The generalised word problem in hyperbolic and relatively hyperbolic groups

Laura Ciobanu, Derek Holt, and Sarah Rees

October 6, 2016

Abstract

We prove that, for a finitely generated group hyperbolic relative to virtually abelian subgroups, the generalised word problem for a parabolic subgroup is the language of a real-time Turing machine. Then, for a hyperbolic group, we show that the generalised word problem for a quasiconvex subgroup is a real-time language under either of two additional hypotheses on the subgroup.

By extending the Muller-Schupp theorem we show that the generalised word problem for a finitely generated subgroup of a finitely generated virtually free group is context-free. Conversely, we prove that a hyperbolic group must be virtually free if it has a torsion-free quasiconvex subgroup of infinite index with context-free generalised word problem.

2010 Mathematics Subject Classification: 20F10, 20F67, 68Q45
Key words: generalised word problem, relatively hyperbolic group, context-free language, real-time Turing machine

1 Introduction

Let $G = \langle X \rangle$ with $|X| < \infty$ be a group and $H \leq G$. The *word problem* $\mathsf{WP}(G, X)$ and *generalised word problem* $\mathsf{GWP}(G, H, X)$ are defined to be the preimages $\phi^{-1}(\{1_G\})$ and $\phi^{-1}(H)$ respectively, where ϕ is the natural map from the set of words over X to G. We are interested in the relationship between the algebraic properties of G (and H) and the formal language classes containing $\mathsf{WP}(G, X)$ and $\mathsf{GWP}(G, H, X)$. These questions have already been well studied for the word problem, but relatively little for the generalised word problem. Since, as is well known for $\mathsf{WP}(G, X)$, the question of membership of $\mathsf{WP}(G, X)$ and $\mathsf{GWP}(G, H, X)$ in a formal language family $\mathcal{F}$ is typically independent of the choice of the finite generating set X, we shall usually use the simpler notations $\mathsf{WP}(G)$ and $\mathsf{GWP}(G, H)$. It will be convenient to assume throughout the paper that all generating sets X of groups G are closed under inversion; i.e. $x \in X \Rightarrow x^{-1} \in X$.

It is elementary to prove that $\mathsf{WP}(G)$ is regular if and only if G is finite and, more generally, that $\mathsf{GWP}(G, H)$ is regular if and only if $|G : H|$ is finite.

It is well-known that $\mathsf{WP}(G)$ is context-free if and only if G is virtually free [14], and it is shown in [10, 11] that $\mathsf{WP}(G)$ is a *real-time* language (that is, the language of a real-time Turing machine) for several interesting classes of groups, including hyperbolic and geometrically finite hyperbolic groups. The solvability of $\mathsf{GWP}(G, H)$ has been established for numerous classes of groups, most recently for all (compact and connected) 3-manifold groups [7].

In this paper, we study the conditions under which $\mathsf{GWP}(G, H)$ is a real-time or a context-free language for subgroups H of hyperbolic and relatively hyperbolic groups. There are several definitions of relatively hyperbolic groups, and the one that we are using here is that of [15]; so in particular the Bounded Coset Penetration Property holds. Our first result is the following.

Theorem 1.1. *Suppose that the finitely generated group G is hyperbolic relative to a set $\{H_i : i \in I\}$ of virtually abelian (parabolic) subgroups of G, and that H is a selected parabolic subgroup. Then $\mathsf{GWP}(G, H)$ is a real-time language.*

The proof uses a combination of results for relatively hyperbolic groups that were developed by Antolín and Ciobanu in [3] and the *extended Dehn algorithms* that were introduced by Goodman and Shapiro in [9]. By choosing $H = H_0$ to be the trivial subgroup of G, we obtain a generalisation of the result proved (also using extended Dehn algorithms) in [11] that $\mathsf{WP}(G)$ is a real-time language when G is a geometrically finite hyperbolic group.

Since Goodman and Shapiro's techniques, such as the N-tight Cannon's algorithm and related results (Theorem 37 in [9]), only apply to virtually abelian groups and not, for example, to all virtually nilpotent groups, our proof cannot be extended to parabolics beyond those that are virtually abelian. Furthermore, even with a different approach to the proof there must be some limitations on the choice of parabolics, since examples exist of relatively hyperbolic groups with generalised word problems that are not real-time. If one considers, for example, the free product $G = H * K$ of two groups H and K, where H is hyperbolic and K has unsolvable word problem, then G is hyperbolic relative to K, but it is easy to see that the subgroup membership problem for K in G is unsolvable, so it cannot be real-time.

Recall that a subgroup $H \leq G$ is called *quasiconvex* in G if geodesic words over X that represent elements of H lie within a bounded distance of H in the Cayley graph $\Gamma(G, X)$. And H is called *almost malnormal* in G if $|H \cap H^g|$ is finite for all $g \in G \setminus H$. We conjecture that, for a hyperbolic group G, the set $\mathsf{GWP}(G, H)$ is a real-time language for any quasiconvex subgroup H of G, but we are currently only able to prove this under either of two additional hypotheses:

Theorem 1.2. *Let G be a hyperbolic group, and H a quasiconvex subgroup of G. Suppose that either*

(i) H is almost malnormal in G; or

(ii) $|C_G(h) : C_H(h)|$ is finite for all $1 \neq h \in H$.

Then $\mathsf{GWP}(G, H)$ is a real-time language.

Our proofs of the result in the two cases are quite distinct, and so we write them separately. Under assumption (i), G is hyperbolic relative to $\{H\}$ [15, Section 1, Example (III)], and the proof is similar to that of Theorem 1.1 (although H is not usually virtually abelian, so we cannot apply that result directly). Under assumption (ii), we make use of some results of Foord [6] about the Schreier graph of G with respect to H. Note that, since centralisers of elements of infinite order in hyperbolic groups are virtually cyclic, $|C_G(h) : C_H(h)|$ is always finite for such elements h and so, in particular, assumption (ii) holds whenever G is torsion-free.

Our next result, which concerns context-free generalised word problems, is straightforward to prove and may be known already, but it does not appear to be in the literature.

Theorem 1.3. *Let G be finitely generated and virtually free, and let H be a finitely generated subgroup of G. Then $\mathsf{GWP}(G, H)$ is deterministic context-free.*

The conclusion of Theorem 1.3 may or may not hold if we drop the condition that H is finitely generated. Suppose that G is the free group on two generators a, b. For $H_1 = [G, G]$, the set $\mathsf{GWP}(G, H_1)$ consists of all words whose exponent sums in both a and b are zero, and is not context-free, but for the subgroup H_2 of words whose exponent sum in a is zero, the set $\mathsf{GWP}(G, H_2)$ is context-free.

We would like to know to what extent Theorem 1.3 is best possible when H is finitely generated. We observe that, where $H_G := \cap_{g \in G} H^g$ is the *core* of H in G, the set $\mathsf{GWP}(G, H, X)$ is the same set of words as $\mathsf{GWP}(G/H_G, H/H_G, X)$. We know of no examples for which H is finitely generated with trivial core and $\mathsf{GWP}(G, H)$ is context-free, but G is not virtually free. The following result is an attempt at a converse to Theorem 1.3.

Theorem 1.4. *Let G be a hyperbolic group, and let H be a quasiconvex subgroup of infinite index in G such that $|C_G(h) : C_H(h)|$ is finite for all $1 \neq h \in H$. If $\mathsf{GWP}(G, H)$ is context-free then G is virtually free.*

Note that we make the same assumption on centralisers of elements $h \in H$ as in Theorem 1.2 (ii), and again we conjecture that this is not necessary for the conclusion of the theorem. However the necessity of quasiconvexity is demonstrated by a construction found in [16], as follows. Let Q be any finitely presented group, and choose $\lambda > 0$. We can define a finitely presented group G and a normal 2-generated subgroup H of G, such that $G/H \cong Q$, and G satisfies the small cancellation condition $C'(\lambda)$; in particular we can choose Q with insoluble word problem, and choose $\lambda \leq 1/6$ to ensure that G is hyperbolic, and in that case $\mathsf{GWP}(G, H)$ is not even recursive.

The closure properties of context-free languages [12, Chapter 11] and of real-time languages [17] ensure that all of the above results are independent of the choice of the finite generating set X of G, so we are free to choose X to suit our own purposes in the proofs. (More generally, for membership of $\mathsf{WP}(G, X)$ or $\mathsf{GWP}(G, H, X)$ in a formal language class $\mathcal{F}$ to be independent of the choice of X, we need $\mathcal{F}$ to be closed under inverse homomorphism. This property holds for all of the most familiar formal language classes, including regular, context-free, deterministic context-free, real-time, context-sensitive, deterministic context-sensitive, and recursive languages.)

