of Lemma 5.5.* Putting all three cases together, we can now prove Lemma 5.5.

Proof of Lemma 5.5. Consider any piece $P \in \mathcal{P}$. By Lemma 5.5, the loop corresponding to the polygonal boundary of P falls into one of three cases. By Lemmas 5.8, 5.9 and 5.11, in each case we have $c(P) \geq \chi(P)$. □

5.3. **The sum $\sum_{T \in \mathcal{T}} c_T t_T$.** Now we turn to verifying the following computation, as the other condition that we need to apply Proposition 4.2. Recall that t_T is the number of turns of type T for each $T \in \mathcal{T}$. We extend it to all turn types T by setting $t_T = 0$ for all T not admissible.

Lemma 5.14. *For any boundary-incompressible w-admissible surface S in simple normal form, we have*

$$\sum_{T \in \mathcal{T}} c_T t_T = \left(\frac{|w|}{2} - 1 + \frac{1}{n}\right) \deg(S).$$

To simplify the notation, for any $1 \leq i, j \leq 2m+1$, let $t_{i,j} = \sum_{\kappa \in C} t_{(\gamma_i, \kappa, \gamma_j)}$, $t_{i,-j} = \sum_{\kappa \in C} t_{(\gamma_i, \kappa, \gamma_j^{-1})}$, and similarly for $t_{-i,j}$ and $t_{-i,-j}$. For convenience, we also set $t_{0,j} = t_{i,0} = 0$ for any $i, j \in \mathbb{Z}$.

Since $c(\gamma_i, \kappa, \gamma_j) = c_{i,j}$ does not depend on κ, we have

$$\sum_{T \in \mathcal{T}} c_T t_T = \sum_{i,j=1}^{2m+1} c_{i,j} t_{i,j} + \sum_{i,j=1}^{2m+1} c_{i,-j} t_{i,-j} + \sum_{i,j=1}^{2m+1} c_{-i,j} t_{-i,j} + \sum_{i,j=1}^{2m+1} c_{-i,-j} t_{-i,-j}. \tag{5.1}$$

For the computation below, we use the following basic facts.

Lemma 5.15. *We have*

(1) $t_{i,j} = t_{j-1,i+1}$,
(2) $t_{i,-j} = t_{-(j+1),i+1}$,
(3) $t_{-i,j} = t_{j-1,-(i-1)}$,
(4) $t_{-i,-j} = t_{-(j+1),-(i-1)}$,

for any $1 \le i, j \le 2m+1$, *where each* $i \pm 1$ *or* $j \pm 1$ *is interpreted mod* $2m+1$.

Proof. Recall that, by the gluing condition, we have $t_T = t_{T'}$ for paired turn types T, T'. A turn of type $(\gamma_i, \kappa, \gamma_j)$ is paired with $(\gamma_{j-1}, \kappa^{-1}, \gamma_{i+1})$ for any $\kappa \in C$. Taking the sum over all $\kappa \in C$ proves the first equality.

The others hold for a similar reason. For instance, a turn of type $(\gamma_i, \kappa, \gamma_j^{-1})$ is paired with $(\gamma_{j+1}^{-1}, \kappa^{-1}, \gamma_{i+1})$ for any $\kappa \in C$. $\square$

Lemma 5.16. *For any* $1 \le i \le 2m+1$, *we have*

$$\sum_{j=-(2m+1)}^{2m+1} t_{i,j} = \sum_{j=-(2m+1)}^{2m+1} t_{j,i} = \frac{1}{2}\deg(S).$$

The same holds with i *replaced by* $-i$.

Proof. The first summation is the total number of turns starting from γ_i, which is exactly the total number of copies of γ_j that appear on ∂S. Equivalently, it is the number of copies of γ on ∂S, which is $\deg_+(S)$. By Lemma 2.3, we have $\deg_+(S) = \frac{1}{2}\deg(S)$ since $p(w) = 1$ for the projection $p : H \to \mathbb{Z}$ taking the standard generator t to 1.

The second summation is the total number of turns ending at γ_i and thus is equal to the previous one.

If we replace i by $-i$, the same argument above holds with γ_i and $\deg +(S)$ replaced by γ_i^{-1} and $\deg_-(S)$ respectively. $\square$

Lemma 5.17. *For any* $1 \le i \le 2m$, *we have* $t_{i,i+1} = 0$, $t_{-(i+1),-i} = 0$, *and* $t_{2m+1,1} = t_{-1,-(2m+1)} = 0$.

Proof. The turn type $(\gamma_i, \kappa, \gamma_{i+1})$ is not admissible for any $\kappa \in C$, since if γ_i ends on one side of the edge space X_C then γ_{i+1} starts on the other side. This shows $t_{i,i+1} = 0$. The others hold for a similar reason. $\square$

Now we compute the four summations on the right hand side of equation (5.1), starting with the first two. By definition, we have

$$\begin{aligned}
\sum_{i,j=1}^{2m+1} c_{i,j} t_{i,j} &= \sum_{i=1}^{2m}\sum_{j>i}\left(1-\frac{1}{n}\right)t_{i,j} + \sum_{i=1}^{2m}\sum_{1\le j\le i}\frac{1}{n}t_{i,j} \\
&= \left(1-\frac{2}{n}\right)\sum_{i=1}^{2m}\sum_{j>i} t_{i,j} + \frac{1}{n}\sum_{i=1}^{2m}\sum_{j=1}^{2m+1} t_{i,j}, \qquad (5.2)\\
&= \left(1-\frac{2}{n}\right)\mathrm{I}_1 + \frac{1}{n}\mathrm{II}_1,
\end{aligned}$$

where $\mathrm{I}_1 = \sum_{i=1}^{2m} \sum_{j>i} t_{i,j}$ and $\mathrm{II}_1 = \sum_{i=1}^{2m} \sum_{j=1}^{2m+1} t_{i,j}$.

