QUASICONVEXITY AND DEHN FILLING

DANIEL GROVES AND JASON FOX MANNING

ABSTRACT. We define a new condition on relatively hyperbolic Dehn filling which allows us to control the behavior of a relatively quasiconvex subgroups which need not be full. As an application, in combination with recent work of Cooper and Futer [6], we provide a new proof of the virtual fibering of non-compact finite-volume hyperbolic 3–manifolds, a result first proved by Wise [28]. Additionally, we explain how the results of [2, Appendix A] can be generalized to the relative setting to control the relative height of relatively quasiconvex subgroups under appropriate Dehn fillings.

1. INTRODUCTION

Dehn filling results for hyperbolic and relatively hyperbolic groups have been used to great effect in recent years, notably in solving the isomorphism problem for a broad class of relatively hyperbolic groups [9], and as part of Agol's proof of the Virtual Haken Conjecture [2] (see particularly the Appendix to [2] and [4]). In many of these results a key ingredient is the control of relatively quasiconvex subgroups under Dehn filling, building on techniques developed in [3].

In previous work, this control was limited by the requirement that the fillings be 'H–fillings', for a relatively quasiconvex subgroup H. This requirement is mild when H is *full* (in the sense that each infinite intersection of H with a maximal parabolic subgroup is finite index in that parabolic), but more restrictive for general relatively quasiconvex subgroups. One way to avoid this issue is to apply combination theorems such as those in [19, 18], etc. to enlarge relatively quasiconvex subgroups to full ones. Even in case this is possible, the methods of this paper are conceptually simpler as they avoid this intermediate enlargement step.

1

In this paper, we propose a new condition on relatively hyperbolic Dehn filling, which we call H–*wide*, which is applicable to relatively quasiconvex subgroups H in much greater generality than previous techniques. We prove that under sufficiently long and H–wide fillings, the same control can be had over the behavior of a relatively quasiconvex subgroup H under filling as could be obtained with sufficiently long H–fillings in the previous works.

We provide two main applications. First, we combine our work with recent work of Cooper and Futer [6] to give a new proof of the following theorem of Wise.

Theorem A. [28, Theorem 14.29] Suppose that G is the fundamental group of a non-compact finite-volume hyperbolic 3–manifold. Then G is virtually compact special.

The study of (virtually) special cube complexes and groups was initiated by Haglund and Wise in [13]; we refer to that paper for the definition and basic properties. The following is an immediate consequence of Theorem A and Agol's criterion for fibering [1, Theorem 1.1].

Corollary B. Suppose that M is a non-compact finite-volume hyperbolic 3–manifold. Then M has a finite-sheeted cover which fibers over the circle.

Previous to this paper, Wise's unpublished manuscript [28] contained the only proof that a non-compact finite-volume hyperbolic 3-manifold virtually fibers over a circle. Our proof does not rely on any results from [28] (and neither does the one in [6]).

The second application we provide is to use H–wide fillings to explain how the results from [2, Appendix A] can be generalized to control the 'relative height' of relatively quasiconvex subgroups under Dehn fillings. We apply this to prove a result (Theorem 7.18) needed by Wilton and Zalesskii in their work [27] on profinite rigidity of 3–manifold groups.

1.1. **On virtual fibering of hyperbolic** 3–**manifolds.** Agol proved in [2] that fundamental groups of closed hyperbolic 3–manifolds are virtually special, which implies that these manifolds are virtually Haken, and virtually fibered. He also proved that Kleinian groups are LERF, and large.

In the non-compact but finite-volume case, the LERF and large results are included in Agol's result, and these manifolds are well known to be Haken. However virtual fibering in the non-compact case is not covered by [2].

We make some comments on Wise's proof of Theorem A in the most recent publicly available version of [28], dated October 29, 2012. This proof relies on [28, Theorem 16.28], which in turn uses [28, Theorem 16.16] in three places. This last result is about the separability of quasiconvex subgroups of certain graphs of virtually sparse special groups. In the proof of [28, Theorem 16.16], Wise asserts that relative quasiconvexity of subgroups of relatively hyperbolic groups persists under sufficiently long fillings and refers to Osin [22]. Such a result is not in [22] and in fact a correct formulation requires some care (see Example 4.8 below).

As explained above, existing results about controlling relatively quasiconvex subgroups under Dehn filling, such as those in [3, 2] and also [28, Theorem 15.6], apply to relatively quasiconvex subgroups which are full, and they do not apply in the setting needed in [28, Theorem 16.16].

We believe the above issue in the proof of [28, Theorem 16.16] can be fixed using either a 'Combination Theorem' approach or techniques as in the current paper. However, one of our goals here is to use the advances of the last five years to give an alternative proof of virtual fibering in the non-compact setting.