This article is structured as follows. In Section 2, we summarise the basic properties of relatively hyperbolic groups that we shall need, and we recall some of their properties that are proved in [3]. In Section 3, we introduce the concept of *extended Dehn algorithms* for solving the word and generalised word problems in groups and we recall some results pertaining to relatively hyperbolic groups that are proved in [9]. Sections 4, 5, 6, 7 and 8 contain the proofs of Theorems 1.1, 1.2 (i), 1.2 (ii), 1.3 and 1.4, respectively. Finally, in Section 9, we sketch a proof of a result of Foord [6, Theorem 4.3.1.1] that we shall need, since the original source might not be readily available to readers.

2 Relatively hyperbolic groups

We follow [15] for notation and the definition of relatively hyperbolic groups. This definition is equivalent to what Farb calls "strong relative hyperbolicity" in [5].

Suppose that G is a group, X a finite generating set, and $\{H_i : i \in I\}$ a collection of subgroups of G, which we call *parabolic* subgroups. We define $X_i := X \cap H_i$ to be the set of generators in X that lie within the subgroup H_i, and $X_I := \cup_{i\in I} X_i$. Then we define $\mathcal{H}$ and $\widehat{X}$ as the sets

$$\mathcal{H} := \bigcup_{i \in I} (H_i \setminus \{1\}), \quad \widehat{X} = X \cup \mathcal{H}.$$

Much of our argument involves the comparison of lengths of various words that represent an element $g \in G$ written over different sets, namely X, $\widehat{X}$, a set $Z \supseteq X$ that is introduced during the extended Dehn algorithm (which, when $Z \supset X$, is not actually a generating set for G, since only some of the words over Z correspond to elements of G), and certain subsets of these sets.

So we shall consider the Cayley graphs Γ and $\widehat{\Gamma}$ for G over the generating sets X and $\widehat{X}$, and view words over X and $\widehat{X}$ also as paths in Γ and $\widehat{\Gamma}$. We denote by d_Γ the graph distance in Γ and by $d_{\widehat{\Gamma}}$ the graph distance in $\widehat{\Gamma}$. For a word w written over a set Y (that is, an element of Y^*), we write $|w|$ to denote the length of w and, for a group element g, we write $|g|_Y$ to denote the length of a shortest word over Y that represents the group element g (assuming that such a word exists). We call $|g|_Y$ the *Y-length* of g, and a shortest word over Y that represents g a *Y-geodesic* for g.

Most words will be written over either X or $\widehat{X}$ and, in order to make a clear distinction between those two types of words we will normally use Roman letters as names for words over X and paths in Γ, and Greek letters as names for words over $\widehat{X}$ and paths in $\widehat{\Gamma}$, with the exception that we will write $\widehat{w}$ for the word over $\widehat{X}$ that is derived from a word w over X by a process called *compression*, which will be described later in this section.

We refer to [15] for a precise definition of relative hyperbolicity of G with respect to $\{H_i : i \in I\}$. Under that definition, relative hyperbolicity is known to be equivalent to the fact that the Cayley graph $\widehat{\Gamma}$ is δ-hyperbolic for some δ together with the Bounded Coset Penetration Property, stated below as Property 2.1.

From now on we shall assume that G is relatively hyperbolic in this sense. It is proved in [15] that, under our assumption that G is finitely generated, the subgroups H_i are finitely generated, the set I is finite, and any two distinct parabolic subgroups have finite intersection. We assume that the generating set X is chosen such that $H_i = \langle X_i \rangle$ for all $i \in I$.

We need some terminology relating to paths in $\widehat{\Gamma}$.

1. We call a subpath of a path π an H_i-*component*, or simply a *component*, of π if it is written as a word over H_i for some $i \in I$, and is not contained in any longer such subpath of π.

2. Two components of (possibly distinct) paths are said to be *connected* if both are H_i-components for some $i \in I$, and both start within the same left coset of H_i.

3. A path is said to *backtrack* if it has a pair of connected components. A path is said to *vertex-backtrack* if it has a subpath of length greater than 1 that is labelled by a word representing an element of some H_i.

   Note that if a path does not vertex-backtrack, then it does not backtrack and all of its components are edges.

4. For $\kappa \geq 0$ we say that two paths are κ-*similar* if the X-distances between their two initial vertices and between their two terminal vertices are both at most κ.

Property 2.1 (Bounded Coset Penetration Property [15, Theorem 3.23]). *For any $\lambda \geq 1, c \geq 0, \kappa \geq 0$, there exists a constant $\epsilon = \epsilon(\lambda, c, \kappa)$ such that, for any two κ-similar (λ, c)-quasi-geodesic paths π and π' in $\widehat{\Gamma}$ that do not backtrack, the following conditions hold:*

(1) *The sets of vertices of π and π' are contained in the closed ϵ-neighbourhoods (with respect to the metric d_Γ) of each other.*

(2) *For any H_i-component σ of π for which the X-distance between its endpoints is greater than ϵ, some H_i-component σ' of π' is connected to σ.*

(3) *Whenever σ and σ' are connected H_i-components of π and π' respectively, the paths σ and σ' are ϵ-similar.*

We shall need to use two more properties of relatively hyperbolic groups that are proved in [3].

Property 2.2 ([3, Theorem 5.2]). *Let Y be a finite generating set for G. Then for some $\lambda \geq 1, c \geq 0$ there exists a finite set Ψ of non-geodesic words over $Y \cup \mathcal{H}$ such that:*

every 2-local geodesic word over $Y \cup \mathcal{H}$ not containing any element of Ψ as a subword labels a (λ, c)-quasi-geodesic path in $\widehat{\Gamma}$ without vertex-backtracking.

Now suppose that v is any word in X^*. Then following [3, Construction 4.1] we define $\widehat{v}$ to be the word over $\widehat{X}$ that is obtained from v by replacing (working from the left) each subword u that is maximal as a subword over some X_i $(i \in I)$ by the element h_u of $\mathcal{H}$ that the subword represents. We call these X_i-subwords u of v its *parabolic segments*, and use the term *compression* for the process that converts v to $\widehat{v}$. A word v is said to have *no parabolic shortenings* if each of its parabolic segments is an X_i-geodesic.

In order to avoid confusion we comment that a parabolic segment (which is a maximum subword over some X_i of a word over X) is not quite the same as a component (which is a maximal subpath/subword over some H_i of a path/word over $\widehat{X}$); but clearly the two concepts are close.

The other required property is proved in [3, Lemma 5.3]; the precise description of Φ is taken from the proof of that lemma, rather than from its statement.

Property 2.3. *Let Y be a finite generating set for G. Then for some $\lambda \geq 1$, $c \geq 0$, some finite subset $\mathcal{H}'$ of $\mathcal{H}$, and any finite generating set X of G with*

$$Y \cup \mathcal{H}' \subseteq X \subseteq Y \cup \mathcal{H},$$

there is a finite subset Φ of non-geodesic words over X such that:

if a word $w \in X^*$ has no parabolic shortenings and no subwords in Φ, then the word $\widehat{w} \in \widehat{X}^*$ is a 2-local geodesic and labels a (λ, c)-quasi-geodesic path in $\widehat{\Gamma}$ without vertex-backtracking.

Furthermore, for every $i \in I$ and $h \in H_i$, we have $|h|_X = |h|_{X_i}$.

In fact $\Phi = \Phi_1 \cup \Phi_2$, where Φ_1 is the set of non-geodesic words in X^* of length 2, and Φ_2 is the set of all words $u \in X^*$ with no parabolic shortening and for which $\widehat{u} \in \Psi$, where Ψ (together with λ and c) is given by Property 2.2.

3 Extended Dehn algorithms

Our proofs of Theorem 1.1 and both parts of Theorem 1.2 depend on the construction of an *extended Dehn algorithm* (eda) [9] for G with respect to H. In each case, we then need to show that the eda satisfies a particular condition that allows us to apply Proposition 3.2 (below) in order to verify both that the algorithm solves $\mathsf{GWP}(G, H)$ and that it can be programmed on a real-time Turing machine. Proposition 3.2 is derived from [11, Theorem 4.1], which was used to prove the solubility of the word problem in real-time for various groups with edas to solve that problem. We restate that result as a proposition in this paper for greater clarity of exposition.

Our definition of an extended Dehn algorithm (which is defined with respect to a specific finite generating set X of G) is modelled on the definition of [9] (where it is called a *Cannon's algorithm*), with the difference that we are using our algorithm to solve a generalised word problem $\mathsf{GWP}(G, H, X)$ rather than a word problem $\mathsf{WP}(G, X)$. Elsewhere in the literature [11] the same concept is called a *generalised Dehn algorithm*; our decision to introduce a new name is based on both our recognition that there are many various different algorithms attributed to (more than one) Cannon, and our desire to avoid overuse of the term 'generalised'.