We also have

$$\begin{aligned}
\sum_{i,j=1}^{2m+1} c_{i,-j} t_{i,-j} &= \sum_{i=1}^{2m} \sum_{j=1}^{2m} \left(1 - \frac{1}{n}\right) t_{i,-j} - \left(1 - \frac{1}{n}\right) t_{2m,-2m} \\
&= \frac{1}{n} \sum_{i=1}^{2m} \sum_{j=1}^{2m+1} t_{i,-j} + \sum_{i=1}^{2m} \sum_{j=1}^{2m} \left(1 - \frac{2}{n}\right) t_{i,-j} - \frac{1}{n} \sum_{i=1}^{2m} t_{i,-(2m+1)} \\
&\quad - \left(1 - \frac{1}{n}\right) t_{2m,-2m} \\
&= \frac{1}{n} \mathrm{II}_2 + \left(1 - \frac{2}{n}\right) \mathrm{I}_2 - \frac{1}{n} \mathrm{III}_1 - \left(1 - \frac{1}{n}\right) t_{2m,-2m},
\end{aligned} \tag{5.3}$$

where $\mathrm{I}_2 = \sum_{i=1}^{2m} \sum_{j=1}^{2m} t_{i,-j}$, $\mathrm{II}_2 = \sum_{i=1}^{2m} \sum_{j=1}^{2m+1} t_{i,-j}$, and $\mathrm{III}_1 = \sum_{i=1}^{2m} t_{i,-(2m+1)}$.

Putting them together, we deduce

Lemma 5.18.

$$\sum_{i,j=1}^{2m+1} c_{i,j} t_{i,j} + \sum_{i,j=1}^{2m+1} c_{i,-j} t_{i,-j} = \left(1 - \frac{2}{n}\right) (I_1 + I_2) + \frac{m}{n} \deg(S) - \frac{1}{n} III_1 - \left(1 - \frac{1}{n}\right) t_{2m,-2m}.$$

Proof. Note by Lemma 5.16 we have $\sum_{j=1}^{2m+1} t_{i,j} + \sum_{j=1}^{2m+1} t_{i,-j} = \frac{1}{2} \deg(S)$ for any $1 \leq i \leq 2m$. Hence

$$\mathrm{II}_1 + \mathrm{II}_2 = 2m \cdot \frac{1}{2} \deg S = m \deg(S),$$

and the result follows by combining equations (5.2) and (5.3). □

Similarly, we compute the third and fourth summation in (5.1) by definition.

$$\sum_{i,j=1}^{2m+1} c_{-i,j} t_{-i,j} = \frac{1}{n} \sum_{i,j=1}^{2m+1} t_{-i,j} + \left(1 - \frac{1}{n}\right) t_{-(2m+1),2m+1} - \left(1 - \frac{1}{n}\right) t_{-1,1}. \tag{5.4}$$

(5.5)

$$\begin{aligned}
\sum_{i,j=1}^{2m+1} c_{-i,-j} t_{-i,-j} &= \sum_{i=1}^{2m+1} \sum_{1 \leq j < i} \left(1 - \frac{1}{n}\right) t_{-i,-j} + \frac{1}{n} \sum_{i=1}^{2m+1} \sum_{j=i}^{2m} t_{-i,-j} \\
&= \sum_{i=1}^{2m+1} \sum_{1 \leq j < i} \left(1 - \frac{2}{n}\right) t_{-i,-j} + \frac{1}{n} \sum_{i=1}^{2m+1} \sum_{j=1}^{2m+1} t_{-i,-j} - \frac{1}{n} \sum_{i=1}^{2m+1} t_{-i,-(2m+1)} \\
&= \left(1 - \frac{2}{n}\right) \mathrm{I}_3 + \frac{1}{n} \sum_{i=1}^{2m+1} \sum_{j=1}^{2m+1} t_{-i,-j} - \frac{1}{n} \mathrm{III}_2,
\end{aligned}$$

where $\mathrm{I}_3 = \sum_{i=1}^{2m+1} \sum_{1 \leq j < i} t_{-i,-j}$ and $\mathrm{III}_2 = \sum_{i=1}^{2m+1} t_{-i,-(2m+1)}$.

Putting these two together, we get

Lemma 5.19.

$$\sum_{i,j=1}^{2m+1} c_{-i,j}t_{-i,j} + \sum_{i,j=1}^{2m+1} c_{-i,-j}t_{-i,-j} = \frac{2m+1}{2n}\deg(S) + \left(1-\frac{2}{n}\right) I_3 - \frac{1}{n} III_2$$
$$+ \left(1-\frac{1}{n}\right) t_{-(2m+1),2m+1} - \left(1-\frac{1}{n}\right) t_{-1,1}$$

Proof. Note that by Lemma 5.16 we have

$$\sum_{j=1}^{2m+1} t_{-i,j} + \sum_{j=1}^{2m+1} t_{-i,-j} = \frac{1}{2}\deg(S)$$

for any $1 \le i \le 2m+1$. Hence

$$\sum_{i,j=1}^{2m+1} t_{-i,j} + \sum_{i=1}^{2m+1}\sum_{j=1}^{2m+1} t_{-i,-j} = \frac{2m+1}{2}\deg(S),$$

and the result follows by combining equations (5.4) and (5.5). □

We now combine Lemmas 5.18 and 5.19 to complete the computation by equation 5.1. To simplify the results, we make one further observation.

Lemma 5.20.

$$I_1 + I_2 + I_3 = \frac{1}{2}\left[(2m-1)\deg(S) + t_{2m+1,-(2m+1)} + t_{-1,1}\right].$$

Proof. Note by Lemma 5.15 we have

$$\mathrm{I}_1 = \sum_{1\le i<j\le 2m+1} t_{i,j} = \sum_{1\le i<j\le 2m+1} t_{j-1,i+1}$$

and

$$2\mathrm{I}_1 = \sum_{1\le i<j\le 2m+1} t_{i,j} + \sum_{1\le i<j\le 2m+1} t_{j-1,i+1} = \sum_{i=1}^{2m}\sum_{j=2}^{2m+1} t_{i,j} + \sum_{i=1}^{2m} t_{i,i+1}.$$

Since $t_{i,i+1} = 0$ for all $1 \le i \le 2m$ by Lemma 5.17, we have

$$2\mathrm{I}_1 = \sum_{i=1}^{2m}\sum_{j=2}^{2m+1} t_{i,j}.$$

A similar computation shows

$$2\mathrm{I}_3 = \sum_{i=2}^{2m+1}\sum_{j=1}^{2m} t_{-i,-j}.$$

Lemma 5.15 also implies

$$\mathrm{I}_2 = \sum_{i=1}^{2m}\sum_{j=1}^{2m} t_{i,-j} = \sum_{i=2}^{2m+1}\sum_{j=2}^{2m+1} t_{-i,j}.$$