1.2. **Outline.** In Section 2 we recall the basic concepts about relatively hyperbolic groups, relatively quasiconvex subgroups, and Dehn filling. In Section 3 we introduce the notion of H–wide fillings. In Section 4 we prove that the behavior of a relatively quasiconvex subgroup H under sufficiently long and H–wide fillings is well-controlled. We also give an example to show that this is not true without the assumption of H–wideness. In Section 5 we prove that in certain circumstances we can ensure the existence of appropriate H–wide fillings. In Section 6 we provide the application to virtual specialness of fundamental groups of finite-volume hyperbolic 3–manifolds. Finally, in Section 7, we prove that the results about height in the hyperbolic setting from [2, Appendix A] can be generalized to control relative height under sufficiently long and H–wide fillings. These results may be of independent interest. As an application, we prove Theorem 7.18, the result required by Wilton and Zalesskii.

2. Background

For background on relatively hyperbolic groups, their associated cusped spaces, and relatively hyperbolic Dehn filling see [11]. For background on relatively quasiconvex subgroups see [3, 15]. We always work in a *combinatorial cusped space* $X = X(G, \mathcal{P}, S)$, where S is some chosen generating set for G which also contains generating sets for the peripheral groups $P \in \mathcal{P}$. This cusped space contains a copy of the Cayley graph of G with respect to the generators S. The *depth* of a vertex of X is its distance to the Cayley graph.

Suppose that $(G, \mathcal{P})$ is relatively hyperbolic. We fix a combinatorial cusped space X for the pair $(G, \mathcal{P})$ as in [11]. Since $(G, \mathcal{P})$ is relatively hyperbolic, X is Gromov hyperbolic. Unless otherwise stated, δ is a hyperbolicity constant for the space X.

Recall that a *Dehn filling* of $(G, \mathcal{P})$ is determined by a collection $\mathcal{N} = \{N_P\}_{P \in \mathcal{P}}$ of normal subgroups $N_P \lhd P$. The Dehn filling is the quotient

$$G(\mathcal{N}) = G / \langle\!\langle \cup N_P \rangle\!\rangle.$$

The *peripheral groups* of $G(\mathcal{N})$ are the images of the elements of $\mathcal{P}$ in $G(\mathcal{N})$. We often abbreviate this as $(G, \mathcal{P}) \to (\overline{G}, \overline{\mathcal{P}})$. A statement S holds *for all sufficiently long fillings* if there is a finite set $B \subset \cup P \smallsetminus \{1\}$ so that S holds for any fillings $G(\mathcal{N})$ so that $N_P \cap B = \emptyset$ for all $P \in \mathcal{P}$. If $(\overline{G}, \overline{\mathcal{P}})$ is a Dehn filling of $(G, \mathcal{P})$, with $\overline{G} = G/K$, then one obtains a combinatorial cusped space $\overline{X}$ for $(\overline{G}, \overline{\mathcal{P}})$ by taking $\overline{X}$ equal to $_K\backslash X$ with self-loops removed. In fact, since the self-loops do not affect the metric on the zero-skeleton, we ignore the issue of removing them and abuse notation by setting $\overline{X} = {}_K\backslash X$.

The following result is key to any approach to relatively hyperbolic Dehn filling theorems using the cusped space.

Theorem 2.1. *Using the cusped spaces just described, let B be a finite metric ball in X. For all sufficiently long fillings the quotient map $X \to {}_K\backslash X$ restricts on B to an isometric embedding whose image is a metric ball.*

Proof. This follows immediately from [22, Theorem 1.1]. □

Using the coarse Cartan-Hadamard Theorem [7, A.1] and the uniform hyperbolicity of combinatorial horoballs [11, Theorem 3.8] we obtain the following corollary, which was stated in a slightly weaker form as [3, Proposition 2.3].

Proposition 2.2. *For all $\delta > 0$ there is a $\delta' > 0$ so that if the combinatorial cusped space X is δ–hyperbolic, then for all sufficiently long*

fillings, the combinatorial cusped space $\overline{X}$ of the Dehn filling is δ'–hyperbolic.

Remark 2.3. Proposition 2.2 implies that $(\overline{G}, \overline{\mathcal{P}})$ is relatively hyperbolic, using only Theorem 2.1. The proof of Theorem 2.1 in [11] used the bicombing of X by *preferred paths*, whereas the proof that the cusped space of $\overline{X}$ is Gromov hyperbolic used a homological bicombing which used an adaptation of results of Mineyev from [21]. Using the above approach allows one to avoid the homological bicombing in [11] entirely.

Definition 2.4. Suppose that $(G, \mathcal{P})$ is relatively hyperbolic with associated cusped space X. Let A be a horoball in X, and let $R > 0$. A geodesic *penetrates A to depth R* or *R–penetrates A* if it contains a point in A at depth R.