For a (Noetherian) rewriting system R with alphabet Z and $w \in Z^*$, we write $R(w)$ for the reduction of the word w using the rules of R. In general, $R(w)$ may depend on the order in which the rules are applied, and we shall specify that order shortly. A word w is called *(R-)reduced* if $R(w) = w$; that is, if w does not contain the left hand side of any rule as a subword.

We define an eda for a finitely generated group $G = \langle X \rangle$ with respect to a subgroup $H \leq G$ to be a finite rewriting system S consisting of rules $u \to v$, where

(i) $u, v \in Z^*$ for some finite alphabet $Z \supseteq X \cup \{H\}$;
(ii) $|u| > |v|$; and
(iii) either $u, v \in (Z \setminus \{H\})^*$ or $u = Hu_1, v = Hv_1$, with $u_1, v_1 \in (Z \setminus \{H\})^*$.

We say that the eda S solves the generalised word problem $\mathsf{GWP}(G, H, X)$ if, for every word w over X, we have $S(Hw) = H$ if and only if w represents an element of H. If $H = \{1\}$ (in which case we may assume only that $Z \supseteq X$), then we call S an eda for G.

As observed earlier, for $S(w)$ to be well-defined, we need to specify the order in which the reduction rules are applied to words w. In the terminology of [9, Section 1.2], S with the order we specify below is an *incremental rewriting algorithm* and, since we shall only apply it to words of the form Hw with $w \in X^*$, the rules $Hu_1 \to Hv_1$ are effectively *anchored* rules. We assume that no two distinct rules have the same left hand sides. Then we require that when a word Hw contains several left hand sides of S, the rule which is applied is one that ends closest to the start of Hw; if there are several such rules, the one with the longest left hand side is selected.

In our applications, the rules of the form $u \to v$ with $u, v \in (Z \setminus \{H\})^*$ will form an eda R for G, of a type that is considered in [9], and which solves the word problem $\mathsf{WP}(G, X)$; i.e. $R(w)$ is the empty word if and only if $w =_G 1$. The properties of the eda R that we shall use are described in more detail in Proposition 3.1 below. Many of the technical results of [9] apply without modification to edas that solve a GWP rather than WP.

Note that the set Z may properly contain $X \cup \{H\}$, and so contain symbols that do not correspond to either elements or subsets of G. But it is a consequence of [9, Proposition 3] that a word in $(Z \setminus \{H\})^*$ that arises from applying these rules to a word $w \in X^*$ unambiguously corresponds to the element of G represented by w, and so we may interpret such words as elements in G. We shall see shortly that our rewrite rules $Hu \to Hv$ will all be of the form $Hz \to H$ for words $z \in (Z \setminus \{H\})^*$ that represent elements of H, and so words derived by applying rules of the eda to Hw with $w \in X^*$ unambiguously represent the coset of H in G defined by Hw.

Following [9, Corollary 27], for a positive integer D, we say that an eda S that solves the word problem for a group $K = \langle Y \rangle$ is D-*geodesic* if a word w that is S-reduced and represents an element g of Y-length at most D must in fact be written over Y, and be a Y-geodesic for g.

In this section so far, we have not made any assumptions on G, H, X, beyond the finiteness of the generating set X. Suppose now that the parabolic subgroups H_i are all virtually abelian, and that R is an eda for G that solves $\mathsf{WP}(G, X)$. For integers $D \geq E \geq 0$, we say that R satisfies $\mathcal{P}(D, E)$ if the following conditions hold.

(1) For each i, $X_i := X \cap H_i$ generates H_i, and the alphabet Z of R has the form $Z = \cup_{i \in I} Z_i \cup X$ with $X_i \subseteq Z_i$.
(2) For each rule $u \to v$ of R, we either have $u, v \in X^*$, or $u, v \in Z_i^*$ for a unique $i \in I$.
(3) For each $i \in I$, the rules $u \to v$ with $u, v \in Z_i^*$ form a D-geodesic eda R_i that solves $\mathsf{WP}(H_i, X_i)$.
(4) All R-reduced words $w \in X^*$ that have length at most E are X-geodesic.

The following result is proved under slightly more general conditions on the parabolic subgroups in [9], and is stated here in the form in which we need it:

Proposition 3.1 ([9, Theorem 37])**.** *Suppose that $G = \langle Y \rangle$ is hyperbolic relative to the virtually abelian parabolic subgroups H_i. Then, there is a finite generating set X of G, consisting of the generators in Y together with some additional elements from the H_i (that include all non-trivial elements from the intersections $H_i \cap H_j$ with $i \neq j$) with the following property: for all sufficiently large integers D, E with $D \geq E \geq 0$, there is an eda for G that solves $\mathsf{WP}(G, X)$ and satisfies $\mathcal{P}(D, E)$.*

We need to extend our definition of the property $\mathcal{P}(D, E)$ to our wider definition of an eda for a group with respect to a subgroup. For G satisfying the above hypotheses, and given any subgroup $H \leq G$, we shall say that an eda S for G with respect to H, with alphabet $Z \cup \{H\}$, satisfies $\mathcal{P}(D, E)$ if the rules in S of the form $u \to v$ with $u, v \in Z^*$ form an eda R satisfying $\mathcal{P}(D, E)$.

In the proofs of each of Theorems 1.1, 1.2 (i), 1.2 (ii), we shall apply the following result, which is essentially (part of) [11, Theorem 4.1].

Proposition 3.2. *Let G be a group, finitely generated over X, H a subgroup of G, and let S be an extended Dehn algorithm for G with respect to H. Suppose that there exists a constant k such that, for any word w over X, we have*

$$|w_1| \leq k \min\{|g|_X : g \in G,\, g \in Hw\},$$

where w_1 is the word over Z defined by $Hw_1 = S(Hw)$. Then S solves the generalised word problem and can be programmed on a real-time Turing machine.

Proof. Since, for any $w \in \mathsf{GWP}(G, H)$ the minimal length representative of Hw is the identity element, it is immediate from the inequality that S solves $\mathsf{GWP}(G, H)$. That an eda satisfying that condition can be programmed in real-time is then an immediate consequence of [11, Theorem 4.1]; in fact as stated that theorem applies only to edas to solve the word problem, but it is clear from the proof that it applies also to $\mathsf{GWP}(G, H)$. $\square$

4 The proof of Theorem 1.1

Suppose that $G = \langle X \rangle$ satisfies the hypotheses of Theorem 1.1, and that $H = H_0$ is the selected parabolic subgroup, generated by $X_0 \subset X$.

We start with some adjustments to X that are necessary to ensure that it satisfies the conditions we need for our arguments. These adjustments all consist of appending generators that lie in one of the parabolic subgroups. Firstly, we extend X to contain the finite subset $\mathcal{H}'$ of $\mathcal{H}$ defined in Property 2.3. Secondly, we adjoin to X the elements of the H_i that are required by Proposition 3.1. (As stated in Proposition 3.1, these include all elements in the finite intersections $H_i \cap H_j$ (for $i \neq j$) of pairs of parabolic subgroups.)

Associated with this choice of X, Properties 2.2 and 2.3 specify sets Ψ and Φ of non-geodesic words over $\widehat{X}$ and X respectively, and associated parameters λ, c. We then define $\epsilon = \epsilon(\lambda, c+1, 0)$ to be the constant in the conclusion of Property 2.1.

We now apply Proposition 3.1 to find an eda R for G that solves $WP(G, X)$ and that satisfies $\mathcal{P}(D, E)$ for parameters D, E with $D \geq E$, where $E > \max\{\epsilon, 2\}$, and E is also greater than the length of any word in Φ, and greater than the X-length of any component of any word in Ψ. (The reasons for these conditions will become clear during the proof.)

We can use all properties of R that are proved in [9, Section 5] and we observe in particular that, by [9, Lemma 48], for a word $w \in X^*$, if $R(w) = u_1 v u_2$ where v is a maximal Z_i-subword for some $i \in I$, then there exists $v' \in X_i^*$ with $R_i(v') = v$, where R_i is the associated eda for H_i. So v unambiguously represents the element $v' \in H_i$.

We create an eda S for G with respect to H by adding to R all rules of the form $Hz \to H$ with $z \in Z_0$. Since none of these new rules actually applies to words over X, it is clear that the eda S also satisfies $\mathcal{P}(D, E)$. We shall verify that S solves $\mathsf{GWP}(G, H, X)$, and that it can be programmed on a real-time Turing machine.

Now suppose that $w \in X^*$ and that $S(Hw) = Hw_1$. In order to verify that our eda S solves $\mathsf{GWP}(G, H, X)$ and can be programmed on a real-time Turing machine, it is sufficient by Proposition 3.2 to establish the existence of a constant k that is independent of the choice of w, such that

$$|w_1| \leq k \min\{|g|_X : g \in Hw\}. \quad (\dagger)$$

So the aim of the rest of the proof is to prove the inequality (†).

