ISSN 1464-8997

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 261–293

Minimal Seifert manifolds for higher ribbon knots

James Howie

Abstract We show that a group presented by a labelled oriented tree presentation in which the tree has diameter at most three is an HNN extension of a finitely presented group. From results of Silver, it then follows that the corresponding higher dimensional ribbon knots admit minimal Seifert manifolds.

AMS Classification 57Q45; 20E06, 20F05, 57M05

Keywords Ribbon knots, Seifert manifolds, LOT groups

1 Introduction

It is well known that every classical knot k (knotted circle in S^3) bounds a compact orientable surface, known as a *Seifert surface* for the knot. A Seifert surface Σ of minimal genus (among all Seifert surfaces for the given knot k) is called *minimal*, and satisfies the following property: the inclusion-induced map $\pi_1(\Sigma\backslash k) \to \pi_1(S^3\backslash k)$ is injective.

For a higher dimensional knot, or more generally a knotted (closed, orientable) n–manifold M in S^{n+2}, a *Seifert manifold* is defined to be a compact, orientable $(n+1)$–manifold W in S^{n+2}, such that $\partial W = M$. A Seifert manifold W for M is defined to be *minimal* if the inclusion-induced map $\pi_1(W\backslash M) \to \pi_1(S^{n+2}\backslash M)$ is injective. In general, any M will always admit Seifert manifolds, but not necessarily minimal Seifert manifolds. For example, Silver [13] has shown that, for any $n \geq 3$, there exist n–knots in S^{n+2} with no minimal Seifert manifolds, and Maeda [9] has constructed, for all $g \geq 1$, a knotted surface of genus g in S^4 that has no minimal Seifert manifold. Further examples of knotted tori in S^4 without minimal Seifert manifolds are constructed by Silver [16].

A theorem of Silver [14] says that, for $n \geq 3$, a knotted n–sphere K in S^{n+2} has a minimal Seifert manifold if and only if its group $G_K = \pi_1(S^{n+2}\backslash K)$ can be expressed as an HNN extension with a *finitely presented* base group. (It is standard that any higher knot group can be expressed as an HNN extension with a *finitely generated* base group.)

As Silver remarks, the proof of his theorem does not extend to the case $n = 2$. However, it remains a *necessary* condition for the existence of a minimal Seifert manifold that the group be an HNN extension with finitely presented base group. This applies also to knotted n–manifolds in S^{n+2}, a fact which is used implicitly by Maeda in the result mentioned above. It remains an open question whether every 2–knot in S^4 has a minimal Seifert manifold. This seems unlikely, however. For example Hillman [5], p. 139 shows that, provided the 3–dimensional Poincaré Conjecture holds, there is an infinite family of distinct 2–knots, all with the same group G, such that the commutator subgroup of G is finite of order 3; and at most one of these knots can admit a minimal Seifert manifold.

In the present article we consider the case of higher dimensional *ribbon knots*, for which the existence of minimal Seifert manifolds is also an open question. Indeed, as we shall point out in the next section, higher ribbon knot groups are special cases of *knot-like groups*, in the sense of Rapaport [12], and Silver [15] has conjectured that every finitely generated HNN base for a knot-like group is finitely presented. It would therefore follow from Silver's conjecture (and his Theorem) that every higher ribbon knot has a minimal Seifert manifold.

Now any higher ribbon knot group has a Wirtinger-like presentation that can be encoded in the form of a *labelled oriented tree* (LOT) [7]. Indeed the LOT encodes not only a presentation for the knot group, but the complete homotopy type of the knot complement. In [7] it was shown that, if the diameter of the tree is at most 3, then the group is locally indicable, and using this that the 2–complex model of the associated Wirtinger presentation is aspherical. A shorter proof of this fact is given in [8], where it is shown that the presentation is in fact diagrammatically aspherical.

In the present paper, we show that, under the same hypothesis on the diameter of the tree, the group is an HNN extension with finitely presented base group, and hence that the higher ribbon knot has a minimal Seifert manifold.

Theorem 1.1 *Let Γ be a labelled oriented tree of diameter at most 3, and $G = G(\Gamma)$ the corresponding group. Then G is an HNN extension with finitely presented base group.*

Corollary 1.2 *Let K be a ribbon n–knot in S^{n+2}, where $n \geq 3$, such that the associated labelled oriented tree has diameter at most 3. Then K admits a minimal Seifert manifold.*

The paper is arranged as follows. In section 2 we recall some basic definitions relating to LOTs and higher ribbon knots. In section 3 we prove some preliminary results about HNN bases for one-relator products of groups, which will allow us to simplify the original problem. In section 4 we reduce the problem to the study of *minimal* LOTs, In section 5 we construct a finitely generated HNN base B for G, and describe a finite set of relators in these generators. In section 6 we prove some technical results about the structure of these relations, which we apply in section 7 to complete the proof of Theorem 1.1 by proving that this finite set is a set of defining relators for B. We close, in section 8, with a geometric description of our generators and relators for the HNN base, and a discussion of how this might be used to generalise Theorem 1.1.

Acknowledgements

In the course of this work I have received useful comments and advice from Nick Gilbert and from Dan Silver. I am grateful to them for their help.

2 LOTs and higher ribbon knots

A *labelled oriented tree* (LOT) is a tree Γ, with vertex set $V = V(\Gamma)$, edge set $E = E(\Gamma)$, and initial and terminal vertex maps $\iota, \tau\colon E \to V$, together with an additional map $\lambda\colon E \to V$. For any edge e of Γ, $\lambda(e)$ is called the *label* of e. In general, one can consider LOTs of any cardinality, but for the purposes of the present paper, every LOT will be assumed to be finite.

To any LOT Γ we associate a presentation

$$\mathcal{P} = \mathcal{P}(\Gamma): \quad \langle\, V(\Gamma) \mid \iota(e)\lambda(e) = \lambda(e)\tau(e) \,\rangle$$

of a group $G = G(\Gamma)$, and hence also a 2–complex $K = K(\Gamma)$ modelled on $\mathcal{P}$. The 2–complex K is a spine of a *ribbon disk complement* $D^4 \backslash k(D^2)$ [7], that is the complement of an embedded 2–disk in D^4, such that the radial function on D^4 composed with the embedding k is a Morse function on D^2 with no local maximum. Conversely, any ribbon disk complement has a 2–dimensional spine of the form $K(\Gamma)$ for some LOT Γ.

By doubling a ribbon disk, we obtain a ribbon 2–knot in S^4, and by successively spinning we can obtain ribbon n–knots in S^{n+2} for all $n \geq 2$. In each case the group of the knot is isomorphic to the fundamental group of the ribbon disk complement that we started with. Conversely, every ribbon n–knot (for $n \geq 2$) can be constructed this way, so that higher ribbon knot groups and LOT groups are precisely the same thing.

Recall [12] that a group G is *knot-like* if it has a finite presentation with deficiency 1 (in other words, one more generator than defining relator), and infinite cyclic abelianisation. It is clear that every LOT group has these properties, so LOT groups are special cases of knot-like groups.

The *diameter* of a finite connected graph Γ is the maximum distance between two vertices of Γ, in the edge-path-length metric. A key factor in our situation is the special nature of trees of diameter 3 or less. For any LOT Γ of diameter 0 or 1, it is easy to see that $G(\Gamma)$ is infinite cyclic, so such LOTs are of little interest.

Remark Every tree of diameter 2 has a single non-extremal vertex. Every tree of diameter 3 has precisely 2 non-extremal vertices.

We recall from [7] that a LOT Γ is *reduced* if:

(i) for all $e \in E$, $\iota(e) \neq \lambda(e) \neq \tau(e)$;

(ii) for all $e_1 \neq e_2 \in E$, if $\lambda(e_1) = \lambda(e_2)$ then $\iota(e_1) \neq \iota(e_2)$ and $\tau(e_1) \neq \tau(e_2)$;

(iii) every vertex of degree 1 in Γ occurs as a label of some edge of Γ.

For every LOT Γ there is a reduced LOT Γ' with the same group as Γ, and the same or smaller diameter, so we may also restrict our attention to reduced LOTs.

A subgraph Γ' of a LOT Γ is *admissible* if $\lambda(e) \in V(\Gamma')$ for all $e \in E(\Gamma')$. If Γ' is connected and admissible, then it is also a LOT. A LOT is *minimal* if every connected admissible subgraph consists only of a single vertex.

If Γ is a LOT and $A \subseteq V(\Gamma)$, we define the *span* of A (in Γ) to be the smallest subgraph Γ' of G such that:

(i) $A \subseteq V(\Gamma')$; and

(ii) if $e \in E(\Gamma)$ with $\lambda(e) \in V(\Gamma')$ and at least one of $\iota(e)$, $\tau(e)$ belongs to $V(\Gamma')$, then $e \in E(\Gamma')$.

We write span(A) for the span of A, and say that A *spans*, or *generates* Γ' if $\Gamma' = \text{span}(A)$. The following is essentially Proposition 4.2 of [7].

Lemma 2.1 *If Γ is a LOT spanned by A, then $\mathcal{P}(\Gamma)$ is Andrews–Curtis equivalent to a presentation with generating set A. If Γ' is an admissible subgraph of Γ with $V(\Gamma') \subseteq A$, then the presentation may be chosen to contain $\mathcal{P}(\Gamma')$, and the Andrews–Curtis moves can be taken relative to $\mathcal{P}(\Gamma')$.*

Corollary 2.2 *If Γ is a LOT spanned by two vertices, then $G(\Gamma)$ is a torsion-free one-relator group.*

Proof Let A be a set of two vertices spanning Γ. Then $\mathcal{P}(\Gamma)$ is Andrews–Curtis equivalent to a presentation $\langle A|R\rangle$. Since $\mathcal{P}(\Gamma)$ has deficiency 1, the same is true of the equivalent presentation $\langle A|R\rangle$. In other words $|R| = 1$, and $G(\Gamma)$ is a one-relator group. But the abelianisation G^{ab} of G is infinite cyclic, so the relator $r \in R$ cannot be a proper power, and so G is torsion-free. □

We will require the following generalisation of Corollary 2.2. Recall that a *one-relator product* of two groups A, B is the quotient of the free product $A * B$ by the normal closure of a single word w, called the *relator*.

Corollary 2.3 *If Γ is a LOT spanned by $V(\Gamma') \cup \{x\}$, where Γ' is an admissible subgraph of Γ and x is a vertex in $V(\Gamma) \setminus V(\Gamma')$, then $G(\Gamma)$ is a one-relator product of $G(\Gamma')$ and $\mathbb{Z}$, where the relator is not a proper power.*

Proof Let $A = V(\Gamma') \cup \{x\}$ and apply the Theorem. Then $\mathcal{P}(\Gamma)$ is equivalent, relative to $\mathcal{P}(\Gamma')$, to a presentation $\mathcal{Q}$ with generating set A and containing $\mathcal{P}(\Gamma')$. Now each of $\mathcal{P}(\Gamma)$, $\mathcal{P}(\Gamma')$ and $\mathcal{Q}$ has deficiency 1. Moreover, $\mathcal{Q}$ has one more generator than $\mathcal{P}(\Gamma')$, so $\mathcal{Q}$ also has one more defining relator than $\mathcal{P}(\Gamma')$. It follows that $G(\Gamma)$ is a one relator product of $G(\Gamma')$ with the infinite cyclic group $\langle x\rangle$. Finally, since the abelianisations of $G(\Gamma)$, $G(\Gamma')$ and $\langle x\rangle$ are all infinite cyclic, it follows that the relator cannot be a proper power. □

3 One-relator groups and one-relator products

The following result is merely a summary of some well-known properties of one-relator groups, which have useful applications to our situation. Recall that a group G is *locally indicable* if, for every nontrivial, finitely generated subgroup H of G, there exists an epimorphism $H \to \mathbb{Z}$.