Combining these, we see that

$$\begin{aligned}
2(\mathrm{I}_1+\mathrm{I}_2+\mathrm{I}_3) &= \sum_{i=1}^{2m}\sum_{j=2}^{2m+1} t_{i,j} + \sum_{i=1}^{2m}\sum_{j=1}^{2m} t_{i,-j} + \sum_{i=2}^{2m+1}\sum_{j=2}^{2m+1} t_{-i,j} + \sum_{i=2}^{2m+1}\sum_{j=1}^{2m} t_{-i,-j} \\
&= \sum_{i=-(2m+1)}^{2m+1}\ \sum_{j=-(2m+1)}^{2m+1} t_{i,j} \\
&\quad - \sum_{i=-(2m+1)}^{2m+1} (t_{i,1}+t_{i,-(2m+1)}) - \sum_{j=-(2m+1)}^{2m+1} (t_{2m+1,j}+t_{-1,j}) \\
&\quad + t_{2m+1,1} + t_{-1,1} + t_{2m+1,-(2m+1)} + t_{-1,-(2m+1)}.
\end{aligned}$$

By Lemma 5.17 we have $t_{2m+1,1} = t_{-1,-(2m+1)} = 0$. Combining with this and Lemma 5.16, the equation above yields

$$2(\mathrm{I}_1+\mathrm{I}_2+\mathrm{I}_3) = (4m+2-4)\cdot\frac{1}{2}\deg(S) + t_{-1,1} + t_{2m+1,-(2m+1)},$$

which is clearly equivalent to the desired formula. □

Now we are ready to prove Lemma 5.14.

Proof of Lemma 5.14. By equation (5.1), using Lemmas 5.18, 5.19, and 5.20, we have

$$\begin{aligned}
\sum_{T\in\mathfrak{T}} c_T t_T &= \left(1-\frac{2}{n}\right)(\mathrm{I}_1+\mathrm{I}_2+\mathrm{I}_3) + \frac{m}{n}\deg(S) - \frac{1}{n}(\mathrm{III}_1+\mathrm{III}_2) - \left(1-\frac{1}{n}\right)t_{2m,-2m} \\
&\quad + \frac{2m+1}{2n}\deg(S) + \left(1-\frac{1}{n}\right)t_{-(2m+1),2m+1} - \left(1-\frac{1}{n}\right)t_{-1,1}. \\
&= \frac{1}{2}\left(1-\frac{2}{n}\right)\left[(2m-1)\deg(S) + t_{2m+1,-(2m+1)} + t_{-1,1}\right] \\
&\quad + \frac{4m+1}{2n}\deg(S) - \frac{1}{n}(\mathrm{III}_1+\mathrm{III}_2) - \left(1-\frac{1}{n}\right)t_{2m,-2m} \\
&\quad + \left(1-\frac{1}{n}\right)t_{-(2m+1),2m+1} - \left(1-\frac{1}{n}\right)t_{-1,1}.
\end{aligned}$$

Note that by Lemma 5.16 we have

$$\mathrm{III}_1+\mathrm{III}_2 = \sum_{i=1}^{2m} t_{i,-(2m+1)} + \sum_{i=1}^{2m+1} t_{-i,-(2m+1)} = \frac{1}{2}\deg(S) - t_{2m+1,-(2m+1)}.$$

Substituting it in the equation above and simplifying, we have

$$\begin{aligned}\sum_{T\in\mathfrak{T}} c_T t_T = &\left(\frac{2m-1}{2}+\frac{1}{n}\right)\deg(S)\\ &+\frac{1}{2}t_{2m+1,-(2m+1)}-\frac{1}{2}t_{-1,1}+\left(1-\frac{1}{n}\right)(t_{-(2m+1),2m+1}-t_{2m,-2m}).\end{aligned}$$

Since $t_{2m+1,-(2m+1)} = t_{-1,1}$ and $t_{-(2m+1),2m+1} = t_{2m,-2m}$ by Lemma 5.15, using $|w| = 2m+1$ we obtain

$$\sum_{T\in\mathfrak{T}} c_T t_T = \left(\frac{|w|}{2}-1+\frac{1}{n}\right)\deg(S).$$

□

5.4. **Proof of Theorem 5.3.** Now we are in a place to prove Theorem 5.3.

Proof of Theorem 5.3. By Lemmas 5.5 and 5.14, we may apply the LP duality method (Proposition 4.2) with $\lambda = |w|/2 - 1 + 1/n$ using the cost function c we defined above. Then Proposition 4.2 shows that

$$-\chi(S) \geq \left(1-\frac{1}{n}\right)\deg(S)$$

for any boundary-incompressible w-admissible surface S as desired. □

5.5. **The n-RF condition.** We give a brief discussion on the n-RF condition (Definition 5.1) in this subsection.

The following observations easily follow from the definition:

Lemma 5.21.

(1) If $a \in A \setminus C$ is n-RF rel C, then it is m-RF rel C for any $m \leq n$.
(2) For a chain of subgroup $C \leq B \leq A$, if $a \in A \setminus B$ is n-RF rel B, then it is also n-RF rel C.
(3) For a chain of subgroup $C \leq B \leq A$, if (A, B) and (B, C) are both n-RF for some $2 \leq n \leq \infty$, then (A, C) is also n-RF.

Proof. Items (1) and (2) follow directly from the definition. For (3), for any $a \in A \setminus C$, if $a \in B$, then it is n-RF rel C since (B, C) is n-RF. If $a \notin B$, then it is n-RF rel B since (A, B) is n-RF, and hence it is n-RF rel C by item (2). □

The next lemma reveals that the 2-RF condition is related to malnormality of C. Recall that a subgroup $C \leq A$ is called malnormal if $aCa^{-1} \cap C = id$ for all $a \notin C$.