Observe that a maximal subword p_i of w_1 that is written over Z_i^* for some i may contain symbols from $Z_i \setminus X_i$, but any such symbols must have arisen from application of the rules in the eda R_i to words over X_i^*, and so p_i unambiguously represents an element of H_i. Following [9], we decompose w_1 as a concatenation

$$w_1 = v_0 p_1 v_1 \cdots p_m v_m, \quad (*)$$

where p_1 is defined to be the first subword of w_1 (working from the left) that is written over Z_{i_1} for some $i_1 \in I$, has maximal length as such a subword, and represents an element of H_{i_1} of X_{i_1}-length greater than E. (Note that the words v_i are denoted by g_i in [9].) The subwords p_i for $i > 1$ are defined correspondingly with respect to the suffix remaining after removing the prefix $v_0 p_1 v_1 \cdots p_{i-1}$ from w_1.

Then any (maximal) subword of any v_j that is written over any Z_i represents an element of H_i of X_i-length at most $E \leq D$, and so $\mathcal{P}(D,E)\,(3)$ ensures that the subword is written over X_i and is an X_i-geodesic. Hence v_j is a word over X with no parabolic shortenings. Then (since $E > 2$) property $\mathcal{P}(D,E)\,(4)$ ensures that v_j is also a 2-local geodesic. Also, since Φ is a set of non-geodesic words over X of length at most E, v_j cannot contain any subword in Φ. It follows by Property 2.3 that each $\widehat{v_j}$ is a 2-local geodesic.

Now, for each j, choose q_j to be a geodesic word over X_{i_j} that represents the same element h_j of H_{i_j} as p_j (so, by Property 2.3, q_j is also an X-geodesic). Define

$$w_2 = v_0 q_1 v_1 \cdots q_m v_m.$$

We already observed that each v_j is written over X, and hence so is w_2. Now according to [9, Lemma 23], the X-lengths of non-identity elements of H_i are bounded below by an exponential function on the lengths of words over Z_i that are their reductions by the eda for H_i. So there is certainly a positive constant k_1 such that $|q_j| \geq k_1 |p_{i_j}|$ for all j, and hence $|w_2| \geq k_1 |w_1|$. Hence it is sufficient to prove the inequality (†) above for the word w_2 rather than w_1, that is, for some k', show that

$$|w_2| \leq k' \min\{|g|_X : g \in Hw\}. \quad (\dagger\dagger)$$

With this in mind, our next step is to construct a word $\widetilde{w_2}$ over $\widehat{X}$, representing the same element of G as w_2, and for which we can use Property 2.2. We define

$$\widetilde{w_2} = \widehat{v_0} h_1 \widehat{v_1} \cdots h_m \widehat{v_m}.$$

Note that we would have $\widetilde{w_2} = \widehat{w_2}$ if the subwords q_j were parabolic segments of w_2, but this might not be true if the first generator in some q_j were in more than one parabolic subgroup, and then we could not be sure that $\widehat{w_2}$ would satisfy the required conditions. We call the process of conversion of w_2 to $\widetilde{w_2}$ *modified compression* and we call the subwords q_j of w_2 together with the parabolic segments of the subwords v_j the *modified parabolic segments* of w_2. Observe that these modified parabolic segments are all X-geodesics.

We want to apply Property 2.2 to $\widetilde{w_2}$, so we must show first that $\widetilde{w_2}$ is a 2-local geodesic over $\widehat{X}$. If not, then $\widetilde{w_2}$ has a non-geodesic subword ζ of length 2, equal in G to an $\widehat{X}$-geodesic word η of length at most 1. We saw earlier that the subwords $\widehat{v_j}$ are 2-local geodesics, so ζ must contain some h_j; that is, $\zeta = yh_j$ or $\zeta = h_j y$, where $y \in \widehat{X}$. Then the definition of the p_j as maximal Z_{i_j}-subwords of w_1 ensures that $y \notin H_{i_j}$, so $|\eta| = 1$. So, since $\lambda, c+1 \geq 1$, ζ and η are both $(\lambda, c+1)$-quasigeodesics. But now, since h_j is a component in ζ of X-length greater than $E > \epsilon = \epsilon(\lambda, c+1, 0)$, we can apply Property 2.1 to the paths in $\widehat{\Gamma}$ labelled by ζ and η, and deduce that η contains a component connected to the component h_j, and so η represents an element of H_{i_j}; hence $\zeta \in H_{i_j}$, and we have a contradiction. (Alternatively, we could apply [3, Lemma 4.2] to deduce that $|\zeta| = 2$, a contradiction.)

To verify the second requirement of Property 2.2, we need to check that $\widetilde{w_2}$ contains no subword in Ψ. So suppose that ξ is such a subword in Ψ. Then ξ cannot contain any of the generators h_j, since h_j would then be a component in ξ of X-length greater than E, by the conditions imposed on the decomposition $(*)$ of w_1; but this contradicts the choice of E earlier in this proof to be greater than the X-length of any component of any word in Ψ. So ξ must be a subword of some $\widehat{v_j}$; but in that case, $\xi = \widehat{u}$ for some subword u of v_j. We saw earlier that v_j has no parabolic shortenings, and hence neither does u. So from the definition of Φ_2 we have $u \in \Phi_2 \subseteq \Phi$. But we also observed earlier that v_j has no subword in Φ, so we have a contradiction.

It now follows using Property 2.2 that $\widetilde{w_2}$ is a (λ, c)-quasigeodesic over $\widehat{X}$ without vertex-backtracking.

Let w_3 be a geodesic over X that represents an element of $Hw_2 = Hw$; since it is geodesic, w_3 cannot contain any subwords in Φ, and it cannot have any parabolic shortenings. So we can apply Property 2.3 to deduce that $\widehat{w_3}$ is a 2-local geodesic over $\widehat{X}$ and a (λ, c)-quasigeodesic without vertex-backtracking.

Now $\widetilde{w_2} =_G h\widehat{w_3}$ for some $h \in H$ and so, since $\widetilde{w_2}$ is a (λ, c)-quasigeodesic,

$$|\widetilde{w_2}| \leq \lambda |h\widehat{w_3}| + c.$$

We note that $\widetilde{w_2}$ and $h\widehat{w_3}$ are both $(\lambda, c+1)$-quasigeodesics over $\widehat{X}$ without vertex-backtracking, and the initial and terminal vertices of the paths in $\widehat{\Gamma}$ that they label coincide. So we have Properties 2.1 (2) and (3) concerning the components of the two paths. (This is why we chose $\epsilon = \epsilon(\lambda, c+1, 0)$ at the beginning of the proof.)

We choose a geodesic word w_h over X_0 that represents h, and consider the words w_2 and $w_h w_3$. We want to compare the lengths of w_2 and $w_h w_3$, and we do this by examining the processes of (modified) compression of w_2 and $w_h w_3$ to $\widetilde{w_2}$ and $\widehat{w_h w_3} = h\widehat{w_3}$. The word w_2 can be decomposed as a concatenation of disjoint subwords that are its *long* modified parabolic segments, its *short* modified parabolic segments and its maximal subwords over $X \setminus X_I$; we define a modified parabolic segment to be *long* if its length is greater than 4ϵ and *short* otherwise.

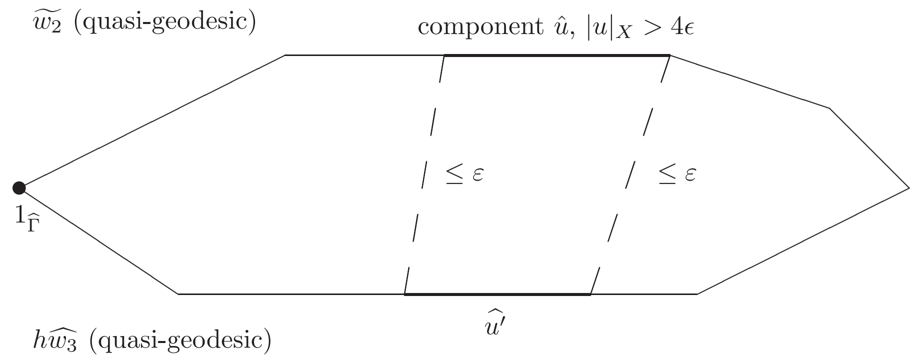

Figure 1: The paths $\widetilde{w_2}$ and $h\widehat{w_3}$ in $\widehat{\Gamma}$: bounded coset penetration

Now modified compression reduces the total length of subwords of the second type, which are replaced by single elements of $\widehat{X}$, by a factor of at most 4ϵ, while the subwords of the third type are unchanged. So the total length of the subwords of w_2 of the second and third types is bounded by

$$4\epsilon|\widetilde{w_2}| \leq 4\epsilon(\lambda|h\widehat{w_3}| + c) \leq 4\epsilon(\lambda|w_h w_3| + c),$$

where the second equality follows from the fact that $h\widehat{w_3} = \widehat{w_h w_3}$.