Theorem 3.1 *Let G be a finitely generated one-relator group. Then*

(i) *G is either a finite cyclic group, or an HNN extension of a finitely presented, one-relator group (with shorter defining relator);*

(ii) *if the defining relator of G is not a proper power, then G is locally indicable.*

Proof See [11] and [3] respectively. □

In order to complete the process of reducing ourselves to a simple special case, we require a generalisation of the above theorem to one-relator products. Suppose that A and B are locally indicable groups, and $N = N(w)$ is the normal closure in $A * B$ of a cyclically reduced word w of length at least 2 that is not a proper power. Then the one-relator product $G = (A * B)/N$ is known [6] to be locally indicable. We show also that G has a finitely presented HNN base, provided that A and B also have this property.

Theorem 3.2 *Let $G = (A * B)/N(w)$ be a one-relator product of two finitely presented, locally indicable groups A and B, each of which has a finitely presented HNN base. Suppose also that G^{ab} is infinite cyclic, with each of the natural maps $A^{ab} \to G^{ab}$ and $B^{ab} \to G^{ab}$ an isomorphism. Then G is a finitely presented, locally indicable group with a finitely presented HNN base.*

Remark The condition on G^{ab} in this theorem is unnecessary for the proof that G has a finitely presented HNN base. It can be removed at the expense of a less straightforward proof. However the condition does hold for all the groups that we are considering in this paper, so there is no loss of generality for us in imposing that condition. The condition also ensures that w cannot be a proper power, so that G is locally indicable by the results of [6].

Proof A presentation for G can be obtained by taking the disjoint union of finite presentations for A and for B, and imposing the single additional relation $w = 1$. Hence G is finitely presented. As pointed out in the remark above, w cannot be a proper power, so G is locally indicable by [6]. It remains only to prove that G has a finitely presented HNN base.

Let

$$A = \langle A_0, a | a^{-1}ga = \alpha(g)\ (g \in A_1)\rangle$$

and

$$B = \langle B_0, b | b^{-1}hb = \beta(h)\ (h \in B_1)\rangle$$

be HNN presentations for A and B with finitely presented bases A_0 and B_0 respectively. Since A and B are finitely presented, it follows also that the associated subgroups A_1 and B_1 are finitely generated.

The commutator subgroup G' of G can be expressed in the form

$$(A' * B' * \langle\ c_n\ (n \in \mathbb{N})\rangle)/N(\{w_n\ (n \in \mathbb{N})\}),$$

where $c_n = a^{n+1}b^{-1}a^{-n}$ and $w_n = a^{-n}wa^n$.

Now A' is an infinite stem product

$$\cdots\quad (a^{-1}A_0a) \underset{(a^{-1}A_1a)}{*} A_0 \underset{A_1}{*} (aA_0a^{-1}) \quad\cdots$$

Since A_0 is finitely presented and A_1 is finitely generated, the subgroup

$$(a^{-k}A_0a^k) \underset{(a^{-k}A_1a^k)}{*} \cdots\cdots \underset{(a^{k-1}A_1a^{1-k})}{*} (a^kA_0a^{-k})$$

is finitely presented for each k. Moreover it is also an HNN base for A. Replacing A_0 by this subgroup, for any sufficiently large k, we may assume that $w_0 \in A_0 * B' * \langle\ c_n\ (n \in \mathbb{N})\rangle$.

Similarly, possibly after replacing B_0 by a sufficiently large finitely presented HNN base for B, we may assume that $w_0 \in A_0 * B_0 * \langle\ c_n\ (n \in \mathbb{N})\rangle$. Now let μ and ν be the least and greatest indices i such that c_i occurs in w_0. (Note that at least one c_i occurs in w_0, for otherwise $w_0 \in A_0 * B_0$, so $w \in A' * B'$, whence $G^{ab} \cong A^{ab} \times B^{ab} \not\cong \mathbb{Z}$, a contradiction.) Define $G_0 = (A_0 * B_0 * \langle c_\mu, \ldots, c_\nu\rangle)/N(w_0)$ and $G_1 = A_0 * B_0 * \langle c_\mu, \ldots, c_{\nu-1}\rangle$, and observe that G_0 is a finitely presented HNN base for G, with associated subgroup G_1. □

4 Reduction of the problem

Recall from section 2 that a LOT Γ is *minimal* if it contains no admissible subtree with more than one vertex. In this section we reduce the proof of the main theorem to the case of a minimal LOT of diameter 3, using the results of section 3. The key point is that a non-minimal LOT can be obtained from a minimal admissible subtree by successively expanding to the span of the existing tree with one extra vertex. By Corollary 2.3, this construction corresponds at the group level to taking a one-relator product of a given group with an infinite cyclic group.

Lemma 4.1 *Let Γ be a LOT of diameter at most 3, containing a proper admissible subtree with more than one vertex. Then there is such an admissible subtree Γ' and a vertex $x \in V(\Gamma)\backslash V(\Gamma')$ such that Γ is spanned by $V(\Gamma') \cup \{x\}$.*

Proof Suppose first that some extremal vertex x of Γ does not occur as a label of any edge of Γ. In this case we take Γ' to consist of Γ with the vertex x and the edge incident to x removed. Clearly Γ' is connected, so a subtree of Γ. Since x is not the label of any edge in $E(\Gamma')$, it follows that Γ' is admissible. Moreover Γ is spanned by $V(\Gamma) = V(\Gamma') \cup \{x\}$, as required.

We may therefore assume that every extremal vertex of Γ occurs at least once as the label of an edge of Γ.

Next suppose that Γ has a proper admissible subtree that contains all the non-extremal vertices of Γ. Let Γ' be a maximal such admissible subtree. The vertices in $V(\Gamma)\backslash V(\Gamma')$ are all extremal in Γ, so occur as labels of edges of Γ. But since Γ' is admissible, no such vertex can be a label of an edge of Γ'. Since the finite sets $V(\Gamma)\backslash V(\Gamma')$ and $E(\Gamma)\backslash E(\Gamma')$ have the same cardinality, it follows that each vertex in $V(\Gamma)\backslash V(\Gamma')$ is the label of precisely one edge in $E(\Gamma)\backslash E(\Gamma')$. In turn, this edge has precisely one endpoint in $V(\Gamma)\backslash V(\Gamma')$, so we can define a permutation σ on $V(\Gamma)\backslash V(\Gamma')$ by defining $\sigma(x)$ to be the extremal endpoint of the unique edge labelled x, for all $x \in V(\Gamma)\backslash V(\Gamma')$. Now fix some vertex $x \in V(\Gamma)\backslash V(\Gamma')$, let t be the size of the orbit of σ that contains x, and define $x_i = \sigma^i(x)$, $i = 1, \dots, t$. Now $\Delta = \mathrm{span}(V(\Gamma') \cup \{x\})$ contains the vertex $x = x_t$, together with any non-extremal vertex of Γ. Hence Δ contains the edge labelled x_t, and hence its endpoint x_1. Similarly Δ contains $x_2, \dots, x_{t-1}$, as well as the edges labelled $x_1, \dots, x_{t-1}$. On the other hand, The vertices $x_1, \dots, x_t$, the edges labelled by them, and the vertices and edges of Γ' together form an admissible subtree of Γ, which by maximality of Γ' must be the whole of Γ. Hence $\Delta = \Gamma$, in other words Γ is spanned by $V(\Gamma') \cup \{x\}$.

Finally, suppose that no proper admissible subtree of Γ contains all the non-extremal vertices of Γ. In particular, Γ must have more than one non-extremal vertex, so has diameter 3. By hypothesis, there is a proper admissible subtree Γ' of Γ that contains more than one vertex. Hence Γ' contains precisely one of the two nonextremal vertices of Γ, say u. As an abstract graph, Γ is the union of Γ' with another tree Γ'', such that $\Gamma' \cap \Gamma'' = \{u\}$. Note that Γ'' contains both of the non-extremal vertices of Γ, so cannot be an admissible subtree, by hypothesis. Hence at least one edge f of Γ'' is labelled by a vertex a of Γ' (other than u). Let e be the edge of Γ that joins the two non-extremal vertices u, v, and let $\Delta = \mathrm{span}(V(\Gamma') \cup \{\lambda(e)\})$. Then Δ contains Γ' and the edge e, and hence v, and hence the edge f. Each extremal vertex of Δ is the label of an edge of Γ, and hence of Δ, since Δ contains at least one endpoint (namely u or v) of every edge of Γ. Moreover there are $|E(\Gamma')|+1$ edges of Δ labelled by the $|V(\Gamma')|=|E(\Gamma')|+1$ vertices of Γ', so an easy counting argument shows that there must be at least $|V(\Delta)|-1$ edges in Δ. In other words Δ is a tree, so the whole of Γ. In other words Γ is spanned by $V(\Gamma')\cup\{\lambda(e)\}$. □

Remark If Γ is a minimal LOT of diameter 2, then the above argument still applies (to the subtree consisting of only the unique non-extremal vertex). In this case we see that the permutation σ is transitive, and that Γ is spanned by two vertices.

Lemma 4.2 *Let Γ be a minimal LOT of diameter 3, and let u,v be the two non-extremal vertices of Γ. Then one of the following holds:*

(i) *One of u,v is a label in Γ, and Γ is spanned by $\{u,v\}$;*

(ii) *Some vertex a occurs twice as a label in Γ, and Γ is spanned by $\{a,u,v\}$.*

Proof By minimality of Γ, every extremal vertex of Γ occurs as a label. There are $|V|-2$ extremal vertices, and $|V|-1$ edges, so either one of u,v occurs as a label or some unique extremal vertex a occurs twice as a label. Note that every edge of Γ is incident to at least one of u,v, so if $u,v\in A\subset V$ then every edge labelled by a vertex of $\mathrm{span}(A)$ is an edge of $\mathrm{span}(A)$.

(i) Suppose that u occurs as a label, and let $\Gamma'=\mathrm{span}(\{u,v\})$. If Γ' has $k+2$ vertices $u,v,x_1,\dots,x_k$, then $x_1,\dots,x_k$ are all extremal in Γ, so each of $u,x_1,\dots,x_k$ is a label of an edge of Γ, which must therefore be an edge of Γ'. Hence Γ' has at least $k-1$ edges, so is connected. By minimality of Γ we have $\Gamma=\Gamma'=\mathrm{span}(\{u,v\})$.

(ii) Suppose that an extremal vertex a appears twice as a label, and let $\Gamma'=\mathrm{span}(\{a,u,v\})$. If Γ' has $k+3$ vertices $a,u,v,x_1,\dots,x_k$, then each of $x_1,\dots,x_k$ is extremal, so the label of an edge of Γ, while a is the label of 2 edges of Γ. Each of these $k+2$ edges is an edge of Γ', so Γ' is connected, and by minimality again we have $\Gamma=\Gamma'=\mathrm{span}(\{a,u,v\})$. □

Corollary 4.3 *If Γ is either a minimal LOT of diameter 2, or a minimal LOT of diameter 3 in which no vertex occurs twice as a label, then $G(\Gamma)$ is a locally indicable group with a finitely presented HNN base.*

Proof By Lemma 4.2 or the remark following Lemma 4.1, Γ is spanned by two vertices. Hence $G = G(\Gamma)$ is a 2–generator, one-relator group. Since G^{ab} is infinite cyclic, G is not finite, and the relator of G cannot be a proper power. The result follows immediately from Theorem 3.1. □

Using the above results, we can reduce our problem to the case of a minimal LOT of diameter 3 that is not spanned by two vertices. In particular, some extremal vertex must occur twice as a label.