Lemma 5.22. *An element $a \in A \setminus C$ is 2-RF rel C if and only if $aCa^{-1} \cap C = id$. Hence if a is n-RF rel C for all $a \in A \setminus C$ for some $n \geq 2$, then C is malnormal.*

Proof. The definition of 2-RF only requires that $ac_1a^{-1}c_2 \neq id$ when $c_1, c_2 \neq id$ (and an equivalent equation with a and a^{-1} swapped). That is, if $ac_1a^{-1} = c_2^{-1}$, then either c_1 or c_2 is the identity, in which case we have $c_1 = c_2 = id$ as they are conjugate. Hence this is equivalent to that $aCa^{-1} \cap C = id$. Then the second assertion easily follows from Lemma 5.21. □

It is also easy to observe that being n-RF rel C is really a condition on the double coset CaC.

Lemma 5.23. *If $a \in A \setminus C$ is n-RF (resp. n-RTF) rel C, then so is any $\tilde{a} \in CaC$.*

Proof. Let $\tilde{a} = cac'$ and suppose $\tilde{a}^{e_1}\tilde{c}_1 \cdots \tilde{a}^{e_k}\tilde{c}_k = id$ for some $k \geq 1$, $e_i = \pm 1$ and $\tilde{c}_i \in C$, where $\tilde{c}_i \neq id$ if $e_i = -e_{i+1}$. The goal is to show that there are at least n e_i's of the same sign. Replacing each $\tilde{a}$ (resp. $\tilde{a}^{-1}$) by cac' (resp. $c'^{-1}a^{-1}c^{-1}$), the equation can be rewritten as $a^{e_1}c_1 \cdots a^{e_k}c_k = id$, where

$$c_i = \begin{cases} c'\tilde{c}_i c & \text{if } e_i = e_{i+1} = 1; \\ c'\tilde{c}_i c'^{-1} & \text{if } e_i = 1, e_{i+1} = -1; \\ c^{-1}\tilde{c}_i c & \text{if } e_i = -1, e_{i+1} = 1; \\ c^{-1}\tilde{c}_i c' & \text{if } e_i = e_{i+1} = -1. \end{cases}$$

Hence when $e_i = -e_{i+1}$, we have $c_i \neq id$ as it is conjugate to $\tilde{c}_i$. Since a is n-RF, there must be at least n e_i's of the same sign as desired. The same proof works for the n-RTF condition. □

Here is the main proposition of this section, which we need in Section 6. Such inheritance should hold more generally for graphs of groups, but we just focus on the case of an amalgam. Similar inheritance for graphs of groups holds for the n-RTF condition; See [CH19, Lemmas 5.17 and 5.18].

Proposition 5.24. *Consider an amalgam $G = A \star_C B$. If for some $2 \leq n \leq \infty$ both (A, C) and (B, C) are n-RF, then (G, A) and (G, B) are n-RF.*

We will prove this proposition using the following more specific statement.

Lemma 5.25. *Let $g = \mathtt{x}_{-r}\mathtt{x}_{-r+1} \cdots \mathtt{x}_0 \cdots \mathtt{x}_{r-1}\mathtt{x}_r$ be a reduced word in $G = A \star_C B$ with $r \geq 1$, where $\mathtt{x}_i \in A \setminus C$ for all $i \equiv r \mod 2$ and $\mathtt{x}_i \in B \setminus C$ otherwise. Suppose $\mathtt{x}_{-r}$ and $\mathtt{x}_r$ are both 2-RF rel C, and $\mathtt{x}_0$ is n-RTF rel C for some $2 \leq n \leq \infty$. Then $g \in G \setminus B$ is n-RF rel B.*

We need one simple observation in the proof.

Lemma 5.26. *Suppose $n \geq 3$ in the setting of Lemma 5.25. Then for any $c \in C$,*

1. *either there is some $1 \leq k \leq r$ such that the element $h = (\mathtt{x}_0 \cdots \mathtt{x}_r)c(\mathtt{x}_{-r} \cdots \mathtt{x}_0) \in G$ is represented by a reduced word $\mathtt{x}_0 \cdots (\mathtt{x}_k c_k \mathtt{x}_{-k}) \cdots \mathtt{x}_0$ for some $c_k \in C$,*
2. *or $h = \mathtt{x}_0 c_0 \mathtt{x}_0 \notin C$ for some $c_0 \in C$.*

Proof. Note that $\mathtt{x}_i$ and $\mathtt{x}_{-i}$ lie in the same factor group for all i in our setup, and $\mathtt{x}_r, \mathtt{x}_{-r} \in A \setminus C$. Let $c_r = c$. If $\mathtt{x}_r c_r \mathtt{x}_{-r} \in A \setminus C$, then the result holds with $k = r$. If not, then $c_{r-1} := \mathtt{x}_r c_r \mathtt{x}_{-r} \in C$. In this case, we have $h = \mathtt{x}_0 \cdots \mathtt{x}_{r-1} c_{r-1} \mathtt{x}_{-r+1} \cdots \mathtt{x}_0$ and we examine whether the element $\mathtt{x}_{r-1} c_{r-1} \mathtt{x}_{-r+1} \in B$ lies in C. Continuing this process inductively,

(1) either we stop at some $1 \le k \le r$ by having $\mathtt{x}_k c_k \mathtt{x}_{-k} \notin C$, which gives the desired result;
(2) or we have reduce h all the way to $h = \mathtt{x}_0 c_0 \mathtt{x}_0$ for some $c_0 \in C$.

In the second situation, we must have $\mathtt{x}_0 c_0 \mathtt{x}_0 \notin C$ as desired since otherwise $\mathtt{x}_0 c_0 \mathtt{x}_0 c' = id$ for some $c' \in C$, contradicting the assumption that $\mathtt{x}_0$ is n-RTF for some $n \ge 3$. □

Now we prove Lemma 5.25.

Proof of Lemma 5.25. Note that $\mathtt{x}_r b \mathtt{x}_r^{-1}$ is a reduced word in G if $b \notin C$ and it lies in $A \setminus C$ if $b \in C \setminus \{id\}$ since $\mathtt{x}_r$ is 2-RF by assumption. Similarly $\mathtt{x}_{-r}^{-1} b \mathtt{x}_{-r}$ is either a reduced word already or lies in $A \setminus C$ for any $b \in B \setminus \{id\}$.