Now let u be a long modified parabolic segment of w_2. Then Property 2.1 applied to $\widetilde{w_2}$ and $h\widehat{w_3}$ ensures that there is a corresponding parabolic segment u' of $w_h w_3$ such that $\widehat{u}$ and $\widehat{u'}$ are connected components of length 1 of $\widetilde{w_2}$ and $h\widehat{w_3}$ Then since $\widehat{u}$ and $\widehat{u'}$ must be ϵ-similar, it follows that the initial and terminal points of u' must be within X-distance ϵ of the initial and terminal points (respectively) of u. Then $4\epsilon < |u| \leq 2\epsilon + |u'|$, and so $|u| \leq 2|u'|$ (see Fig. 1).

Since $\widetilde{w_2}$ does not backtrack, distinct components of $\widetilde{w_2}$ must correspond to distinct components of $h\widehat{w_3}$, and we deduce that the total length of the long parabolic segments in w_2 is bounded above by $2|w_h w_3|$. So

$$|w_2| \leq (4\epsilon\lambda + 2)|w_h w_3| + 4\epsilon c.$$

Now we consider w_h, which is a parabolic segment of $w_h w_3$. If $|w_h| > \epsilon$ then Property 2.1 applied to (the paths labelled by) $h\widehat{w_3}$ and $\widetilde{w_2}$ ensures the existence of a corresponding parabolic segment u_2 in w_2, whose initial vertex is within X-distance ϵ of the basepoint $1_{\widehat{\Gamma}}$ of the Cayley graph $\widehat{\Gamma}$, and whose terminal vertex must be in H. Let w_2 factorise as a concatenation of subwords $u_1 u_2 u_3$. Since $u_1 u_2$ is a prefix of w_2 that represents an element of H, the fact that $\widetilde{w_2}$ does not vertex-backtrack ensures that its subword $\widetilde{u_1 u_2}$ must have length at most 1, and hence u_1 is empty. But now the prefix u_2 of w_2 is written over the generators of H, and w_2 cannot have a non-trivial such prefix, since w_1 (from which it was derived) was reduced by the eda S, and we have a contradiction.

So now $|w_h| \leq \epsilon$, and we can deduce from the inequality above that

$$|w_2| \leq A|w_3| + B$$

for some constants A, B. Provided that $|w_3| \neq 0$, it follows that

$$|w_2| \leq (A+B)|w_3|.$$

But if $|w_3| = 0$, then w_2 must represent an element of H and so, since $\widetilde{w_2}$ has already been proved not to vertex-backtrack, $\widetilde{w_2}$ must be a word written over H of length at most 1. It follows that w_2 is a word over X_0, and so, just as above, we deduce that $|w_2| = 0$, and so the same inequality holds. This completes our verification of the condition of (††) (and hence (†)), and the theorem is proved.

5 The proof of Theorem 1.2 (i)

Let H be a quasiconvex and almost malnormal subgroup of a hyperbolic group $G = \langle X \rangle$. It is observed in [15, Section 1, Example (III)] that G is hyperbolic relative to $\{H\}$ so we can apply the results of Section 2. The proof of Theorem 1.2 (i) is very similar to that of Theorem 1.1 (although we are no longer assuming that H is virtually abelian) but is more straightforward, so we shall only summarise it here. In particular, we have $Z = X$ so the complications arising from the elements of $Z \setminus X$ that do not necessarily represent group elements do not arise.

We start by extending X as before to include the finite subset $\mathcal{H}'$ of $\mathcal{H}$ defined in Property 2.3. Since we are not applying Proposition 3.1 in this proof, the other adjustment to X is not necessary. We define $\Psi, \Phi, \lambda, c, \epsilon, E$ as before and put $D = E$.

The standard Dehn algorithm for solving $\mathsf{WP}(G, X)$ consists of all rules $u \to v$ with $u, v \in X^*$ such that $u =_G v$ and $4\delta \geq |u| > |v|$, where δ is the 'thinness' constant of G with respect to X (i.e. all geodesic triangles in Γ are δ-thin); see [1, Theorem 2.12]. Words w that are reduced by this algorithm are 4δ-local geodesics, and it is proved in [10, Proposition 2.1] that, if w represents the group element g, then $|w| \leq 2|g|$.

We define our Dehn algorithm R for $\mathsf{WP}(G, X)$ to consist of all rules $u \to v$ as above, with $k \geq |u| > |v|$, where $k = \max(2D, 4\delta)$. Then R-reduced words have the property that subwords representing group elements of X-length at most D are X-geodesics. Since $D = E$, it is also true that R-reduced words of length at most E are geodesic, so R has the required property $\mathcal{P}(D, E)$. As in the previous proof, we define S to be the eda for $\mathsf{GWP}(G, H, X)$ consisting of R together with rules $Hx \to H$ for all $x \in X_0 := X \cap H$.

As before, we suppose that S reduces the input word Hw to Hw_1 and define the decomposition $(*)$ of w_1 with p_i being maximal X_0-subwords of w_1 that represent group elements of X-length greater than E. Again we let q_i be geodesic words over X_0 (and hence also over X) with $q_i =_G p_i$. By [10, Proposition 2.1], we have $|q_i| \geq k_1|p_i|$ with $k_1 = 1/2$, so again we have $|w_2| \geq k_1|w_1|$. The remainder of the proof is identical to that of Theorem 1.1.

6 Proof of Theorem 1.2 (ii)

As we did for Part (i) of this theorem, we prove Theorem 1.2 (ii) by constructing an eda over the alphabet $X \cup \{H\}$, where $G = \langle X \rangle$. As in the two earlier proofs, we verify that the conditions of Proposition 3.2 hold, to complete the proof.

For a finite (inverse-closed) generating set X of an arbitrary group G, we define an *X-graph* to be a graph with directed edges labelled by elements of X, in which, for each vertex p and each $x \in X$, there is a single edge labelled x with source p and, if this edge has target q, then there is an edge labelled x^{-1} from q to p. So the *Cayley graph* $\Gamma(G, X)$ and, for a subgroup $H \leq G$, the *Schreier graph* $\Sigma(G, H, X)$ of G with respect to H are examples of X-graphs. We shall denote the base points of the Cayley and Schreier graphs by 1_Γ and 1_Σ respectively.

Following [6, Chapter 4], for $k \in \mathbb{N}$, we define the condition GIB(k) (which stands for *group isomorphic balls*) for $\Sigma := \Sigma(G, H, X)$ as follows.

> GIB(k): there exists $K \in \mathbb{N}$ such that, for any vertex p of Σ with $d(1_\Sigma, p) \geq K$, the closed k-ball $B_k(p)$ of Σ is X-graph isomorphic to the k-ball $B_k(1_\Gamma)$ of $\Gamma(G, X)$.

We say that Σ satisfies GIB(∞) if it satisfies GIB(k) for all $k \geq 0$. The following result is proved in [6, Theorem 4.3.1.1]. Since its proof may not be readily available, we shall sketch it in Section 9.

Proposition 6.1. *Let H be a quasiconvex subgroup of the hyperbolic group G. Then $\Sigma(G, H, X)$ satisfies* GIB(∞) *if and only if, for all $1 \neq h \in H$, the index $|C_G(h) : C_H(h)|$ is finite.*

Suppose that G, H satisfy the hypotheses of Theorem 1.2 (ii). So $\Sigma := \Sigma(G, H, X)$ satisfies GIB(∞) and, by [6, Theorem 4.1.3.3] or [13], Σ is δ-hyperbolic for some $\delta > 0$ (that is, geodesic triangles in Σ are δ-thin).

Let k be an integer with $k \geq 4\delta$. Let K be an integer that satisfies the condition in the definition of GIB(k), and let $R = 2K$. We can assume that $K \geq \max(k, 2)$. We define our eda to consist of all rules of the following two forms:

$$Hv_1 \to Hv_2, \qquad |v_2| < |v_1| \leq R, \quad (1)$$
$$u_1 \to u_2, \qquad |u_2| < |u_1| \leq k, \quad (2)$$

where $v_1, v_2, u_1, u_2 \in X^*, v_1 v_2^{-1} \in H, u_1 =_G u_2$.

In order to apply Proposition 3.2 we need to verify that, whenever $w \in X^*$ and Hw is reduced according to the above eda, the length of the shortest string v over X with $Hv = Hw$ is bounded below by a linear function of $|w|$.

We shall use [10, Proposition 2.1]: if u (of length > 1) is a k-local geodesic in a δ-hyperbolic graph, with $k \geq 4\delta$, then the distance between the endpoints of u is at least $|u|/2 + 1$.

So suppose that Hw is reduced according to the eda. If w has length at most R, then since Hw is reduced by rules of type (1), Hw is geodesic in Σ, and the inequality in Proposition 3.2 holds with $k = 1$.