Corollary 4.4 *If the group of every reduced, minimal LOT of diameter 3 which is not spanned by two vertices is locally indicable with finitely presented HNN base, then the same is true for every LOT of diameter 3 or less.*

Recall [7] that the *initial graph* $I(\Gamma)$ of Γ is the graph with the same vertex and edge sets as Γ, but with incidence maps ι, λ. Similarly the *terminal graph* $T(\Gamma)$ of Γ has the same vertex and edges sets as Γ, but incidence maps λ, τ. It was shown in [7] that the commutator subgroup of $G(\Gamma)$ is locally free if either $I(\Gamma)$ or $T(\Gamma)$ is connected. (If $I(\Gamma)$ and $T(\Gamma)$ are both connected, then $G(\Gamma)'$ is free of finite rank.) In particular, any finitely generated HNN base for $G(\Gamma)$ is free, so automatically finitely presented.

Hence we can concentrate attention on the case of a minimal LOT Γ of diameter 3, not spanned by any two of its vertices, such that neither $I(\Gamma)$ nor $T(\Gamma)$ is connected. Our next result gives a detailed description of the structure of $I(\Gamma)$. In particular it will show us that $I(\Gamma)$ has precisely two connected components, one containing each of the nonextremal vertices of Γ. A similar statement holds for $T(\Gamma)$.

Lemma 4.5 *Let Γ be a minimal LOT of diameter 3, with nonextremal vertices u and v, and an extremal vertex a that occurs twice as a label of edges of Γ. Then:*

- (i) *u and v are sources in $I(\Gamma)$;*
- (ii) *no vertex other than u or v is the initial vertex of more than one edge of $I(\Gamma)$;*
- (iii) *a is the terminal vertex of precisely two edges of $I(\Gamma)$;*
- (iv) *each vertex other than a, u, v is the terminal vertex of precisely one edge of $I(\Gamma)$;*
- (v) *any directed cycle in $I(\Gamma)$ contains a;*
- (vi) *each component of $I(\Gamma)$ contains at least one of u, v;*

(vii) $I(\Gamma)$ *has at most two connected components.*

Proof (i) Since $\lambda(e) \neq u$ for all $e \in E(\Gamma)$, u is not the terminal vertex of any edge in $I(\Gamma)$, in other words u is a source. Similarly v is a source in $I(\Gamma)$.

(ii) Any vertex x of Γ, with the exception of u and v, is extremal in Γ, so the initial vertex of at most one edge of Γ. Hence x is also the initial vertex of at most one edge in $I(\Gamma)$.

(iii) $a = \lambda(e)$ for precisely two edges $e \in E(\Gamma)$.

(iv) If $x \in V(\Gamma)\backslash\{a,u,v\}$ then $x = \lambda(e)$ for precisely one edge $e \in E(\Gamma)$.

(v) Suppose $(e_1, e_2, \ldots, e_n)$ is a directed cycle in $I(\Gamma)$. Then there are vertices $x_1, \ldots, x_n \in V(\Gamma)$ with $x_i = \iota(e_i)$ for all i, $\lambda(e_i) = x_{i+1}$ for $i < n$, and $\lambda(e_n) = x_1$. Now each x_i is extremal since it occurs as a label. If no x_i is equal to a then we can remove the vertices $x_1, \ldots, x_n$ and the edges $e_1, e_2, \ldots, e_n$ from Γ to form a connected, admissible subgraph Γ' that contains at least three vertices (a, u, v). This contradicts the minimality of Γ, and so $x_i = a$ for some i, as claimed.

(vi) By (iv) if $x \notin \{a, u, v\}$ then x is the terminal vertex in $I(\Gamma)$ of a unique edge. If the initial vertex of this edge is not one of a, u, v then it also is the terminal vertex of a unique edge. Continuing in this way, we can construct a directed path that ends at x, and either begins at one of a, u, v or contains a cycle. By (v) any directed cycle contains a, so in any case we have a directed path from one of a, u, v to x. It suffices therefore to find a path in $I(\Gamma)$ from u or v to a. But a is the terminal vertex in $I(\Gamma)$ of precisely two edges, with initial vertices x_1 and x_2 say. Now apply the above argument to each of x_1, x_2. If there is a path from u or v to x_1 or x_2 then we are done. Otherwise there are directed paths from a to each of x_1, x_2. Neither u nor v can belong to these paths, since they are sources in $I(\Gamma)$. But then from (ii) it follows that there is at most one directed path of any given length beginning at a, whence $x_1 = x_2$, a contradiction. Hence there is a directed path in $I(\Gamma)$ from u or v to a, as claimed.

(vii) This follows immediately from (vi). □

A similar result holds for $T(\Gamma)$.

Lemma 4.6 *Let Γ be a minimal LOT of diameter 3, with nonextremal vertices u and v, and an extremal vertex a that occurs twice as a label of edges of Γ. Then:*

(i) *u and v are sinks in $T(\Gamma)$;*

(ii) *no vertex other than u or v is the terminal vertex of more than one edge of $T(\Gamma)$;*

(iii) *a is the initial vertex of precisely two edges of $T(\Gamma)$;*

(iv) *each vertex other than a, u, v is the initial vertex of precisely one edge of $T(\Gamma)$;*

(v) *any directed cycle in $T(\Gamma)$ contains a;*

(vi) *each component of $T(\Gamma)$ contains at least one of u, v;*

(vii) *$T(\Gamma)$ has at most two connected components.*

Corollary 4.7 *Suppose that Γ is a reduced, minimal LOT of diameter 3, which is not spanned by two vertices, and such that neither $I(\Gamma)$ nor $T(\Gamma)$ is connected. Then*

(i) *There is a unique extremal vertex a of Γ that is the label of two distinct edges of Γ. One of these edges has an extremal initial vertex, and the other has an extremal terminal vertex.*

(ii) *$I(\Gamma)$ has precisely two connected components, each containing one of the two nonextremal vertices u, v of Γ.*

(iii) *There is a unique cycle in $I(\Gamma)$, which is either a directed cycle containing a, or consists of two directed paths (one of length 1, the other of length at least 2), from u or v to a.*

(iv) *$T(\Gamma)$ has precisely two connected components, each containing one of the two nonextremal vertices u, v of Γ.*

(v) *There is a unique cycle in $T(\Gamma)$, which is either a directed cycle containing a, or consists of two directed paths (one of length 1, the other of length at least 2), from a to u or v.*

(vi) *The cycles in $I(\Gamma)$ and $T(\Gamma)$ are not both directed.*

Proof (i) We already know that there is an extremal vertex a occurring twice as a label, by Lemma 4.2, since otherwise Γ can be spanned by two vertices. We also know that a is unique, since every extremal vertex occurs at least once as a label. Now suppose that neither of the edges labelled a has extremal initial vertex. The initial vertices of these two edges must be distinct, since Γ is reduced, and so must be the two nonextremal vertices u, v of Γ. But then there are edges of $I(\Gamma)$ from both u and v to a. Hence u and v belong to the same connected component of $I(\Gamma)$. By Lemma 4.5, (vi) it follows that $I(\Gamma)$ is connected, a contradiction.

A similar contradiction arises if neither edge has an extremal terminal vertex.

(ii) This is just a restatement of Lemma 4.5, (vi), together with the hypothesis that $I(\Gamma)$ is not connected.

(iii) Since $I(\Gamma)$ has the same vertex and edge sets as Γ, it has the same euler characteristic, namely 1. Since $I(\Gamma)$ has two components, it follows that $H_1(\Gamma) \cong \mathbb{Z}$, so there is a unique cycle in $I(\Gamma)$. If this cycle is directed, then it must contain a, by Lemma 4.5, (v). Otherwise it must contain at least two vertices at which the orientation of the edges of the cycle changes. This is possible only at a vertex which is either the initial vertex of at least two edges or the terminal vertex of at least two edges, and by Lemma 4.5 the only such vertices are a, u, v. Let us assume that a is in the same component of $I(\Gamma)$ as u. Then the cycle must contain both a and u, and indeed must consist of two directed paths from u to a. By uniqueness of the cycle (or directly from Lemma 4.5), we see that there only two directed paths in $I(\Gamma)$ from u to a. Moreover, precisely one of these paths is of length 1, since precisely one of the edges of Γ labelled a has a nonextremal initial vertex.

(iv) Similar to (ii).

(v) Similar to (iii).

(vi) If the cycle in $I(\Gamma)$ is directed, then there is an edge of $I(\Gamma)$ with initial vertex a, and so also there is an edge of Γ with initial vertex a. Similarly, if the cycle in $T(\Gamma)$ is directed, then there is an edge of Γ with terminal vertex a. Since a is extremal in Γ, these cannot both occur. $\square$

5 Construction of the HNN base

In this section, we construct a presentation of a group that will turn out to be an HNN base for G. As a first step, we fix names for the various vertices of Γ. Throughout we make the following assumptions:

- Γ is a minimal LOT of diameter 3, which cannot be spanned by fewer than three vertices.
- The non-extremal vertices of Γ are u and v.
- The unique vertex of Γ that appears twice as a label is a.
- Of the edges labelled a, one has its initial vertex in $\{u, v\}$ and its terminal vertex extremal, while the other has its initial vertex extremal and its terminal vertex in $\{u, v\}$.

- Neither $I(\Gamma)$ nor $T(\Gamma)$ is connected.

We know from Lemma 4.2 that Γ is then spanned by $\{a,u,v\}$. Let Δ denote the subtree of Γ whose vertex set is $\{a,u,v\}$. We give inductive definitions of two sequences $\{b_1,b_2,\dots,b_P\}$ and $\{c_1,c_2,\dots,c_Q\}$ of vertices of Γ, and two sequences $\{e_0,\dots,e_P\}$, $\{f_0,\dots,f_Q\}$ of edges of Γ as follows.

Define e_0 to be the edge of Γ whose label is a and whose terminal vertex is in $\{u,v\}$. For $i\geq 0$, assume inductively that e_i has been defined. If e_i is an edge of Δ, then we define $P=i$ and stop the construction of the sequences $\{b_1,b_2,\dots,b_P\}$ and $\{e_0,\dots,e_P\}$. Otherwise e_i joins one of $\{u,v\}$ to an extremal vertex other than a, and we define b_{i+1} to be that extremal vertex, and e_{i+1} to be the unique edge of Γ labelled b_{i+1}.

Similarly, define f_0 to be the edge of Γ whose label is a and whose initial vertex is in $\{u,v\}$. For $i\geq 0$, assume inductively that f_i has been defined. If f_i is an edge of Δ, then we define $Q=i$ and stop the construction of the sequences $\{c_1,c_2,\dots,c_Q\}$ and $\{f_0,\dots,f_Q\}$. Otherwise f_i joins one of $\{u,v\}$ to an extremal vertex other than a, and we define c_{i+1} to be that extremal vertex, and f_{i+1} to be the unique edge labelled by c_{i+1}.