If $n = 2$, it suffices to show that $gbg^{-1}b' \neq id$ for all $b, b' \neq id \in B$. By moving $\mathtt{x}_{-r}$ the end, $gbg^{-1}b'$ is conjugate to

$$\mathtt{x}_{-r+1} \cdots \mathtt{x}_{r-1} (\mathtt{x}_r b \mathtt{x}_r^{-1}) \mathtt{x}_{r-1}^{-1} \cdots \mathtt{x}_{-r+1}^{-1} (\mathtt{x}_{-r}^{-1} b' \mathtt{x}_{-r}),$$

which is cyclically reduced by the observation above and thus nontrivial.

If $n \ge 3$, consider $w = g^{e_1} b_1 \cdots g^{e_k} b_k \in G$ for some $k \ge 1$, $e_i = \pm 1$, and $b_i \in B$ with the property that $b_i \neq id$ when $e_i = -e_{i+1}$ (indices taken mod k). The goal is to show that $w \neq id$ assuming there are no n e_i's of the same sign. The cyclic sequence of e_i's can be cut into (cyclic) subsequences so that all e_i's in each subsequence are equal and each subsequence has maximal length with this property. Corresponding to a subsequence where all $e_i = 1$, we have a subword of the form $u = g b_{i+1} \cdots g b_{i+\ell}$ for some $1 \le \ell < n$. The word is already reduced near each b_{i+j} if $b_{i+j} \notin C$. When $b_{i+j} \in C$, the word reduces as in Lemma 5.26. Summarizing all cases, the subword u reduces to

$$u = \mathtt{x}_{-r} \cdots \mathtt{x}_{-2} \mathtt{x}_{-1} \left(\prod_{j=1}^{\ell-1} w_j \right) \mathtt{x}_0 \cdots \mathtt{x}_r c_{i+\ell}$$

for some $0 \le k_j \le r$, where

$$w_j = \mathtt{x}_0 \cdots \mathtt{x}_{k_j - 1} (\mathtt{x}_{k_j} d_j \mathtt{x}_{-k_j}) \mathtt{x}_{-k_j+1} \cdots \mathtt{x}_{-1}, \text{ if } k_j \ge 1, \quad \text{and} \quad w_j = \mathtt{x}_0 d_j, \text{ if } k_j = 0,$$

for all $1 \le j < \ell$, and

(1) either $d_j \in B \setminus C$ and $k_j = r$,
(2) or $d_j \in C$ and $\mathtt{x}_{k_j} d_j \mathtt{x}_{-k_j} \notin C$.

Note that this is clearly a reduced word in G, except that we may have a subword of the form $\mathtt{x}_0 d_{j+1} \cdots \mathtt{x}_0 d_{j+m}$ for some $m \ge 1$ when some k_j's are zero. The subword $\mathtt{x}_0 d_{j+1} \cdots \mathtt{x}_0 d_{j+m}$ in the exceptional case lies in $A \setminus C$ if $r \equiv 0 \mod 2$ and in $B \setminus C$ if $r \equiv 1 \mod 2$ since $m \le \ell < n$ and $\mathtt{x}_0$ satisfies n-RTF rel C. Thus u is reduced in both situations. Moreover, the starting $\mathtt{x}_{-r}$ (resp. the ending $\mathtt{x}_r c_{i+\ell}$) ensures that the word representing u above has no cancellation with the tail of the preceding $g^{-1} c_{i-1}$ (resp. the head of the succeeding g^{-1})

by the observation above. By taking inverse, a similar reduction holds for a subsequence where all $e_i = -1$. It follows that w is a nontrivial reduced word, so $w \neq id$. □

Then we deduce Proposition 5.24 from Lemma 5.25.

Proof of Proposition 5.24. By symmetry, it suffices to show that (G, B) is n-RF. For any $g \in G \setminus B$, it can be written as a word $g = \mathtt{x}_{-r-1}\mathtt{x}_{-r} \cdots \mathtt{x}_0 \cdots \mathtt{x}_r\mathtt{x}_{r+1} \in G$ for some $r \geq 0$, where $\mathtt{x}_i \in A$ if $i \equiv r \mod 2$ and $\mathtt{x}_i \in B$ otherwise, and $\mathtt{x}_i \notin C$ for all i except that possibly $\mathtt{x}_{-r-1} = id$ or $\mathtt{x}_{r+1} = id$. We may assume $\mathtt{x}_{-r-1} = \mathtt{x}_{r+1} = id$ since g is n-RF rel B if and only if the same holds for any $g' \in BgB$ by Lemma 5.23. Now if $r \geq 1$, then g is in the form of Lemma 5.25 and our assumption implies that $\mathtt{x}_{-r}, \mathtt{x}_r, \mathtt{x}_0$ satisfy the requirements. Hence g is n-RF rel B.

If $r = 0$, then $g = \mathtt{x}_0 \in A \setminus C$, consider a word $w = g^{e_1}b_1 \cdots g^{e_k}b_k$ with $k \geq 1$, $b_i \in B$, and $b_i \neq id$ if $e_i = -e_{i+1}$. Those b_i's with the property $b_i \in B \setminus C$ (if exist) cut w into subwords, each of the form $u = \mathtt{x}_0^{f_1}c_1\mathtt{x}_0^{f_2} \cdots c_r\mathtt{x}_0^{f_{r+1}}$ for some $r \geq 0$ with $c_i \in C$, $f_i = \pm 1$ (which is equal to some $e_{i'}$) and $c_i \neq id$ if $f_i = -f_{i+1}$. Since $\mathtt{x}_0 \in A \setminus C$ is n-RF rel C, we see that $u \in A \setminus C$ unless we have n equal f_i's, which means we have n equal e_i's in w. Hence if there are no n equal e_i's, the subwords in between those b_i's with $b_i \in B \setminus C$ each lies in $A \setminus C$, so w must be a nontrivial element of G as desired. This completes the proof. □

6. Application to the Kervaire conjecture and related problems

Now we deduce from Theorem 5.3 results about a general word in an HNN extension $H = A\star_C$ with t-exponent sum ± 1. Throughout this section, let $p : H \to \mathbb{Z}$ be the epimorphism sending t to the generator $1 \in \mathbb{Z}$ and restricting to the trivial homomorphism on A.