So suppose that $|w| > R$, and let w_1 be the prefix of length R of w. We aim to show that every vertex of Σ that comes after w_1 on the path from 1_Σ labelled w lies outside of $B_K(1_\Sigma)$. Choose w_2 so that w_1w_2 is maximal as a prefix of w subject to all vertices of w_2 lying outside of $B_K(1_\Sigma)$. Then, since w_1 is geodesic of length $R = 2K$, we have $|w_2| \geq K - 1$, and so $|w_1w_2| \geq 3K - 1$. Since w_1 is geodesic in Σ, and that part of the path labelled w_1w_2 that lies outside of the K-ball is a k-local geodesic in Σ (because, by GIB(k), it is isometric to the corresponding word in the Cayley graph, and our inclusion of the rules of type (2) in the eda ensures that the reduced words over X of length $\leq k$ are geodesics), we see that the whole of the path labelled w_1w_2 is a k-local geodesic in Σ. So we can apply [10, Proposition 2.1] to deduce from the δ-hyperbolicity of Σ that

$$d_\Sigma(1_\Sigma, Hw_1w_2) \geq |w_1w_2|/2 + 1 \geq (3K + 1)/2 > K + 1.$$

It follows that, if w_1w_2 were not already equal to w, then it would be extendible to a longer prefix of w subject to all vertices of w_2 lying outside of $B_K(1_\Sigma)$. So $w_1w_2 = w$, and the above inequality gives us the linear lower bound $d_\Sigma(1_\Sigma, Hw) \geq |w|/2 + 1 \geq |w|/2$ on $d_\Sigma(1_\Sigma, Hw)$. So the inequality in Proposition 3.2 holds with $k = 2$ and hence, by the preceding paragraph, it holds with $k = 2$ for all words w such that Hw is reduced according to the eda. The result now follows from Proposition 3.2, and this completes the proof of Theorem 1.2 (ii).

7 Proof of Theorem 1.3

Let F be a free subgroup of $G = \langle X \rangle$ with $|G : F|$ finite, and let Y be the inverse closure of a free generating set for F; that is the union of a free generating set with its inverses. Let $K = F \cap H$. The subgroup K has finite index in H, and so must (like H) be finitely generated. It easy to see that the elements in any right transversal of K in H lie in different cosets of F in G, so we can extend a right transversal $T' = \{t_1, \ldots, t_m\}$ of K in H to a right transversal $T = \{t_1, \ldots, t_n\}$ of F in G. Then any word $w \in X^*$ can be expressed (in G) as a word in U^*T, where U is the set $\{u(i, x) : 1 \leq i \leq n,\ x \in X\}$ of Schreier generators for F in G defined by the equations $t_ix = u(i, x)t_j$. (Note that X inverse-closed implies that U is inverse-closed.) Then, by substituting the reduced word in Y^* for each $u(i, x)$, the word w can be written as a word vt in Y^*T (where v is not necessarily freely reduced).

The first step to recognise whether $w \in H$ is to rewrite it to the form vt, as above, using a transducer. Then $w \in H$ if and only if $t \in T'$ and $v \in K$. It remains for us to describe the operation of a deterministic pushdown automaton (pda) N to recognise those words v in Y^* that lie in the subgroup K of the free group F. Note that this machine operates simultaneously, rather than sequentially, with the transducer, and it follows from the fact that context-free languages are closed under inverse gsms [12, Example 11.1, Theorem 11.2] that the combination of the two machines is a pda.

By [2] or [8, Proposition 4.1], any finitely generated subgroup K of a free group is L-*rational*, where L is the set of freely reduced words over a free generating set; that is, the set $K \cap L$ is a regular language. In our case, we choose L to ber the freely reduced words over Y.

We shall build our pda N out of a finite state automaton (fsa) M for which $L(M) \cap L = K \cap L$ (where $L(M)$ is the set of words accepted by M). The construction is (in effect) described in the proof of [8, Theorem 2.2] that K is L-quasiconvex. The L-quasiconvexity condition is equivalent to the property that all prefixes of freely reduced words that represent elements of K lie within a bounded distance of K in the Schreier graph $\Sigma := \Sigma(F, K, Y)$. Equivalently, a freely reduced word v over Y represents an element of K precisely if it labels a loop in Σ from 1_Σ to 1_Σ that does not leave a particular bounded neighbourhood $B = B_d(1_\Sigma)$ of 1_Σ.

Suppose that $g_1 = 1$ and that $K = Kg_1, Kg_2, \ldots, Kg_r$ are the right cosets corresponding to the bounded neighbourhood B of 1_Σ within Σ that is identified above. The fsa M is defined as follows.

(i) The states of M are denoted by $\sigma_1, \ldots, \sigma_r, \hat{\sigma}$.

(ii) The states $\sigma_1, \ldots, \sigma_r$ correspond to the right cosets Kg_i of K in F, where each g_i is in the finite subset B identified above; indeed we may identify σ_i with the coset Kg_i, and then use the name B both for the set $\{Kg_1, \ldots, Kg_r\}$ of cosets and for the set $\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$ of states. The state σ_1 (which corresponds to the subgroup K) is the start state and the single accepting state.

(iii) For $1 \le i, j \le r$ and $y \in Y$, there is a transition $\sigma_i^y = \sigma_j$ if and only if $Kg_i y = Kg_j$. It follows from this that $\sigma_i^y = \sigma_j$ if and only if $\sigma_j^{y^{-1}} = \sigma_i$.

(iv) $\hat{\sigma}$ is a failure state, and is the target of all transitions that are not defined in (iii), including those from $\hat{\sigma}$.

We see that, as a word is read by M, the automaton keeps track of the coset of Σ that contains Kw, where w is the prefix that has been read so far, so long as that coset is within the finite neighbourhood B of 1_Σ, and in addition so are all cosets Kw' for which w' is a prefix of w. The L-quasiconvexity of K ensures that a word in L is accepted by M if and only if it represents an element of K. In fact any word over Y that is accepted by M must represent an element of K.

However words over Y that are not freely-reduced (that is, not in L) and do not stay inside of B will be rejected by M, even when they represent elements of K. In order to construct a machine that accepts all words v over Y within K, and not simply those that are also freely-reduced, we need to combine the operation of the fsa M above with a stack, which we use to compute the free reduction.

We construct our pda N to have the same state set $B \cup \{\hat{\sigma}\}$ as M, again with $\sigma_1 = Kg_1 = K$ as the start state and sole accepting state. The transitions from the states $\sigma_i = Kg_i$ are as in M. We need however to describe the operation of the stack, and transitions from the state $\hat{\sigma}$, which is non-accepting, but no longer a failure state.

The stack alphabet is the set $Y \cup (Y \times B)$. The second component of an element of $Y \times B$ is used to record the state M is in immediately before it enters the state $\hat{\sigma}$. In addition, we use the stack to store the free reduction of the prefix of v that has been read so far.

The operations of the pda N that correspond to the various transitions of M are described in the following table. The absence of an entry in the 'push' column indicates that nothing is pushed.

Transition of M	Input state	Input symbol	Pop	Push	Output state
$\sigma_i^y = \sigma_j$	σ_i	y	y^{-1}		σ_j
	σ_i	y	$y' \neq y^{-1}$	$y'y$	σ_j
$\sigma_i^y = \hat{\sigma}$	σ_i	y	y'	$y'(y, \sigma_i)$	$\hat{\sigma}$
$\hat{\sigma}^y = \hat{\sigma}$	$\hat{\sigma}$	y	(y^{-1}, σ_i)		σ_i
	$\hat{\sigma}$	y	(y', σ_i), $\quad y' \neq y^{-1}$	$(y', \sigma_i)y$	$\hat{\sigma}$
	$\hat{\sigma}$	y	y^{-1}		$\hat{\sigma}$
	$\hat{\sigma}$	y	$y' \neq y^{-1}$	$y'y$	$\hat{\sigma}$

Note that we have not specified that $y' \neq y^{-1}$ in line 3, but in fact the condition $y' = y^{-1}$ does not arise in this situation. Since there can be a symbol in $Y \times B$ on the stack only when N is in state $\hat{\sigma}$, it is not possible to pop such a symbol when N is in state σ_i, so there are no such entries in the table.

The fact that N recognises $\mathsf{GWP}(F, K, Y)$ follows from the fact that N accepts w if and only if M accepts $\overline{w}$, where $\overline{w}$ is the free reduction (in L) of w. We prove this by induction on the number k of reductions of the form $w_1yy^{-1}w_2 \to w_1w_2$ with $y \in Y$ that we need to apply to reduce w to $\overline{w}$.