Note that the $P+Q+3$ vertices $\{u,v,a,b_1,\dots,b_P,c_1,\dots,c_Q\}$ and the $P+Q+2$ edges $\{e_0,\dots,e_P,f_0,\dots,f_Q\}$ together form an admissible subgraph of Γ, which has euler characteristic 1 and hence is connected, and hence by minimality of Γ must be the whole of Γ. In other words

$$V=V(\Gamma)=\{u,v,a,b_1,\dots,b_P,c_1,\dots,c_Q\},$$

and

$$E=E(\Gamma)=\{e_0,\dots,e_P,f_0,\dots,f_Q\}.$$

We also introduce the following notation. For $i=1,\dots,P$, x_i denotes the unique non-extremal vertex of Γ (ie $x_i\in\{u,v\}$) incident with the edge e_{i-1}. For $i=1,\dots,Q$, y_i denotes the unique non-extremal vertex of Γ incident with the edge f_{i-1}. In other words, x_i is the vertex adjacent to b_i in Γ, and y_i is the vertex adjacent to c_i.

Lemma 5.1 (i) *If $x_2=\dots=x_P=u$, then $x_1=v$ and e_P is incident at v.*

(ii) *If $x_2=\dots=x_P=v$, then $x_1=u$ and e_P is incident at u.*

(iii) *If $y_2=\dots=x_Q=u$, then $y_1=v$ and f_Q is incident at v.*

(iv) *If $y_2=\dots=y_Q=v$, then $y_1=u$ and f_Q is incident at u.*

Proof We prove (i). The other proofs are similar.

Suppose first that $x_1 = x_2 = \ldots = x_P = u$, and consider the subgraph $\Gamma_0 = \mathrm{span}\{a,u\}$ of Γ. Since $\lambda(e_0) = a$ and e_0 is incident to u, we have $e_0 \in E(\Gamma_0)$, and since b_1 is an endpoint of e_0 we have $b_1 \in V(\Gamma_0)$. Similarly $e_1 \in E(\Gamma_0)$ and $b_2 \in V(\Gamma_0)$, and so on, until $e_P \in E(\Gamma_0)$. If e_P is incident with v, then $v \in V(\Gamma_0)$, and since Γ is spanned by $\{a,u,v\}$ it follows that $\Gamma = \Gamma_0$ is spanned by $\{a,u\}$, a contradiction. Otherwise, e_P joins a to u, in which case the vertices $a, u, p_1, \ldots, b_P$ and the edges $e_0, \ldots, e_P$ form an admissible subtree of Γ of diameter two, which again is a contradiction.

Now suppose that $x_1 = v$ and $x_2 = \ldots = x_P = u$, and let $\Gamma_0 = \mathrm{span}\{b_1, u\}$. Arguing as above, we see that Γ_0 contains the edges $e_1, \ldots, e_{P-1}$ and the vertices $u, b_1, \ldots, b_P$. If e_P is not incident at v, then it joins u to a, so e_P and a also belong to Γ_0. But then e_0 joins b_1 to v and has label a, so we also have $v \in V(\Gamma_0)$. Hence $\Gamma = \Gamma_0$ since Γ is spanned by $\{a,u,v\}$, and so Γ is spanned by $\{b_1, u\}$, a contradiction. $\square$

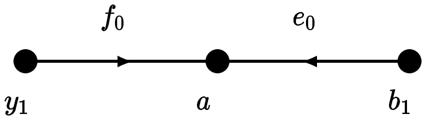
We next subdivide each of the sequences $\{b_i\}$, $\{c_i\}$ into two subsequences, depending on the orientation of the edges labelled by these vertices. Specifically, let:

- $p(1), \ldots, p(s)$ be the sequence, in ascending order, of integers i such that $0 < i \le P$ and $b_i = \tau(e_{i-1})$;
- $p'(1), \ldots, p'(s')$ be the sequence, in ascending order, of integers i such that $0 < i \le P$ and $b_i = \iota(e_{i-1})$;
- $q(1), \ldots, q(t)$ be the sequence, in ascending order, of integers i such that $0 < i \le Q$ and $c_i = \iota(f_{i-1})$; and
- $q'(1), \ldots, q'(t')$ be the sequence, in ascending order, of integers i such that $0 < i \le Q$ and $c_i = \tau(f_{i-1})$.

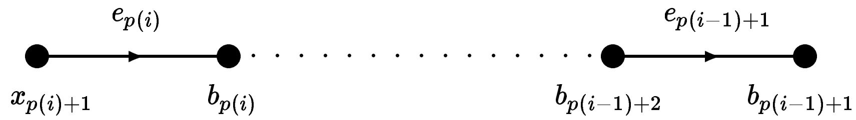
For consistency of notation in what follows, we set $p(0) = p'(0) = q(0) = q'(0) = 0$.

Thus each b_i, for $i = 1, \ldots, P$, can be written uniquely as $b_{p(j)}$ or as $b_{p'(j)}$, and each c_i, for $i = 1, \ldots, Q$, can be written uniquely as $c_{q(j)}$ or as $c_{q'(j)}$.

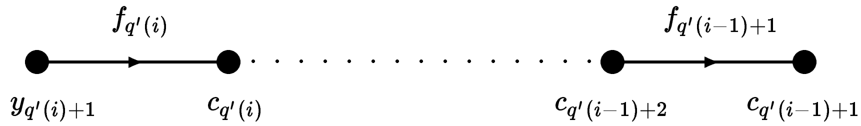
This notation allows us to give a more precise description of the structure of the initial and terminal graphs of Γ. Specifically, $I(\Gamma)$ is constructed from the vertices $\{a,u,v\}$ by adding two edges



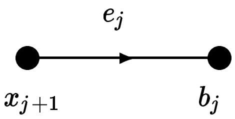
together with directed chains



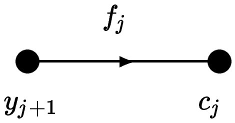
for each $i = 1, \ldots, s$, and



for each $i = 1, \ldots, t'$; and finally single edges



for $p(s) < j \leq P$ and



for $q'(t') < j \leq Q$.

In the above diagrams x_{P+1} and y_{Q+1} (which have not been defined) should be interpreted as $\iota(e_P)$ and $\iota(f_Q)$ respectively. Note that at most one of these is equal to a. (This happens if and only if a is the initial vertex of its incident edge in Γ.) All other x_j and y_j belong to $\{u, v\}$.

If $I(\Gamma)$ contains a directed cycle, for example, then this cycle must contain a. From the above, we see that this can happen only if $s = 1$, $p(1) = P$, and $x_{P+1} = a$.

The structure of $T(\Gamma)$ is entirely analogous, and similar remarks apply. We omit the details.

Now we are ready to construct a specific presentation for an HNN base for $G = G(\Gamma)$. Recall that G is given by a finite presentation

$$\mathcal{P}(\Gamma) = \langle V(\Gamma) \mid \iota(e)\lambda(e) = \lambda(e)\tau(e),\ e \in E(\Gamma)\rangle.$$

Since Γ is connected, we have $G^{ab} \cong \mathbb{Z}$, and the commutator subgroup G' is the normal closure in G of the subgroup $B = B(\Gamma)$ generated by the finite set $\{xy^{-1}\ ;\ x, y \in V(\Gamma)\}$. A theorem of Bieri and Strebel [2] says that G is an HNN extension of B with stable letter t (which can be taken to be any element of $V(\Gamma)$) and associated subgroups $A_0 = B \cap tBt^{-1}$ and $A_1 = B \cap t^{-1}Bt$:

$$G = \langle B, t \mid t^{-1}\alpha t = \phi(\alpha),\ \alpha \in A_0\rangle,$$

where $\phi\colon A_0 \to A_1$ is the isomorphism induced by conjugation by t.

Clearly B is finitely generated. It remains to prove that B is finitely presentable, and we do this by constructing an explicit set of defining relators.

Recall that our assumptions on Γ imply that each of $I(\Gamma)$ and $T(\Gamma)$ has precisely two connected components, with the vertices u, v belonging to separate components in each case.

Let F denote the subgroup of the free group on $V(\Gamma)$ generated by

$$\{xy^{-1}\ ;\ x, y \in V(\Gamma)\}.$$

Then F is free of rank $|V(\Gamma)| - 1 = |E(\Gamma)|$, and any basis for F can be chosen as a finite generating set for B. Rather than fix a specific basis for F, we proceed as follows. Let $\bar{K} = \bar{K}(\Gamma)$ be the maximal abelian cover of the 2–complex $K = K(\Gamma)$ associated to Γ (which is the standard 2–complex model of the presentation $\mathcal{P}(\Gamma)$). Then since K has a single 0–cell, we identify the 0–cells of $\bar{K}$ with integers, via the isomorphism $H_1(K) \cong G^{ab} \cong \mathbb{Z}$. The 1–cells of $\bar{K}$ with initial vertex $i \in \mathbb{Z}$ can be denoted w_i, where $w \in V(\Gamma)$, and each w_i has terminal vertex $i+1 \in \mathbb{Z}$. Let L be the 1–subcomplex of $\bar{K}$ with 0–cells $0, 1$ and 1–cells $\{w_0,\ w \in V(\Gamma)\}$. Then F is naturally identified with $\pi_1(L, 0)$.

We also construct a graph $\hat{L}$ and an immersion $\pi\colon \hat{L} \to L$ as follows. $V(\hat{L}) = \{0,1\} \times \{u, v\}$, $E(\hat{L}) = E(L)$, $\iota(w_0) = (0, x)$ where $x \in \{u, v\}$ belongs to the same component of $I(\Gamma)$ as w, and $\tau(w_0) = (1, y)$ where $y \in \{u, v\}$ belongs to the same component of $T(\Gamma)$ as w. The graph homomorphism π is defined to be the identity map on edges, and is defined on vertices by $\pi(i, u) = \pi(i, v) = i$, $i = 0, 1$. It is not difficult to see that $\hat{L}$ is connected. Indeed, if the edge of

Γ between u and v has label w, then the edges u, v, w of $\hat{L}$ form a spanning tree. Since π is bijective on edges, it is an immersion, and hence injective on fundamental groups. Indeed, the fundamental group $\hat{F}$ of $\hat{L}$ embeds as a free factor of $F = \pi_1(L)$ via π_*, as we can see by the following construction: add an edge X to $\hat{L}$ with $\iota(X) = (0, u)$ and $\tau(X) = (0, v)$, and an edge Y with $\iota(Y) = (1, u)$, $\tau(Y) = (1, v)$, to form a larger graph $\tilde{L}$. The immersion $\pi\colon \hat{L} \to L$ extends to a homotopy equivalence $\pi\colon \tilde{L} \to L$ that shrinks the edge X to the vertex 0, and the edge Y to the vertex 1. Hence we have

$$F = \pi_1(L) \cong \pi_1(\tilde{L}) = \pi_1(\hat{L}) * \langle X, Y \rangle.$$

Since the map $\pi\colon \hat{L} \to L$ is bijective on edges, any path in L which lifts to a path in $\hat{L}$ does so uniquely. Given a closed path C in L that lifts to a closed path $\hat{C}$ in $\hat{L}$, we define two related paths in L, namely the *forward derivative* $\partial_+ C$ of C and the *backward derivative* $\partial_- C$ of C, as follows. For $\partial_+ C$ we first fix a maximal subforest Φ_I of $I(\Gamma)$. Next, we cyclically permute $\hat{C}$ so that it begins and ends at one of the vertices $(1, u)$ or $(1, v)$. Hence $\hat{C}$ is a concatenation of length two subpaths of the form $x^{-1}y$, where $x, y \in E(\hat{L}) = V(\Gamma)$ belong to the same component of $I(\Gamma)$. The next step is to replace each such subword $x^{-1}y$ by the product

$$(x^{-1}z_0)(z_0^{-1}z_1)\dots(z_m^{-1}y),$$

where $(x, z_0, z_1, \dots, z_m, y)$ is the geodesic from x to y in Φ_I. We now have a concatenation of length 2 subwords of the form $x^{-1}y$ where x and y are joined by an edge in Φ_I. This edge corresponds to an edge of Γ, and hence to a defining relation in $\mathcal{P}(\Gamma)$ that can be written

$$x^{-1}y = gh^{-1}$$

for some $g, h \in V(\Gamma)$. The final step is to replace each such word $x^{-1}y$ by the corresponding word gh^{-1}. The result is a closed path $\partial_+ C$ in L.