Theorem 6.1. *Let $H = A\star_C$ be an HNN extension associated to injections $i_P, i_N : C \to A$ with standard presentation (3.1). Suppose for some $2 \leq n \leq \infty$, the group-subgroup pairs $(A, i_P(C))$ and $(A, i_N(C))$ are n-RF (Definition 5.1). Then for any $w \in H$ with $p(w) = \pm 1$ and not conjugate to $at^{\pm 1}$ for some $a \in A$, any boundary-incompressible w-admissible surface S has*

$$-\chi(S) \geq \left(1 - \frac{1}{n}\right) \deg(S).$$

The proof reduces the problem to the case of Theorem 5.3 using a somewhat standard algebraic trick, which at least goes back to Klyachko's original proof of the Kervaire–Laudenbach conjecture for torsion-free groups [Kly93, Lemma 3]. Such a statement for a free HNN extension and its proof can be found in [FR96, Lemma 4.2]. We use a similar argument to produce the desired Lemma 6.2 for a general HNN extension.

The trick is to express a conjugate of the given word w into the special form in Theorem 5.3 at the cost of passing to a different HNN extension structure of $H = A\star_C$. To see the different HNN extension structures, for each $k \in \mathbb{Z}_{\geq 0}$, let A_k be the subgroup generated by all words $t^{-i}at^i$ with $a \in A$ and $0 \leq i \leq k$. Note that $A_0 = A$ and A_k is the free product of $k+1$ copies of A amalgamated over k copies of C when $k \geq 1$. For convenience, let $A_{-1} = i_N(C) = A_0 \cap t^{-1}A_0 t$. Then $H = A\star_C$ is also the HNN extension of A_k over the subgroups A_{k-1} and $t^{-1}A_{k-1}t$ for any $k \geq 0$.

Lemma 6.2. *Let $H = A\star_C$ be an HNN extension. Let $w \in H$ be an element with $p(w) = 1$, where $p : H \to \mathbb{Z}$ is the epimorphism mentioned above. Then either w is conjugate to at for some $a \in A$, or there is some $k \in \mathbb{Z}_+$ such that a conjugate of w can be written as $a_1 t^{-1} b_1 t \cdots a_m t^{-1} b_m t x t$ for some $a_i \in A_{k-1} \setminus t^{-1}A_{k-2}t$, $b_i \in A_{k-1} \setminus A_{k-2}$, and $x \in A_{k-1}$ for some $m \geq 1$.*

Proof. Let $K = \ker p$, which is an amalgamated free product of infinitely many A's over C's, generated by elements of the form $t^{-i}at^i$ with $a \in A$ and $i \in \mathbb{Z}$. For any integers $k \leq \ell$, let $A_{k,\ell} \leq K$ be the subgroup generated by elements of the form $t^{-i}at^i$ with $a \in A$ and $k \leq i \leq \ell$. Comparing to the notation introduced above, we have $A_k = A_{0,k}$ for all $k \in \mathbb{Z}_{\geq 0}$. Note that $A_{k,\ell} \leq A_{k',\ell'}$ if $k' \leq k$ and $\ell \leq \ell'$, and $t^{-i}A_{k,\ell}t^i = A_{k+i,\ell+i}$.

Each $g \neq id \in K$ is contained in some $A_{k,\ell}$ where we take k to be maximal and ℓ to be minimal with this property, which we refer to as the support of g.

For any $h \in H$, we have $g = hwh^{-1}t^{-1} \in K$ supported in some $A_{\ell,\ell+k}$, where $\ell \in \mathbb{Z}$ and $k \in \mathbb{Z}_{\geq 0}$. Consider all conjugates hwh^{-1} of w such that the number k is minimal. Up to replacing h by ht^ℓ, we may assume that $g = hwh^{-1}t^{-1} \in A_k$.

If $k = 0$, then $hwh^{-1} = at$ for some $a \in A_0 = A$. So it suffices to consider the case $k \geq 1$. In such cases, A_k is the amalgamated free product of $U := A_{k-1} = A_{0,k-1}$ and $V := t^{-1}A_{k-1}t = A_{1,k}$ over W, where $W = A_{1,k-1}$ when $k \geq 2$ or $W = i_P(C)$ when $k = 1$. Hence $g = hwh^{-1}t^{-1}$ can be written as a reduced word in $U \star_W V$, which has length at least two as A_k is the support of g.

Among all conjugates hwh^{-1} of w with the property that $g = hwh^{-1}t^{-1} \in A_k$ for the minimal number k above, choose one such that the reduced word representing g is the shortest. There are two cases:

(1) If the reduced word representing g starts with some element in $U \setminus W$, i.e. $g = u_1 v_1 \cdots u_m v_m$, where $m \geq 1$, $u_i \in U \setminus W$ and $v_i \in V \setminus W$ for all i except that possibly $v_m = id$, in which case we must have $m \geq 2$.

 When $v_m \in V \setminus W$, simply let $a_i = u_i$ and $b_i = t v_i t^{-1}$ and $x = id$ for $1 \leq i \leq m$, which gives rise to the desired expression of $hwh^{-1} = gt$. In the case $v_m = 1$, define a_i, b_i in the same way for $i \leq m-1$ and let $x = u_m$.

(2) If the reduced word representing g starts with some element in $V \setminus W$, i.e. $g = v_0 u_1 \cdots v_m$, where $m \geq 1$, $u_i \in U \setminus W$ and $v_i \in V \setminus W$ for all i except that possibly $v_m = id$.

 If $v_m = id$, then $hwh^{-1} = gt$ is conjugate to $u_1 v_1 \cdots u_m (t v_0 t^{-1}) t$. Note that $t v_0 t^{-1} \in tVt^{-1} = A_{k-1} = U$, so $g' := u_1 v_1 \cdots v_{m-1}[u_m(t v_0 t^{-1})]$ is a (not necessarily reduced) word in A_k of length strictly less than that of the reduced word g, contradicting our choice of $g = hwh^{-1}t^{-1}$.

 If $v_m \in V \setminus W$, then $hwh^{-1} = gt$ is conjugate to $u_1 v_1 \cdots u_m v_m (t v_0 t^{-1}) t$. We have $u_{m+1} := t v_0 t^{-1} \in U$ as noted above, so $g' := u_1 v_1 \cdots u_m v_m u_{m+1}$ must be a reduced word in A_k since otherwise its word length in reduced form is strictly smaller than that of g. Hence we can proceed as in case (1).