The case $k = 0$ of our induction follows from the fact that $L(M) \cap L = K \cap L$, combined with the observation that, if w is freely reduced, then w leads to the same state of M as it does of N. For in that case the only possible transitions as we read w are of the types described in lines 2,3,5,7 of the table.

For $k > 0$ it is enough to prove the statement

> $(*)$: if $wy \in L$, then the configuration of N after reading wyy^{-1} is identical to the configuration after reading w.

It follows from $(*)$ that a word $w_1yy^{-1}w_2$ in which yy^{-1} is the leftmost cancelling pair is accepted by N if and only if w_1w_2 is accepted by N, and hence we have the inductive step we need.

We can check the statement $(*)$ with reference to the table. There are up to seven possibilities for the type of transition of N as the final symbol y of wy is read.

For the first two of these, N is in state σ_i after reading w, and moves to a state σ_j. Since wy is in L, the top stack symbol after reading w is not y^{-1}. Hence the transition must be of the type described in line 2 of the table, and not as in line 1, that is, $y' \neq y^{-1}$ is popped, and then $y'y$ is pushed. Recalling that $\sigma_i^y = \sigma_j$ in M if and only if $\sigma_j^{y^{-1}} = \sigma_i$, we see that the next transition of N, from wy on y^{-1}, is of the type described in line 1. Then the symbol y is popped, the symbol y' is again on the top of the stack, and N returns to the state σ_i

We consider similarly the remaining five possibilities for the transition from w on y', and the subsequent transitions on y'^{-1}, and verify $(*)$ for each of those configurations. This completes the proof of Theorem 1.3.

8 Proof of Theorem 1.4

Our proof of Theorem 1.4 has the same structure as the proof in [14] that groups with $\mathsf{WP}(G)$ context-free are virtually free, and it would be helpful for the reader to be familiar with that proof.

We shall prove that G has more than one end. Assuming that to be true, we use Stalling's theorem [18] to conclude that G has a decomposition as an amalgamated free product $G = G_1 *_K G_2$, or as an HNN-extension $G = G_1 *_{K,t}$, over a finite subgroup K. Since G_1 (and G_2) are easily seen to be quasiconvex subgroups of G, it is not hard to show that the hypotheses of the theorem are inherited by the subgroup $H \cap G_1$ of G_1 (and $H \cap G_2$ of G_2), and so they too have more than one end, and we can apply the Dunwoody accessibility result to conclude that G is virtually free.

So we just need to prove that G has more than one end. Fix a finite inverse-closed generating set X of G. Then, as in [14], we consider a context-free grammar in Chomsky normal form with no useless variables that derives $\mathsf{GWP}(G, H, X)$. More precisely, we suppose that each rule has the form $S \to \varepsilon$ (where S is the start symbol), $z \to z'z''$ or $z \to a$, where z, z', z'' are variables, and a is terminal, and we assume that S does not occur on the right hand side of any derivation. When a word w' can be derived from a word w by application of a single grammatical rule we write $w \Rightarrow w'$, and when a sequence of such rules is needed we write $w \Rightarrow^* w'$.

Let $z_1, \ldots, z_n$ be the variables of the grammar other than S and, for each z_i let u_i be a shortest word in X^* with $z_i \Rightarrow^* u_i$. Let L be the maximum length of the words u_i.

Let $w \in \mathsf{GWP}(G, H, X)$ with $|w| > 3$, and fix a derivation of w in the grammar. We shall define a planar X-graph Δ with an associated X-graph homomorphism $\phi : \Delta \to \Sigma := \Sigma(G, H, X)$. We start with a simple plane polygon with a base point, and edges labelled by the letters of w, and with ϕ mapping the base-point of Δ to 1_Σ. Note that ϕ is not necessarily injective.

If z_i occurs in the chosen derivation of w, then we have $w = vv_iv'$ with $z_i \Rightarrow^* v_i$; two such words v_i and v_j are either disjoint as subwords of w or related by containment. Since $z_i \Rightarrow^* u_i$, we also have $vu_iv' \in \mathsf{GWP}(G,H,X)$. So we can draw a chord labelled u_i in the interior of Δ between the two ends of the subpath labelled v_i, and ϕ extends to this extension of Δ. If we do this for each such z_i for which $1 < |v_i| < |w| - 1$ then, as in [14, Theorem 1], we get a 'diagonal triangulation' of Δ, in which the sides are either boundary edges of Δ or internal chords of length at most L. (But note that, for the first derivation $S \to z_1z_2$, say, if $|v_1| > 1$ and $|v_2| > 1$ then, to avoid an internal bigon, we omit the chord labelled u_2.)

Suppose, for a contradiction, that G has just one end; that is, for any R, the complement in $\Gamma(G,X)$ of any ball of radius R is connected. Then, for any R, we can find a word $w_1w_2w_3$ over X with $w_1w_2w_3 =_G 1$ which, starting at 1_Γ, labels a simple closed path in $\Gamma(G,X)$, where $|w_1| = |w_3| = R$, w_3w_1 is geodesic, and no vertex in the path labelled w_2 is at distance less than R from 1_Γ.

Choose such a path with $R = 3L+1$. Choose k' such that the whole path lies in the ball $B_{k'}(1_\Gamma)$ of $\Gamma(G,X)$, and let $k = k' + L$. Then, since by Proposition 6.1 $\Sigma(G,H,X)$ satisfies GIB(k), there exists K such that, for any vertex p of $\Sigma(G,H,X)$ with $d(1_\Sigma, p) \geq K$, the ball $B_k(p)$ of $\Sigma(G,H,X)$ is X-graph isomorphic to the ball $B_k(1_\Gamma)$ of $\Gamma(G,X)$. Choose such a vertex p, and consider the path labelled $w_1w_2w_3$ of $\Sigma(G,H,X)$ that is based at p, as in Fig. 2.

Choose a vertex q on the path labelled $w_1w_2w_3$ with $d(1_\Sigma, q)$ minimal, and let w_4 be the label of a geodesic path in $\Sigma(G,H,X)$ from 1_Σ to q. Then, for some cyclic permutation w' of $w_1w_2w_3$, we have a closed path in $\Sigma(G,H,X)$ based at 1_Σ and labelled $w_4w'w_4^{-1}$.

We apply the above triangulation process to a planar X-graph Δ for $w_4w'w_4^{-1}$. Since q is the closest vertex to 1_Σ on the loop labelled by w', and the path from 1_Σ to q in the Cayley graph labelled by w_4 is geodesic, every vertex on that path is as close to q as to any other vertex of w', and so any vertex of w_4 that can be connected by the image of a chord of Δ' to a vertex of w' must be within distance at most L of q. Let r be the first such vertex on w_4 (as we move from 1_Σ to q), and let w_5 be the suffix of w_4 that labels the path along w_4 from r to q. Then $|w_5| \leq L$.

So we can derive from our triangulation of Δ a triangulation of a planar diagram Δ' for the word $w_5w'w_5^{-1}$, and there is an associated X-graph homomorphism ϕ' that maps this to the corresponding subpath in $\Sigma(G,H,X)$. (Note that the images of w_5 and w_5^{-1} under ϕ' are equal, but that ϕ' is injective when restricted to w'.) By our choice of $k = k' + L$, the image of ϕ' lies entirely within $B_k(p)$, which is X-graph isomorphic to $B_k(1_\Gamma)$. So the distances in $\Sigma(G,H,X)$ between vertices in this image are the same as in any path with the same label in $\Gamma(G,X)$.

As in [14], we colour, using three colours, the vertices of the boundary paths of Δ' that are labelled w_1, w_2, w_3 (where vertices on two of these subwords get both associated colours), and we colour the vertices on w_5 and w_5^{-1} with the same colour (or colours) as q. As in [14, Lemma 5], we conclude that there is a triangle in the triangulation whose vertices use all three colours between them. One (or even two) of these vertices could be on the subpath labelled w_5, and two 2-coloured vertices in the triangle might coincide, but, since any vertex on

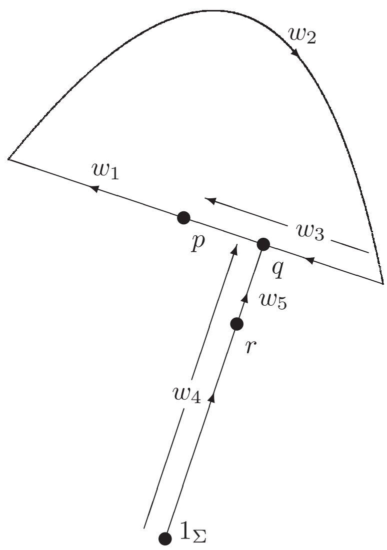

Figure 2: Triangulation in $\Sigma(G, H, X)$

w_5 is within distance L of q, replacing vertices on w_5 by q as necessary, we end up with a triangle of three (not necessarily distinct) vertices p_1, p_2, p_3 with p_i on w_i, and with $d(p_i, p_j) \leq 2L$ for each i, j. At least one of p_1, p_3 must be within distance L of p. But then $d(p, p_2) \leq 3L$, contradicting our assumption that w_2 is outside $B_{3L}(p)$. This completes the proof of Theorem 1.4.