Remarks (i) $\partial_+ C$ depends on the choice of maximal forest Φ_I, and then is well-defined only up to cyclic permutation.

(ii) If C' is a cyclic permutation of C, then C' also lifts to a closed path in $\hat{L}$, so $\partial_+ C'$ is defined. It is equal to (a cyclic permutation of) $\partial_+ C$.

(iii) The definition of $\partial_+ C$ does not depend on C being (cyclically) reduced. Indeed the insertion into C of a cancelling pair xx^{-1} may alter $\partial_+ C$. However, the insertion of a cancelling pair $x^{-1}x$ will *not* alter $\partial_+ C$.

(iv) C and $\partial_+ C$ are (freely) homotopic in $\bar{K}$ (since the last part of the construction involves replacing a path $x^{-1}y$ by a homotopic path gh^{-1}). In particular, if C is nullhomotopic in $\bar{K}$, then so is $\partial_+ C$.

(v) The unique lift of $\partial_+ C$ in $\tilde{L}$ does not contain the edge Y.

The backward derivative $\partial_- C$ is defined similarly. This time we fix a maximal forest Φ_T of $T(\Gamma)$, and choose a cyclic permutation of $\hat{C}$ beginning at $(0,u)$ or $(0,v)$, split $\hat{C}$ into subpaths of the form xy^{-1} with x,y in the same component of $T(\Gamma)$, and then use relations of $\mathcal{P}$ corresponding to edges of Φ_T to transform $\hat{C}$. Remarks analogous to the above hold also for $\partial_- C$.

Now consider the unique cycle in $T(\Gamma)$. If $z_0,\dots,z_m$ are the vertices of this cycle in cyclic order, define $\hat{R}_0$ to be the nullhomotopic path

$$(z_m z_0^{-1})(z_0 z_1^{-1})\dots(z_{m-1} z_m^{-1})$$

in $\hat{L}$ and $R_0 = \pi(\hat{R}_0)$ the corresponding nullhomotopic path in L. Now define $R_1 = \partial_- R_0$. If R_1 lifts to $\hat{L}$ then define $R_2 = \partial_- R_1$, and so on. In this way we obtain either an infinite sequence $R_1, R_2, \dots$ of paths in L, or a finite sequence $R_1,\dots,R_M$ such that R_M does not lift to $\hat{L}$.

In a similar way, the unique cycle in $I(\Gamma)$ determines a nullhomotopic closed path S_0 in L that lifts to $\hat{L}$, so a sequence $S_1,\dots$ of closed paths in L (finite or infinite), such that $S_i = \partial_+ S_{i-1}$ for each $i \geq 1$, and if the sequence is finite with final term S_N then S_N does not lift to $\hat{L}$.

Lemma 5.2 *The paths R_i and S_j are all nullhomotopic in $\bar{K}$.*

Proof This follows by induction and Remark (iv) above, since R_0 and S_0 are nullhomotopic. □

Now suppose that the sequence $\{R_i\}$ contains at least m terms. We construct a 2–complex L_m as follows. The 1–skeleton of L_m is the subcomplex of $\bar{K}$ consisting of L, together with the 0–cells $2,\dots,m+1$ and the 1–cells $u_1, v_1, \dots, u_m, v_m$. Then L_m has precisely m 2–cells attached to L using the paths $R_1,\dots,R_m$. We also consider the full subcomplex $\bar{K}_m$ of $\bar{K}$ on the set $\{0,1,\dots,m+1\}$ of 0–cells.

Lemma 5.3 *The 2–complexes L_m and $\bar{K}_m$ are homotopy equivalent.*

Proof We argue by induction on m, there being nothing to prove in the case $m=0$. Let γ denote the covering transformation of $\bar{K}$ that sends a 0–cell $n \in \mathbb{Z}$ to $n+1$. Note that the link of the 0–cell $m+1$ in $\bar{K}_m$ is naturally identifiable with the graph $T(\Gamma)$. Let d be the unique edge in $E(\Gamma) = E(T(\Gamma))$ that does not belong to the maximal forest $\Phi_T \subset T(\Gamma)$. Then d is contained in the unique cycle in $T(\Gamma)$, so R_0 has a subword xy^{-1}, where x,y are the endpoints of d in $T(\Gamma)$. Corresponding to d is a relator $xy^{-1}h^{-1}g$ in $\mathcal{P}$, which lifts to a 2–cell α with boundary path $x_m y_m^{-1} h_{m-1}^{-1} g_{m-1}$ in $\bar{K}_m$. Modulo the other 2–cells of $\bar{K}_m$, the boundary path of α is homotopic to $\gamma^m(R_0)^{-1}\cdot\gamma^{m-1}(R_1)$. Since R_0 is nullhomotopic in the 1–skeleton of $\bar{K}$, this is in fact homotopic to $\gamma^{m-1}(R_1)$. This in turn is homotopic (in $\bar{K}_{m-1}$) to $\gamma^{m-2}(R_2)$, etc. Repeating this argument, we see that the boundary path of α is homotopic in $\bar{K}_m\backslash\alpha$ to R_m. A simple homotopy move allows us to replace α by a 2–cell whose boundary path is R_m.

The link of $m+1$ in the resulting 2–complex K' is then isomorphic to $T(\Gamma)\backslash d = \Phi_T$. Since Φ_T is a forest with two components (one containing u and the other containing v), it collapses to the graph with no edges and vertex set $\{u,v\}$. Each move in this collapsing process (removing a vertex and an edge from the graph) can be mirrored by a collapse in the 2–complex K' (removing a 1–cell and a 2–cell that are incident at the 0–cell $m+1$). After performing all these collapsing moves, we are left with a 2–complex K'', simple homotopy equivalent to $\bar{K}_m$. By inspection, K'' is formed from $\bar{K}_{m-1}$ by adding a 2–cell with boundary path R_m, a 0–cell $m+1$, and two 1–cells u_m, v_m, each joining m to $m+1$.

By inductive hypothesis, $\bar{K}_{m-1}$ is homotopy equivalent to L_{m-1}, so $\bar{K}_m$ is homotopy equivalent to the 2–complex obtained from L_{m-1} by adding a 2–cell with boundary path R_m, a 0–cell $m+1$, and two 1–cells u_m, v_m, each joining m to $m+1$. But this 2–complex is precisely L_m, and the proof is complete. $\square$

Remark An analogous result holds for the S_j. We omit the details, but will use this result implicitly in what follows.

Corollary 5.4 *If $R_1,\dots,R_m$ and $S_1,\dots,S_n$ are all defined, then $m+n < |V(\Gamma)|$.*

Proof By the Lemma and its analogue for the S_j, $\bar{K}_m$ is homotopy equivalent to a 2–complex formed from L by attaching m 2–cells and then wedging on m circles; and $\gamma^{-n}(\bar{K}_n)$ is homotopy equivalent to a complex obtained from L by adding n 2–cells and then wedging on n circles. Since $\gamma^{-n}(\bar{K}_{m+n}) = \gamma^{-n}(\bar{K}_n)\cup \bar{K}_m$, with $\gamma^{-n}(\bar{K}_n)\cap\bar{K}_m = \bar{K}_1 = L$, it follows that $\gamma^{-n}(\bar{K}_{m+n})$ is homotopy equivalent to a complex formed from L by adding $m+n$ 2–cells and then wedging on $m+n$ circles. Hence $\beta_1(\bar{K}_{m+n}) \geq m+n$. Now $H_2(K)=0$, and $\bar{K}$ is a $\mathbb{Z}$–cover of K, so $H_2(\bar{K})=0$ by [1], Proposition 1. Hence also $H_2(K')=0$

for any subcomplex $K' \subseteq K$. In particular $H_2(\bar{K}_{m+n}) = 0 = H_2(L)$. Since also $H_0(\bar{K}_{m+n}) = \mathbb{Z} = H_0(L)$ and $\chi(\bar{K}_{m+n}) = \chi(L) = 2 - |V(\Gamma)|$, it follows that

$$m + n \leq \beta_1(\bar{K}_{m+n}) = \beta_1(L) = |V(\Gamma)| - 1. \qquad \square$$

Corollary 5.5 *Each of the sequences $\{R_i\}$ and $\{S_j\}$ are finite, and if the final terms are R_M and S_N respectively then $M + N < |V(\Gamma)|$.*

We claim that the finite sequences $\{R_i\}$ and $\{S_j\}$ form a full set of defining relators for the HNN base B of G, which completes the proof of our Theorem 1.1. In order to prove this claim, we need to derive some further information about the structure of the words R_i and S_j.

Remark The definitions of R_i and S_i depend, *a priori*, on specific choices for the maximal forests Φ_T and Φ_I respectively. Suppose we were to choose a different maximal tree Φ'_I in $I(\Gamma)$, for example. Then geodesics in Φ_I and Φ'_I would differ at most by the unique cycle in $I(\Gamma)$. It follows from this that the resulting definitions of $\partial_+ C$, for any closed path C in L that lifts to $\hat{L}$, are equal modulo the normal closure of S_1. An easy induction shows that, for any i, the definitions of S_i resulting from different choices of Φ_I are equal modulo the normal closure of $\{S_1, \dots, S_{i-1}\}$. Hence our set of defining relators does not depend in an essential way upon the choices of maximal forests Φ_I and Φ_T.

6 Structure of the relations

In this section we examine the structure of the proposed defining relators R_i and S_i of the HNN base B for G. Recall that each of R_i and S_i is a closed path in the 2–complex L, and that we have a homotopy equivalence $\pi\colon \tilde{L} \to L$, which restricts to an edge-bijective graph immersion on $\hat{L} = \tilde{L} \backslash \{X, Y\}$ and shrinks each of the 1–cells X, Y to a point. Let $\tilde{C}$ denote the unique (up to cyclic permutation) cyclically reduced closed path in $\tilde{L}$ that maps to a given cyclically reduced closed path C in L. Then C lifts to $\hat{L}$ if and only if $\tilde{C}$ is a path in $\hat{L}$, in which case $\tilde{C}$ is the unique lift. By definition, each R_i (resp S_i) is defined if and only if R_{i-1} (resp S_{i-1}) lifts to $\hat{L}$. Hence $\tilde{R}_i$ is a path in $\hat{L}$ for $1 \leq i \leq M-1$, and S_i is a path in $\hat{L}$ for $1 \leq i \leq N-1$. Moreover, the path $\tilde{R}_M$ involves Y but not X, while the path $\tilde{S}_N$ involves X but not Y.