□

To prove Theorem 6.1, we need to check that the reduced word in the new HNN extension structure as in Lemma 6.2 satisfies the conditions in Theorem 5.3. This easily follows from Proposition 5.24 by induction.

Lemma 6.3. *For an amalgam $G_{k+1} = H_1 \star_{C_1} H_2 \star_{C_2} \cdots H_k \star_{C_k} H_{k+1}$ with $k \geq 1$, suppose from some $2 \leq n \leq$ the image of each C_i in H_i (resp. H_{i+1}) is n-RF. Then the subgroups $G_k = H_1 \star_{C_1} H_2 \star_{C_2} \cdots H_k$ and H_{k+1} are both n-RF in G_{k+1}.*

Proof. We proceed by induction on k. The base case $k = 1$ is for an amalgam of $k + 1 = 2$ factor groups, and we know both factor groups are n-RF under our assumption by Proposition 5.24. Suppose the result holds for amalgams with no more than k factor groups and we show G_k and H_{k+1} are n-RF in G_{k+1}. Note that G_{k+1} is the amalgam of G_k with H_{k+1} over C_k, so the result follows from Proposition 5.24 once we show C_k is n-RF in G_k. By the induction hypothesis H_k is n-RF in G_k, and we know by assumption C_k is n-RF in H_k. Applying Lemma 5.21 (3) to the subgroup chain $C_k \leq H_k \leq G_k$, we see that C_k is n-RF in G_k as desired, which completes the proof. □

Corollary 6.4. *In the notation above for an HNN extension $G = A\star_C$, if for some $2 \leq n \leq \infty$ both $(A, i_P(C))$ and $(A, i_N(C))$ are n-RF, then the pairs $(A_{k-1}, t^{-1}A_{k-2}t)$ and (A_{k-1}, A_{k-2}) are also n-RF for all $k \geq 1$.*

Proof. There is nothing to prove for the case $k = 1$ as the assertion agrees with the assumption. For $k \geq 2$, as we mentioned earlier, A_{k-1} is the amalgam of $k + 1$ copies of A over k copies of C, where $t^{-1}A_{k-2}t$ and A_{k-2} are identified with the subgroups given by the amalgam of the first and last k copies of A. Hence the assertion follows from Lemma 6.3. □

Now we prove Theorem 6.1.

Proof of Theorem 6.1. Up to replacing w by w^{-1} we may assume $p(w) = 1$. By our assumption, w is not conjugate to at for any $a \in A$, thus by Lemma 6.2 w is conjugate to a reduced word $w' = a_1t^{-1}b_1t \cdots a_mt^{-1}b_mtxt$ in the HNN extension of A_{k-1} over the subgroups A_{k-2} and $t^{-1}A_{k-2}t$ for some $k \in \mathbb{Z}_+$ and $m \geq 1$. It is guaranteed by Lemma 6.2 that each $a_i \in A_{k-1} \setminus t^{-1}A_{k-2}t$, so it is n-RF and thus also n-RTF rel $t^{-1}A_{k-2}t$ by Corollary 6.4. Similarly each b_i is n-RF and n-RTF rel A_{k-2}. Note that any w-admissible surface S for the HNN extension $H = A\star_C$ is also w-admissible for the HNN extension structure of H above with vertex group A_{k-1} by enlarging the proper subgroup A to A_{k-1} in Definition 2.1. Moreover, the notion of boundary incompressiblity stays the same in the process as well as the quantities $\deg(S)$ and $-\chi(S)$. Thus the result follows directly from Theorem 5.3. □

Remark 6.5. It is essential to exclude the case where w is conjugate to at in Theorem 6.1. Note that $p(w)$ can be thought of as the algebraic intersection number of the loop γ

representing w with the edge space, which is assumed to be 1 (up to a sign), and in the case $w = at$ the geometric intersection number is also equal to 1, which is known to be a bad situation; see [HS09, Theorem 1.2 and Exmaple in Section 2] for instance. Such cases need to be considered separately.

For a free HNN extension $H = A \star \mathbb{Z}$, the assumptions in Theorem 6.1 above are easy to check, so we immediately obtain the following corollary:

Corollary 6.6. *Let A be an arbitrary group and let $p : A \star \mathbb{Z} \to \mathbb{Z}$ be the retract to $\mathbb{Z}$ induced by the trivial homomorphism $A \to \mathbb{Z}$ and $id_{\mathbb{Z}}$. If $w \in A \star \mathbb{Z}$ has $p(w) = \pm 1$, then any boundary-incompressible w-admissible surface S has*

$$-\chi(S) \geq \frac{1}{2} \deg(S).$$

Moreover, if each nontrivial element in A has order at least n for some $2 \leq n \leq \infty$, then we have a strengthened inequality

$$-\chi(S) \geq \left(1 - \frac{1}{n}\right) \deg(S).$$

Proof. Think of $H = A \star \mathbb{Z}$ as a free HNN extension. Then clearly by definition any $a \neq id \in A$ is n-RF rel the trivial subgroup id if a has order at least n. So the group-subgroup pair (A, id) is $2 - RF$ in all cases and n-RF if each nontrivial element in A has order at least n. Now if w is not conjugate to $at^{\pm 1}$, then Theorem 6.1 implies the desired bound. If $w = at^{\pm 1}$, the only two turn types (t, id, t^{-1}) and (t^{-1}, id, t), neither of which is an admissible turn (Section 3.4). Hence there is no boundary-incompressible w-admissible surface and thus the assertion is vacuous. $\square$

Now we turn to applications related to the Kervaire–Laudenbach conjecture. We deduce Klyachko's theorem as a consequence of Corollary 6.6.