9 Sketch of proof of Proposition 6.1

Suppose first that $|C_G(h) : C_H(h)|$ is infinite for some $1 \neq h \in H$, and let w be a word representing h. Then, for any $K > 0$, there exists a word $v \in C_G(h)$ labelling a path in $\Sigma := \Sigma(G, H, X)$ from 1_Σ to a vertex p with $d(1_\Sigma, p) > K$, and there is a loop labelled w based at p in Σ, but no such loop based at 1_Γ in $\Gamma := \Gamma(G, X)$. So GIB($|w|$) fails in Σ.

Suppose conversely that GIB(k) fails in Σ for some k. Then there are vertices p of Σ at arbitrarily large distance from 1_Σ such that the ball $B_k(p)$ in Σ is not X-graph isomorphic to the ball $B_k(1_\Gamma)$ in Γ. So the natural labelled graph morphism $B_k(1_\Gamma) \to B_k(p)$ with $1_\Gamma \mapsto p$ is not injective, and hence two distinct vertices of $B_k(1_\Gamma)$ map to the same vertex of $B_k(p)$. So, for any such vertex p, there is at least one labelled loop based at p, within $B_k(p)$, such that the corresponding labelled path based at in $B_k(1_\Gamma)$ in $B_k(1_\Gamma)$ is not a loop in Γ. Since the number of words that can label loops in a ball of radius k in Σ is finite, some word w with $w \neq_G 1$ must label loops based at p for infinitely many vertices p of Σ, and we can choose w to be geodesic over X. Now for any integer N, there is a word v of length greater that N, labelling a geodesic in Σ from 1_Σ to a vertex p, from which there is a loop in Σ labelled by w.

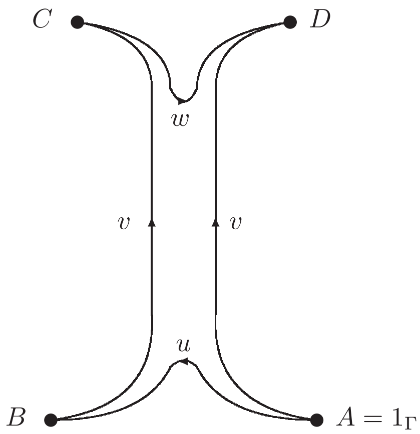

Figure 3: The geodesic quadrilateral $ABCD$

For such a word v, we have $hvw = v$ for some $h \in H$. Let u be a geodesic word labelling h. Then we have a geodesic quadrilateral with vertices $A = 1_\Gamma, B, C, D$ in $\Gamma(G, X)$ with sides AB, BC, CD, AD labelled u, v, w, v, respectively, as shown in Fig. 3.

By the hyperbolicity of G, each vertex of AB lies within a distance 2δ of some vertex on BC, CD or DA, where δ is the constant of hyperbolicity. Furthermore, since H is quasiconvex in G, there is a constant λ, such that each vertex of AB is within a distance λ of a vertex of Γ representing an element of H. Since each vertex of w lies at distance at least $|v| - k$ from any vertex in H, by choosing $|v| > k + 2\delta + \lambda$ we can ensure that none of the vertices of AB is 2δ-close to any vertex of CD. So the vertices of AB must all be 2δ-close to vertices in BC or DA. But, since v labels a geodesic path from 1_Σ in Σ, at most $2\delta + \lambda$ vertices on BC or on DA can be within $2\delta + \lambda$ of a vertex in H. So each vertex of AB is at distance at most 2δ from one of at most $4\delta + 2\lambda$ vertices and, since the total number of vertices in Γ with that property is bounded, we see that $|AB| = |u|$ is bounded by some expression in $|X|$, δ and λ.

By hyperbolicity of G, the two paths BC and AD labelled v must synchronously L-fellow travel for some L (which depends on the upper bounds on $|w|$ and $|u|$). Let $m > 0$. Then, by choosing v sufficiently long, we can ensure that some word u' appears as a word-difference between BC and AD at least m times; that is, v has consecutive subwords $v_0, \ldots, v_m, v'$, such that $v = v_0 v_1 v_2 \cdots v_m v'$, and $h v_0 v_1 v_2 \cdots v_i u' =_G v_0 v_1 v_2 \cdots v_i$ for each i with $0 \leq i \leq m$. The case $i = 0$ gives $u' = v_0^{-1} h^{-1} v_0$, and it follows from this that $g_i := v_0 (v_1 v_2 \cdots v_i) v_0^{-1} \in C_G(h)$ for $1 \leq i \leq m$. Also, since v labels a geodesic in Σ, the elements $v_0 v_1, v_0 v_1 v_2, \ldots v_0 v_1 v_2 \cdots v_m$ lie in distinct cosets of H and hence so do the g_i. Since we can choose m arbitrarily large, this contradicts the finiteness of $|C_G(h) : C_H(h)|$.

Acknowledgements

The first author was supported by the Swiss National Science Foundation grant Professorship FN PP00P2-144681/1, and would like to thank the mathematics departments of the Universities of Newcastle and Warwick for their support and hospitality.

References

[1] J. Alonso, T. Brady, D. Cooper, V. Ferlini, M. Lustig, M. Mihalik, M. Shapiro and H. Short, "Notes on word-hyperbolic groups", in E. Ghys, A. Haefliger and A. Verjovsky, eds., *"Proceedings of the Conference "Group Theory from a Geometric Viewpoint" held in I.C.T.P., Trieste, March 1990*, World Scientific, Singapore, 1991.

[2] A.V. Anisimov and F.D. Seifert. Zur algebraischen Charakteristik der durch kontext-freie Sprachen definierten Gruppen. *Elektron. Informationsverarb. Kybernet.*, 11:695–702, 1975.

[3] Y. Antolín and L. Ciobanu, Finite generating sets for relatively hyperbolic groups and applications to conjugacy languages, *Trans. Amer. Math. Soc.*, Volume 368, Number 11 (2016), 7965–8010.

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

[5] B. Farb, Relatively hyperbolic groups. Geom. Funct. Anal. 8 (1998) no. 5, 810–840.

[6] R. Foord. Automaticity and growth in certain classes of groups and monoids. PhD thesis, University of Warwick, 2000.

[7] S. Friedl and H. Wilton. The membership problem for 3-manifold groups is solvable. *Alg. & Geom. Topology*, 16 (2016) 1827–1850.

[8] S.M. Gersten and H.B. Short. Rational subgroups of biautomatic groups. *Ann. of Math.*, 134(1):125–158, 1991.

[9] O. Goodman and M. Shapiro, On a generalization of Dehn's algorithm. *Internat. J. Algebra Comput.*, 18 (2008) 1137–1177.

[10] D.F. Holt, Word-hyperbolic groups have real-time word problem. *Internat. J. Algebra Comput.* 10 (2000) 221–227.

[11] D.F. Holt and S. Rees, Solving the word problem in real time, *J. London Math Soc.* 63 (2001) 623–639.

[12] John E. Hopcroft and Jeffrey D. Ullman, *Introduction to automata theory, languages and computation*, Addison-Wesley, 1979.

[13] I. Kapovich. The geometry of relative Cayley graphs for subgroups of hyperbolic groups, `http://arxiv.org/abs/math/0201045v2`.

[14] D.E. Muller and P.E. Schupp. Groups, the theory of ends, and context-free languages. *J. Comp. System Sci.*, 26:295–310, 1983.

[15] D. Osin, Relatively hyperbolic groups: intrinsic geometry, algebraic properties and algorithmic problems. Mem. Amer. Math. Soc. 179 (2006) no. 843, vi+100pp.

[16] E. Rips. Subgroups of small cancellation groups, *Bulletin London Math. Soc.*, 14(1):45–47, 1982.

[17] Arnold L. Rosenberg, Real-time definable languages, J. Assoc. Comput. Mach. 14 (1967) 645–662.

[18] J. Stallings. *Group Theory and Three-Dimensional Manifolds*, volume 4 of *Yale Mathematical Monographs*. Yale University Press, 1971.

Laura Ciobanu, Mathematical and Computer Sciences, Colin McLaurin Building, Heriot-Watt University, Edinburgh EH14 4AS, UK

E-mail address: `l.ciobanu@hw.ac.uk`

Derek Holt, Mathematics Institute, Zeeman Building, University of Warwick, Coventry CV4 7AL, UK

E-mail address: `D.F.Holt@warwick.ac.uk`

Sarah Rees, School of Mathematics and Statistics, University of Newcastle, Newcastle NE1 7RU, UK

E-mail address: `Sarah.Rees@ncl.ac.uk`