For any group A and letter Z, we say that a word $w \in A * \langle Z \rangle$ is *positive* (resp *negative*) in Z if only positive (resp negative) powers of Z occur in w. We say that w is *strictly positive* (resp *strictly negative*) if in addition at least one positive (resp negative) power of Z does occur in w, in other words $w \notin A$.

We will concentrate our attention on the relators S_i. The analysis of the R_i is entirely analogous.

We first treat the case where $I(\Gamma)$ contains a directed cycle C.

Theorem 6.1 *Suppose that the unique cycle C in $I(\Gamma)$ is directed. Then:*

- $N = 1$;
- $\tilde{S}_1$ *is either strictly positive or strictly negative in* X;
- S_1 *involves each of* $a, b_1, \ldots, b_P$ *exactly once, and no* c_j;
- *each of* $a, b_1, \ldots, b_P$ *is an extremal source in* Γ.

Proof The vertex a is contained in C, by Lemma 4.5, (v). Since $\iota(f_0) \in \{u, v\}$, f_0 is not an edge of C, so the edge of C coming into a is e_0. Hence $b_1 = \iota(e_0)$ is a vertex of C, and since e_1 is the only edge with $\lambda(e_1) = b_1$, it is also an edge of C, and so on. Hence each of $b_1, \ldots, b_P$ are vertices of C, $\iota(e_P) = a$, and the edges of C are precisely $e_P, \ldots, e_0$ (in the order of the orientation of C). Each of the vertices of C is extremal in Γ, and since it is the initial vertex of an edge of $I(\Gamma)$ it is also the initial vertex of an edge of Γ, ie a source in Γ. Moreover

$$S_0 = (a^{-1}b_P)(b_P^{-1}b_{P-1}) \ldots (b_1^{-1}a),$$

so

$$S_1 = \partial_+ S_0 = (b_P \tau(e_P)^{-1})(b_{P-1}x_P^{-1}) \ldots (b_1 x_2^{-1})(a x_1^{-1}),$$

where each $x_i \in \{u, v\}$.

Suppose that S_1 lifts to $\hat{L}$. Then $\tau(e_P)$ belongs to the same component of $I(\Gamma)$ as b_{P-1}, x_P to the same component as b_{P-2}, and so on. Since $a, b_1, \ldots, b_P$ all belong to the same component of $I(\Gamma)$, it follows that the x_i also all belong to the same component. But u and v belong to different components of $I(\Gamma)$, and so the x_i are all equal, which contradicts Lemma 5.1.

Hence S_1 does not lift to $\hat{L}$, and so $N = 1$. Moreover, by the above argument, some of the x_i belong to the opposite component of $I(\Gamma)$ from a. If a, u belong to the same component of $I(\Gamma)$, this means that some of the x_i are equal to v. Then $\tilde{S}_1$ is formed from S_1 by replacing each occurrence of v^{-1} by $v^{-1}X^{-1}$, and so $\tilde{S}_1$ is strictly negative in X. Similarly, if a, v belong to the same component of $I(\Gamma)$, then $\tilde{S}_1$ is strictly positive in X. □

For the rest of the section, we can assume that the cycle C is not directed. Then $y_1 = \iota(f_0) = \iota(e_{p(1)}) \in \{u, v\}$. We may assume that $y_1 = u$. Then C has the form

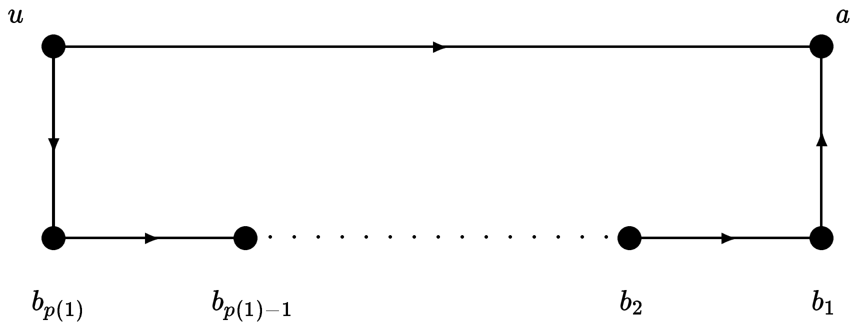


Figure 1

For the purpose of defining forward derivatives, and hence the S_i, we fix Φ_I to be the maximal subforest of $I(\Gamma)$ obtained by removing the edge f_0 (the edge joining u to a in C).

For $k \le \min(s, t' + 1)$, let $I_k(\Gamma)$ denote the subgraph of Φ_I consisting of the edges $\{e_i, 0 \le i \le p(k)\}$ and $\{f_i, 1 \le i \le q'(k-1)\}$, together with all their incident vertices. Note that I_k contains no more than two components, one contained in each component of Φ_I. Hence whenever two vertices of I_k belong to the same component of Φ_I, then the geodesic between them is also contained in I_k.

Theorem 6.2 *Suppose that the cycle in $I(\Gamma)$ has the form shown in Figure 1. Then:*

(i) *Each S_i can be written, up to cyclic permutation, in the form $aU_ia^{-1}V_i$, where U_i is a word in*
$$\{a, u, v, c_1, \ldots, c_{q'(i-1)+1}\};$$
and V_i is a word in
$$\{a, u, v, b_1, \ldots, b_{p(i)+1}\}.$$

(ii) *If $p(i) < P$, then V_i contains a single occurrence of $b_{p(i)+1}$ and does not contain a.*

(iii) *If $q'(i-1) < Q$, then U_i contains a single occurrence of $c_{q'(i-1)+1}$ and does not contain a.*

(iv) *Every letter occurring in S_i, other than $b_{p(i)+1}$ and $c_{q'(i-1)+1}$, is a vertex of the subgraph $I_i \subseteq I(\Gamma)$.*

(v) *If* $p(i) = P$ *or* $q'(i-1) = Q$ *then* $i = N$.

Proof We prove this by induction on i, the initial case being when $i = 1$. We have

$$S_0 = (u^{-1}a)(a^{-1}b_1)(b_1^{-1}b_2)\dots(b_{p(1)}^{-1}u),$$

so

$$S_1 = \partial_+ S_0 = (ac_1^{-1})(x_1a^{-1})(x_2b_1^{-1})\dots(x_{p(1)}b_{p(1)-1}^{-1})(b_{p(1)+1}b_{p(1)}^{-1})$$

(if $p(1) < P$). The vertices $a, u, b_1, \dots, b_{p(1)}$ are contained in I_1, but not c_1, $b_{p(1)+1}$. The first four statements of the result (for $i = 1$) follow, setting $U_1 = c_1^{-1}x_1$ and

$$V_1 = (x_2b_1^{-1})\dots(x_{p(1)}b_{p(1)-1}^{-1})(b_{p(1)+1}b_{p(1)}^{-1}).$$

For the last statement, certainly $Q > 0 = q'(0)$. Suppose that $p(1) = P$ and $i < N$. Then

$$S_1 = (ac_1^{-1})(x_1a^{-1})(x_2b_1^{-1})\dots(x_Pb_{P-1}^{-1})(\tau(e_P)b_P^{-1})$$

lifts to $\hat{L}$, so each of $x_2, \dots, x_P$ belongs to the same component of $I(\Gamma)$ as $a, b_1, \dots, b_{P-1}$, in other words $x_2 = \dots = x_P = u$. By Lemma 5.1 we have $x_1 = v$ and e_P incident with v. But $\iota(e_P) = u$ so $\tau(e_P) = v$, which does not belong to the same component of $I(\Gamma)$ as b_{P-1}. It follows that S_1 does not, after all, lift to $\hat{L}$, a contradiction.

This completes the proof of the initial case of the induction.

Now assume inductively that $i > 1$ and the result is true for $i-1$. In particular, $i - 1 < N$, so $p(i-1) < P$ and $q'(i-2) < Q$. Hence U_{i-1} contains a single occurrence of $c_{q'(i-2)+1}$, V_{i-1} contains a single occurrence of $b_{p(i-1)+1}$, and every other letter occurring in S_{i-1} is a vertex of the subgraph I_{i-1} of $I(\Gamma)$. Consider the construction of $S_i = \partial_+ S_{i-1}$ from S_{i-1}. We first write a suitable cyclic permutation of S_{i-1} as a product of length two subwords of the form $g^{-1}h$. For all but two of these subwords, both g and h are vertices of I_{i-1}. (There are precisely two exceptions, since the occurrences of $b_{p(i-1)+1}$ and $c_{q'(i-2)+1}$ in S_{i-1} are separated at least by an occurrence of $a^{\pm 1}$.)

Suppose first that g, h are vertices of I_{i-1}. The next step is to replace $g^{-1}h$ by the product

$$(g^{-1}z_1)(z_1^{-1}z_2)\dots(z_t^{-1}h)$$

where $g, z_1, z_2, \dots, z_t, h$ are the vertices on the geodesic from g to h in Φ_I. This geodesic is contained in I_{i-1}, so each bracketed term here is $(\iota(e)^{-1}\lambda(e))^{\pm 1}$ for some edge e of I_{i-1}. The final step is to replace this by $(\lambda(e)\tau(e)^{-1})^{\pm 1}$. Note that $\tau(e)$ is a vertex of I_i, and $\tau(e) \neq a$. Also, none of the intermediate vertices z_i in the geodesic is equal to a, since a is an extremal vertex of Φ_I. Note that, if $g^{-1}h$ is a subword of U_{i-1}, then all letters in the resulting subword of S_i come from $\{u, v, c_1, \ldots, c_{q'(i-1)}\}$, while if it is a subword of $a^{-1}V_{i-1}a$ then all letters come from $\{a, u, v, b_1, \ldots, b_{p(i)}\}$.

A similar argument holds if, say $g = b_{p(i-1)+1}$. Here, however, the geodesic from g to h is not contained in I_{i-1}. It is the union of the geodesic from $b_{p(i-1)+1}$ to z in I_i, where $z \in \{u, v\}$, with the geodesic (in I_{i-1}) from z to h. Edges in I_{i-1} give rise to length 2 subwords of S_i consisting of letters which are vertices in I_i. The same is true for an edge e_j from b_j to b_{j+1}, for $p(i-1) < j < p(i)$. (The corresponding word is $x_j b_j^{-1}$.) Finally, the edge $e_{p(i)}$ (from $b_{p(i)}$ to z) contributes a subword $\tau(e_{p(i)})b_{p(i)}^{-1}$. If $p(i) < P$ then $\tau(e_{p(i)}) = b_{p(i)+1}$; otherwise $\tau(e_{p(i)}) \in \{a, u, v\}$.

The analysis if $h = b_{p(i-1)+1}$, or if one of g, h is $c_{q'(p-2)+1}$ is similar to the above.

Each of the two subwords $g^{-1}h$ of S_{i-1} that contain the letter a gives rise to a subword of S_i containing an occurrence of a with the same exponent. If $g = a$ then the subword begins $(x_1 a^{-1})\ldots$, while if $h = a$ then the subword ends $\ldots(a x_1^{-1})$. If $p(i) < P$ and $q'(i-1) < Q$ then this will be the only occurrence of a in this subword of S_i.

Statements (i)–(iv) follow.

To prove (v), suppose for example that $i < N$ and $p(i) = P$. Another induction on i shows that $x_2 = \ldots = x_P = u$. An argument similar to that given above in the initial case of the induction again gives rise to a contradiction: by Lemma 5.1, $\tau(e_P) = v$, which does not belong to the same component of $I(\Gamma)$ as b_{P-1}, so S_i does not lift to $\hat{L}$ and $i = N$.