Theorem 6.7 (Klyachko [Kly93])**.** *Let A be a torsion-free group and let $p : A \star \mathbb{Z} \to \mathbb{Z}$ be the retract to $\mathbb{Z}$ induced by the trivial homomorphism $A \to \mathbb{Z}$ and $id_{\mathbb{Z}}$. If $w \in A \star \mathbb{Z}$ has $p(w) = \pm 1$, then the natural map $A \to (A \star \mathbb{Z})/\langle\langle w \rangle\rangle$ is injective.*

Proof. Suppose the natural map is not injective, that is, there is some $a \neq id \in A$ that lies in $\langle\langle w \rangle\rangle$. As seen in Example 2.5 (with $H = A \star \mathbb{Z}$), this gives rise to an equation (2.1), which provides a w-admissible surface S of degree $\deg(S) = k$ with $-\chi(S) = k - 1$ (as it is a sphere with $k + 1$ disks removed) for some $k \in \mathbb{Z}_+$. Moreover, as explained in Example 2.5, when k is minimal among all equations of this form, S is boundary-incompressible. Hence by Corollary 6.6, we have $k - 1 = -\chi(S) \geq \deg(S) = k$ by taking $n = \infty$ as A is torsion-free, which leads to a contradiction. Thus the natural map must be injective. $\square$

The following result seems to be new, where the relator is a proper power but the factor group A is arbitrary.

Theorem 6.8. *Let A be an arbitrary group and let $p : A \star \mathbb{Z} \to \mathbb{Z}$ be the retract to $\mathbb{Z}$ induced by the trivial homomorphism $A \to \mathbb{Z}$ and $id_{\mathbb{Z}}$. If $w \in A \star \mathbb{Z}$ has $p(w) = \pm 1$, then the natural map $A \to (A \star \mathbb{Z})/\langle\langle w^m \rangle\rangle$ is injective for any $m \geq 2$.*

Proof. The proof is similar to the one above. If the map is not injective, we obtain a surface S that is a sphere with $k+1$ disks removed, where k boundary components each represents $w^{\pm m}$ and the remaining one represents some $a \neq id \in A$ for some $k \geq 1$. We consider S as a w-admissible surface, then it has $\deg(S) = km$. Moreover, S is boundary-incompressible when we take an equation with k minimal. Hence by Corollary 6.6, we have

$$k - 1 = -\chi(S) \geq \frac{1}{2}\deg(S) = \frac{mk}{2} \geq k$$

as $m \geq 2$, and we get a contradiction. Thus the map must be injective. $\square$

This proves Conjecture 1.5 when the free product has a $\mathbb{Z}$ factor other than A_λ and w projects to a generator of this $\mathbb{Z}$ factor. To the best knowledge of the author, the strongest previous result regarding this conjecture is a theorem of Howie [How90, Theorem A], which proves the case when $m \geq 4$. Note that for quotients by high powers, the problem is easier when the power m gets larger. For instance, the case $m \geq 7$ follows from small-cancellation theory [LS77, Corollary 9.4], and the case $m \geq 6$ was proved earlier by Gonzales-Acuña and Short [GAnS86b].

7. Questions

References

[Bav91] Christophe Bavard. Longueur stable des commutateurs. *Enseign. Math. (2)*, 37(1-2):109–150, 1991.

[Bro84] S. D. Brodskiĭ. Equations over groups, and groups with one defining relation. *Sibirsk. Mat. Zh.*, 25(2):84–103, 1984.

[Cal09a] Danny Calegari. *scl*, volume 20 of *MSJ Memoirs*. Mathematical Society of Japan, Tokyo, 2009.

[Cal09b] Danny Calegari. Stable commutator length is rational in free groups. *J. Amer. Math. Soc.*, 22(4):941–961, 2009.

[CH19] Lvzhou Chen and Nicolaus Heuer. Spectral gap of scl in graphs of groups and 3-manifolds, 2019.

[Che20] Lvzhou Chen. Scl in graphs of groups. *Invent. Math.*, 221(2):329–396, 2020.

[FR96] Roger Fenn and Colin Rourke. Klyachko's methods and the solution of equations over torsion-free groups. *Enseign. Math. (2)*, 42(1-2):49–74, 1996.

[GAnS86a] Francisco González-Acuña and Hamish Short. Knot surgery and primeness. *Math. Proc. Cambridge Philos. Soc.*, 99(1):89–102, 1986.

[GAnS86b] Francisco González-Acuña and Hamish Short. Knot surgery and primeness. *Math. Proc. Cambridge Philos. Soc.*, 99(1):89–102, 1986.

[Gor83] C. McA. Gordon. Dehn surgery and satellite knots. *Trans. Amer. Math. Soc.*, 275(2):687–708, 1983.

[GR62] Murray Gerstenhaber and Oscar S. Rothaus. The solution of sets of equations in groups. *Proc. Nat. Acad. Sci. U.S.A.*, 48:1531–1533, 1962.

[How82] James Howie. On locally indicable groups. *Math. Z.*, 180(4):445–461, 1982.

[How90] James Howie. The quotient of a free product of groups by a single high-powered relator. II. Fourth powers. *Proc. London Math. Soc. (3)*, 61(1):33–62, 1990.

[How02] James Howie. A proof of the Scott-Wiegold conjecture on free products of cyclic groups. *J. Pure Appl. Algebra*, 173(2):167–176, 2002.

[HS09] James Howie and Muhammad Sarwar Saeed. Freiheitssätze for one-relator quotients of surface groups and of limit groups. *Q. J. Math.*, 60(3):313–325, 2009.

[Ker65] Michel A. Kervaire. On higher dimensional knots. In *Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse)*, pages 105–119. Princeton Univ. Press, Princeton, N.J., 1965.

[Kir78] Rob Kirby. Problems in low dimensional manifold theory. In *Algebraic and geometric topology (Proc. Sympos. Pure Math., Stanford Univ., Stanford, Calif., 1976), Part 2*, Proc. Sympos. Pure Math., XXXII, pages 273–312. Amer. Math. Soc., Providence, R.I., 1978.

[Kly93] Anton A. Klyachko. A funny property of sphere and equations over groups. *Comm. Algebra*, 21(7):2555–2575, 1993.

[LS77] Roger C. Lyndon and Paul E. Schupp. *Combinatorial group theory.* Springer-Verlag, Berlin-New York, 1977. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 89.

[Rom12] Vitaliĭ Roman'kov. Equations over groups. *Groups Complex. Cryptol.*, 4(2):191–239, 2012.

Department of Mathematics, Purdue University, West Lafayette, Indiana, USA
Email address, L. Chen: `lvzhou@purdue.edu`