If $i < N$ and $q'(i-1) = Q$ then a similar argument applies. Here we can show that $y_1 = \ldots = y_Q = x_1 \in \{u, v\}$, which contradicts Lemma 5.1. □

This result contains all the necessary information about S_i if $i < N$. We now need to investigate further the structure of $\tilde{S}_N$, particularly as regards occurrences of X. Note that, up to cyclic permutation, we have $\tilde{S}_N = a\tilde{U}_N a^{-1}\tilde{V}_N$, by Theorem 6.2 (i).

Lemma 6.3 *Each of $\tilde{U}_N$, $\tilde{V}_N$ is either positive or negative in X.*

Proof As indicated in the proof of Theorem 6.2, all of V_N, except for the part arising from the geodesic γ from $b_{p(N-1)+1}$ to u, consists of letters which are vertices in I_{N-1}. All of these vertices are in the same component of $I(\Gamma)$ as u. The part of V_N arising from γ is

$$[(x_{p(N-1)+2}b^{-1}_{p(N-1)+1})\dots(x_{p(N)}b^{-1}_{p(N)-1})(\tau(e_{p(N)})b^{-1}_{p(N)})]^{\pm 1},$$

or, if γ passes through a (ie if $\iota(e_{p(N)})=a$):

$$[(x_{p(N-1)+2}b^{-1}_{p(N-1)+1})\dots(\tau(e_{p(N)})b^{-1}_{p(N)})(x_1a^{-1})\dots(b_{p(1)+1}b^{-1}_{p(1)})]^{\pm 1}.$$

The expression in square brackets is a product of terms gh^{-1} with h in the same component of $I(\Gamma)$ as u. To lift to $\tilde{L}$, we replace $h^{-1}g$ by $h^{-1}Xg$ whenever g belongs to the same component of $I(\Gamma)$ as v and h to the same component as u, and by $h^{-1}X^{-1}g$ if g belongs to the same component as u and h to the same component as v. Hence $\tilde{V}_N$ is either positive or negative in X

A similar argument applies to $\tilde{U}_N$, replacing u by x_1 in the above. □

We will also need to investigate possible occurrences of a in S_N other than those indicated in Theorem 6.2.

Lemma 6.4 *The words $\tilde{U}_N$ and $\tilde{V}_N$ contain in total at most one occurrence of a.*

Proof From the discussion in the proof of Lemma 6.3, the word V_N (and hence also $\tilde{V}_N$) contains a single occurrence of a if $e_{p(N)}$ is incident with a in Γ, and no occurrence of a otherwise. Similarly U_N (and hence also $\tilde{U}_N$) contains a single occurrence of a if $f_{q'(N-1)}$ is incident with a in Γ, and no occurrence of a otherwise. The result now follows from the fact that a is extremal in Γ. □

7 Completion of the proof

Define

$$\begin{aligned} G_0 &= \pi_1(\hat{L})/\{R_1,\dots,R_{M-1},S_1,\dots,S_{N-1}\},\\ G_+ &= (G_0 * \langle X\rangle)/\{\tilde{S}_N\},\\ G_- &= (G_0 * \langle Y\rangle)/\{\tilde{R}_M\},\end{aligned}$$

and

$$G_1 = (G_0 * \langle X,Y\rangle)/\{\tilde{R}_M,\tilde{S}_N\} \cong (\pi_1(L))/\{R_1,\dots,R_M,S_1,\dots,S_N\}.$$

Lemma 7.1 *The group G_0 is free.*

Proof By Theorems 6.1 and 6.2, and the analogous results for the R_i, the set of $M+N-2$ distinct numbers $\mathcal{B} = \{p(1)+1,\dots,p(N-1)+1, p'(0)+1,\dots,p'(M-2)+1\}$ has the property that each $j \in \mathcal{B}$ is the greatest index of a b–letter occurring in a unique relator R_i or S_i, and moreover that relator contains precisely one occurrence of b_j.

It follows that the 1–complex L' obtained from $\hat{L}$ by removing the 1–cells b_j, $j \in \mathcal{B}$ is connected, with fundamental group isomorphic to G_0. □

Lemma 7.2 *The natural maps $G_0 \to G_+$ and $G_0 \to G_-$ are injective.*

Proof We show that the map $G_0 \to G_+$ is injective. The proof of injectivity of $G_0 \to G_-$ is entirely analogous. Since G_0 is a free group and G_+ is a one-relator group $G_+ = (G_0 * \langle X \rangle)/\{\tilde{S}_N\}$, we need only show that $\tilde{S}_N$, regarded as a word in $(G_0 * \langle X \rangle)$, genuinely involves X. The result then follows from the Freiheitssatz for one-relator groups [10].

Consider the various possibilities for the structure of $\tilde{S}_N$. If the initial graph $I(\Gamma)$ contains a directed cycle, then $N=1$ and $\tilde{S}_1$ is a strictly positive (or strictly negative) word in X, by Theorem 6.1. Thus $\tilde{S}_1$, regarded as a word in the free product $G_0 * \langle X \rangle$, is also strictly positive (or strictly negative) in X, and so genuinely involves X.

Suppose then that $I(\Gamma)$ does not contain a directed cycle. By Theorem 6.2 (i) and Corollary 6.3 we have (up to cyclic permutation) $\tilde{S}_N = a\tilde{U}_N a^{-1}\tilde{V}_N$, with each of $\tilde{U}_N$ and $\tilde{V}_N$ being either positive or negative in X. We also have $\tilde{S}_N$ definitely involving X, since otherwise S_N would lift to $\hat{L}$.

If X occurs in $\tilde{S}_N$ with nonzero exponent-sum, then occurrences of X survive modulo the relators $R_1,\dots,R_{M-1},S_1,\dots,S_{N-1}$, so we may assume that X appears with exponent-sum zero. Thus one of $\tilde{U}_N$, $\tilde{V}_N$ is strictly positive, and the other is strictly negative, with precisely the same number of occurrences of $X^{\pm 1}$. We may rewrite $\tilde{S}_N$ (again, up to cyclic permutation) as

$$\tilde{S}_N = XA_1X\dots A_tXW_1X^{-1}B_tX^{-1}\dots B_1X^{-1}W_2$$

for some $t \geq 0$ and words A_i, B_i and W_1, W_2 that do not involve X. If we can show that neither W_1 nor W_2 is equal to the identity element in G_0, then it will follow that the above expression for $\tilde{S}_N$ does not allow for cancellation of X–symbols, when reducing modulo the relators of G_0. The result will follow.

Now a occurs with exponent-sum zero in each of the relators $R_1, \ldots, R_{M-1}$ and $S_1, \ldots, S_{N-1}$ of the group G_0, by Theorem 6.2. If neither U_N nor V_N contains the letter a, then each of W_1, W_2 contains precisely one occurrence of a, and so has infinite order in G_0. In particular, they are nontrivial in G_0, as required.

This reduces us to the case where one of U_N, V_N involves the letter a. By Corollary 6.4 we know that this can happen for only one of U_N, V_N.

First suppose that a occurs in U_N. Then $q'(N-1) = Q$ (and so also $N > 1$). As in the proof of Corollary 6.3, the part of U_N that gives rise to occurrences of X comes from the geodesic δ in Φ_I from $c_{q'(N-2)+1}$ to x_1. The relevant subword of U_N has the form:

$$[(y_{q'(N-2)+2}c_{q'(N-2)+1}^{-1}) \ldots (y_Q c_{Q-1}^{-1})(\tau(f_Q)c_Q^{-1})]^{\pm 1},$$

or, if δ passes through a:

$$[(y_{q'(N-2)+2}c_{q'(N-2)+1}^{-1}) \ldots (\tau(f_Q)c_Q^{-1})(x_1 a^{-1}) \ldots (b_{p(1)+1}b_{p(1)}^{-1})]^{\pm 1}.$$

The occurrences of X in $\tilde{U}_N$ correspond to those y_j, $j \geq q'(N-2)+2$ that are not equal to x_1, and also from $\tau(f_Q)$ if this is not in the same component of $I(\Gamma)$ as x_1. In the case where δ passes through a, we see that, in $\tilde{S}_N = a\tilde{U}_N a^{-1}\tilde{V}_N$ the a–letters that occur in the same W_i have the same exponent, and hence the W_i are both nontrivial in G_0, as required. In the other case, $\tau(f_Q) = a$ and the unique occurrence of c_Q in $\tilde{V}_N$ lies on the same side of all the X–letters as the unique occurrence of a. Hence c_Q occurs (precisely once) in the same W_i that contains two a–letters. To prove that this W_i is nontrivial in G_0, it suffices to show that c_Q does not occur in any of the relators $R_1, \ldots, R_{M-1}$ or $S_1, \ldots, S_{N-1}$. But c_Q can occur in S_j $(j < N)$ only if $j = N-1$ and $q'(N-2) = Q-1$, while c_Q can occur in R_j $(j < M)$ only if $j = M-1$ and $q(M-1) = Q-1$. In either case $y_2 = \ldots = y_Q = x_1$ (since R_{M-1} and S_{N-1} lift to $\hat{L}$) and f_Q joins a to x_1, which contradicts Lemma 5.1.

Suppose next that a occurs in V_N. Then $p(N) = P$. The occurrences of X in $\tilde{V}_N$ arise as indicated in the proof of Corollary 6.3. The relevant subword of V_N has the form:

$$[(x_{p(N-1)+2}b_{p(N-1)+1}^{-1}) \ldots (x_P b_{P-1}^{-1})(\tau(e_P)b_P^{-1})]^{\pm 1},$$

or, if γ passes through a:

$$[(x_{p(N-1)+2}b_{p(N-1)+1}^{-1}) \ldots (\tau(e_P)b_P^{-1})(x_1 a^{-1}) \ldots (b_{p(1)+1}b_{p(1)}^{-1})]^{\pm 1}.$$

The occurrences of X in $\tilde{V}_N$ correspond to those x_j, $j \geq p(N-1)+2$ in this subword that are equal to v, and also to $\tau(e_P)$ if $\tau(e_P) = v$. If $a = \tau(e_P)$ then since

$$\tilde{S}_N \sim a\tilde{U}_N a^{-1}\tilde{V}_N \sim XA_1X \dots A_tXW_1X^{-1}B_tX^{-1} \dots B_1X^{-1}W_2$$

we see that the two a–letters that occur in the same W_i have the same exponent, and hence both W_i are nontrivial in G_0, as required.

If $a = \iota(e_P)$ then γ passes through a. Assume for the moment that $x_1 = u$. Then the unique occurrence of b_P in $\tilde{U}_N$ lies on the same side of all the X–letters as the unique occurrence of a. Hence the W_i that contains two a–letters also contains a single occurrence of b_P. To prove that this W_i is nontrivial in G_0, it suffices to show that b_P does not occur in any of the relators $R_1, \dots, R_{M-1}$ or $S_1, \dots, S_{N-1}$ of G_0. But b_P can occur in S_j $(j < N)$ only if $j = N-1$ and $p(N-1) = P-1$, while if b_P occurs in R_j $(j < M)$, then $j = M-1$ and $p'(M-2) = P-1$. In either case $x_1 = \dots = x_P = u$, contradicting Lemma 5.1.

This last argument does not apply if $x_1 = v$. In this case we still have $x_2 = \dots = x_P = u$, and since $a = \iota(e_P)$ it follows from Lemma 5.1 that $\tau(e_P) = v$. If, say, $W_1 = 1$ in G_0, then $A_t = vb_P^{-1}$ and $A_tW_1B_t = A_tB_t \neq 1$ in G_0, since this word contains a single occurrence of b_P, which by similar arguments to the above cannot occur in any of the relators of G_0. Hence no more than one pair of letters $X^{\pm 1}$ in S_N can cancel modulo the relators of G_0, and so S_N, as a word in $G_0 * \langle X \rangle$, definitely involves X, as required.

This completes the proof of the Lemma. □

Corollary 7.3 *The maps $G_{\pm} \to G_1$ are injective.*

Proof The commutative square

$$\begin{tikzcd} G_0 \arrow[r] \arrow[d] & G_+ \arrow[d] \\ G_- \arrow[r] & G_1 \end{tikzcd}$$

is a pushout, and the maps $G_0 \to G_{\pm}$ are injective by the lemma. Hence G_1 is the free product of G_+ and G_-, amalgamated over G_0. □

Let L_+ be the 1–complex obtained from $\hat{L}$ by identifying the 0–cells $(0,u)$ and $(0,v)$ to a single 0–cell 0. Then L_+ is homotopy equivalent to the subcomplex $\hat{L}\cup X$ of $\tilde{L}$, and G_+ is a homomorphic image of the free group $\pi_1(\hat{L})*\langle X\rangle$, which is naturally identifiable with $\pi_1(L_+)$. Let us fix the 0–cell 0 as a basepoint for L_+, and consider the generating set

$$B_+=\{\theta_e=\tau(e)\lambda(e)^{-1}\ ;\ e\in E(\Gamma)\}$$

for $\pi_1(L_+,0)$. Note that B_+ is not a basis, since the unique cycle in $T(\Gamma)$ gives rise to a relation R_0 among the θ_e. However, this is the only relation, in the sense that $\pi_1(L_+,0)$ has a one-relator presentation $\langle B_+\mid R_0\rangle$.

Similarly, if L_- is obtained from $\hat{L}$ by identifying the 0–cells $(1,u)$ and $(1,v)$ to a single 0–cell 1, then G_- is a homomorphic image of the free group $\pi_1(L_-,1)$, which is generated by

$$B_-=\{\phi_e=\lambda(e)^{-1}\iota(e)\ ;\ e\in E(\Gamma)\}$$

modulo a single relator S_0 arising from the unique cycle in $I(\Gamma)$.

Theorem 7.4 *The correspondence $\theta_e\leftrightarrow\phi_e$ $(e\in E(\Gamma))$ induces a group isomorphism $G_+\leftrightarrow G_-$.*

Proof The relation R_0 among the generators B_+ is precisely the nullhomotopic path R_0 in L, which lifts to L_+ (indeed to $\hat{L}$). Under the isomorphism $\Psi\colon F(B_+)\to F(B_-)$ induced by the map $\theta_e\mapsto\phi_E$, this relation R_0 is mapped to $\partial_-R_0=R_1$, which is a relation in G_-. Hence we have an induced homomorphism $\pi_1L_+\to G_-$. In order to show that this in turn induces a homomorphism $G_+\to G_-$, we must show that each relation of G_+ is mapped to a relation of G_-.

Each word R_i, $1\le i\le M-1$ is mapped under Ψ to $\partial_+R_i=R_{i+1}$, which is a relation in G_-. Similarly, for $1\le j\le N$ we have $\Psi^{-1}(S_{j-1})=\partial_-S_{j-1}=S_j$, so $\Psi(S_j)=S_{j-1}$, which is also a relation in G_-. Hence Ψ induces a group homomorphism $G_+\to G_-$, as claimed. Similarly Ψ^{-1} induces a group homomorphism $G_-\to G_+$, and these homomorphisms are mutually inverse isomorphisms, by standard arguments. □

Corollary 7.5 *$G(\Gamma)$ is isomorphic to an HNN extension of the finitely presented group G_1, with associated subgroups $G_\pm$.*

Proof This is an easy exercise, given the isomorphism described in the previous lemma. □

This completes the proof of our main result, Theorem 1.1.

8 Further remarks

In the proof of Theorem 1.1, we have relied heavily on one-relator theory to show that our HNN base G_1 is indeed defined by the relators R_i and S_i. If we look at LOTs of larger diameter, we no longer have these tools at our disposal.

As long as $I(\Gamma)$ and $T(\Gamma)$ each have only two components (and hence only one cycle), a great deal of the proof goes through. Certainly the forward and backward derivatives give rise to two finite sequences R_i and S_i of relators for G_1, but in order to prove that these relations are sufficient to define G_1 we would need to prove a Freiheitssatz for the one-relator products $(G_0 * \langle X \rangle)/S_N$ and $(G_0 * \langle Y \rangle)/R_M$. In our case, we have used the combinatorics of the diameter 3 situation in a nontrivial way to show that G_0 is free and that S_N properly involves X (resp R_M properly involves Y) modulo the relations of G_0, from which the Freiheitssatz follows.

It seems reasonable to conjecture in more generality that the HNN base B for G, generated by $\{xy^{-1},\ x, y \in V\}$ will be finitely presented. One may construct sets of relations on this generating set analogous to the R_i and S_i above, by repeatedly applying the forward derivative construction to nullhomotopic paths arising from closed paths in $I(\Gamma)$ (analogous to our S_0), and the backward derivative construction to nullhomotopic paths arising from closed paths in $T(\Gamma)$ (analogous to our R_0). Provided we restrict attention to simple closed paths, only finitely many relations arise in this way, and one can conjecture that these form a set of defining relators for B.

Before making this conjecture precise, let us first give a geometric interpretation of these relations. On the 2–complex $K = K(\Gamma)$ we define a *track* $\mathbf{T}$ in the sense of Dunwoody [4] as follows: $\mathbf{T}$ intersects each 1–cell in a single point, and each 2–cell in two arcs as in the diagram below.

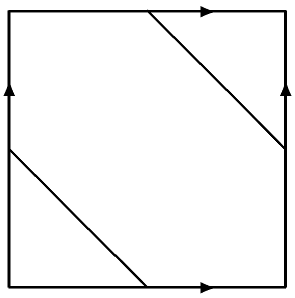


Figure 2

The initial graph $I(\Gamma)$ is naturally embedded as a subgraph of the link of the 0–cell in K. Corresponding to a cycle

$$C = (x_1, \ldots, x_n)$$

in $I(\Gamma)$ is a Dehn diagram D_1 over $\mathcal{P}(\Gamma)$ with a single interior vertex (whose link maps isomorphically to C). We also have a nullhomotopic closed path

$$S_0 = (x_1^{-1}x_2)\ldots(x_n^{-1}x_1)$$

in $K^{(1)}$. The boundary label of D_1 is $S_1 = \partial_+ S_0$. Moreover, if we regard D_1 as a map from the disc D^2 to K, then the track $\mathbf{T}$ on K induces a track on D^2. This track consists of a single circle in the interior of D^2, together with a collection of arcs, each connecting two adjacent track points on ∂D^2.

Now suppose that S_1 lifts to $\hat{L}$. Then the Dehn diagram D_1 can be extended to a diagram D_2 with boundary label $S_2 = \partial_+ S_1$, and so on. On any Dehn diagram arising in this way, the track induced by $\mathbf{T}$ consists of a collection of concentric circles in the interior of D^2, together with a collection of arcs, each connecting two adjacent track points on ∂D^2.

Dual to the track $\mathbf{T}$ is a flow on K, indicated on the boundary of the 2–cells by the arrows in Figure 2. The flow induced on D^2 by any of the Dehn diagrams obtained as above has only one singular point in the interior of D^2, which is a sink.

We can perform a similar construction for any cycle in $T(\Gamma)$. The boundary label of the resulting Dehn diagram is obtained by repeatedly applying the backward derivative operator to a nullhomotopic closed path in $K^{(1)}$. Again, the induced track on D^2 consists of a collection of concentric circles in the interior of D^2, together with a collection of arcs, each connecting two adjacent track points on ∂D^2. The induced flow has only one singular point in the interior of D^2, which is a source.

Let us define a Dehn diagram to be *tame* if the induced track on D^2 consists of a collection of concentric circles in the interior of D^2, together with a collection of arcs, each connecting two adjacent track points on ∂D^2. This is equivalent to the induced flow having only one singular point in the interior of D^2, which is either a sink or a source. It is not difficult to show that every tame Dehn diagram arises by the above construction from a cycle in $I(\Gamma)$ or $T(\Gamma)$, and that its boundary label is an alternating word in the generators $V(\Gamma)$ of $G(\Gamma)$.

Conjecture 8.1 *Let B be the subgroup of $G(\Gamma)$ generated by the alternating words in $V(\Gamma)$. Then B has a finite presentation in which the defining relators are the boundary labels of tame Dehn diagrams.*

References

[1] **J F Adams**, *A new proof of a theorem of W H Cockcroft*, J. LMS 49 (1955) 482–488

[2] **R Bieri**, **R Strebel**, *Almost finitely presented soluble groups*, Comment. Math. Helv. 65 (1990) 243–254

[3] **S D Brodskiĭ**, *Equations over groups and groups with a single defining relator*, Siberian Mathematical Journal 25 (1984) 231–251

[4] **M J Dunwoody**, *The accessibility of finitely presented groups*, Invent. Math. 81 (1985) 449–457

[5] **J Hillman**, *2-Knots and their Groups*, Austral. Math. Soc. Lecture Series 5, Cambridge University Press (1989)

[6] **J Howie**, *On locally indicable groups*, Math. Z. 180 (1982) 445–461

[7] **J Howie**, *The asphericity of ribbon disc complements*, Trans. AMS 289 (1985) 281–302

[8] **S V Ivanov**, *On asphericity of group presentations given by labelled oriented trees*, preprint (1996)

[9] **T Maeda**, *Knotted surfaces in the 4-sphere with no minimal Seifert manifolds*, from: "Combinatorial and Geometric Group Theory" (A J Duncan, N D Gilbert and J Howie, editors), LMS Lecture Note Series 204, Cambridge University Press (1994) 239–246

[10] **W Magnus**, *Über diskontinuierliche Gruppen mit einer definierenden Relation (der Freiheitssatz)*, J. reine angew. Math. 163 (1930) 141–165

[11] **D I Moldavanskiĭ**, *Certain subgroups of groups with one defining relation* (Russian), Sibirsk. Mat. Zh. 8 (1967) 1370–1384

[12] **E S Rapaport**, *Knot-like groups*, from: "Knots, Groups and 3-Manifolds" (L P Neuwirth, editor), Annals of Mathematics Studies, 84 Princeton University Press (1975) 119–133

[13] **D Silver**, *Examples of 3-knots with no minimal Seifert manifolds*, Math. Proc. Cambridge Phil. Soc. 110 (1991) 417–420

[14] **D Silver**, *On the existence of minimal Seifert manifolds*, Math. Proc. Cambridge Phil. Soc. 114 (1993) 103–109

[15] **D Silver**, *HNN bases and high-dimensional knots*, Proc. Amer. Math. Soc. 124 (1996) 1247–1252

[16] **D Silver**, *Free group automorphisms and knotted tori in S^4*, J. Knot Theory Ramifications (1997) 95–103

Department of Mathematics, Heriot-Watt University
Edinburgh EH14 4AS, Scotland

Email: `J.Howie@hw.ac.uk`

Received: 2 June 1997