Suppose $H \leq G$. Then A is *R–penetrated by H* if there is a geodesic γ with endpoints in H which R–penetrates A.

Recall the following result from [18].

Proposition 2.5. [18, Proposition A.6] *Let $(G, \mathcal{P})$ be relatively hyperbolic and $H \leq G$ be relatively quasiconvex. There is a constant R so that whenever a horoball A of X is R–penetrated by H then the intersection of H with the stabilizer of the horoball is infinite.*

The following is a combination of Lemma 3.3 from [12], and a statement implicit in its proof.

Lemma 2.6. [12, Lemma 3.3] *Suppose $(G, \mathcal{P})$ is relatively hyperbolic, with δ–hyperbolic combinatorial cusped space X. Suppose further that P_1 and P_2 are distinct conjugates of elements of $\mathcal{P}$, and that $F = P_1 \cap P_2$. Then F acts freely on some set Q in X which lies in the Cayley graph and has diameter (in X) at most $2\delta + 1$.*

In particular, there is a constant C depending only on δ and the cardinality of the generating set S so that $\#F \leq C$.

The second part of the Lemma says that if P_1 and P_2 are distinct maximal parabolics, then $\#P_1 \cap P_2 \leq C$. In other words, the family $\mathcal{P}$ is *C–almost malnormal*. In particular, for a parabolic subgroup A of size more than C, there is no ambiguity about which $g \in G$ and which $P \in \mathcal{P}$ has $A \leq P^g$ (up to the choice of conjugating element in gP).

The following result was stated without proof as [12, Proposition 3.4]. The proof that we provide here is more elementary than the one suggested in [12].

Proposition 2.7. *If $(G, \mathcal{P})$ is relatively hyperbolic, and $\mathcal{P}$ is C–almost malnormal, then for all sufficiently long fillings $(\overline{G}, \overline{\mathcal{P}})$ of $(G, \mathcal{P})$, the collection $\overline{\mathcal{P}}$ is C–almost malnormal.*

Proof. By Proposition 2.2, there is a δ' so that for all sufficiently long fillings $(\overline{G}, \overline{\mathcal{P}})$ of $(G, \mathcal{P})$, the cusped space $\overline{X}$ for $(\overline{G}, \overline{\mathcal{P}})$ is δ'–hyperbolic. Fix a filling $(\overline{G}, \overline{\mathcal{P}})$ so that the induced map between cusped spaces is injective on any ball of radius $100\delta'$ centered in the Cayley graph of X (see Theorem 2.1).

Suppose that $\overline{P}_1$ and $\overline{P}_2$ are distinct conjugates of elements of $\overline{\mathcal{P}}$, and let $\overline{F} = \overline{P}_1 \cap \overline{P}_2$. There are horoballs $\overline{A}_1$ and $\overline{A}_2$ in $\overline{X}$ so that $\overline{P}_i$ stabilizes $\overline{A}_i$. As explained in the proof of [12, Lemma 3.3], the subgroup $\overline{F}$ acts freely on a subset of the Cayley graph of $\overline{G}$ in $\overline{X}$ of diameter at most $2\delta' + 1$. Moreover, it is clear by considering the $\overline{F}$–orbit of a geodesic between the limit points of $\overline{A}_1$ and $\overline{A}_2$ in ∂X that there are also subsets Q_1 and Q_2 of diameter at most $2\delta' + 1$ so that $\overline{F}$ acts freely on each Q_i and Q_i is contained in $\overline{A}_i$ at depth $5\delta'$.

Suppose that a geodesic between Q_1 and Q_2 $10\delta'$–penetrates some other horoball B. Then let B be the closest such horoball to $\overline{A}_1$, and replace $\overline{A}_2$ by B and Q_2 by an $\overline{F}$–invariant subset Q of B at depth $5\delta'$ and diameter at most $2\delta' + 1$. In this manner, we may suppose that any geodesic between Q_1 and Q_2 stays within a $10\delta'$–neighborhood of the Cayley graph in $\overline{X'}$.

We may thus lift Q_1, Q_2 and the geodesics between them to X. To see that this is possible, consider that any pair of points in Q_1 and pair of points in Q_2 are the vertices of a geodesic quadrilateral with two sides of length at most $5\delta'$ and so can be filled with a disk which lies entirely within a $20\delta'$–neighborhood of the Cayley graph. In an entirely similar way to the proof of [12, Theorem 4.1] (a result whose proof did not rely on the result we are currently trying to prove), it now follows that $\overline{F}$ can be lifted bijectively to a finite subgroup F of G which stabilizes two distinct horoballs. Because $\mathcal{P}$ is C–almost malnormal, it follows that $|\overline{F}| = |F| \leq C$, which is what we were required to prove. $\square$

Definition 2.8. Suppose that $(G, \mathcal{P})$ is relatively hyperbolic and that X is a cusped space for $(G, \mathcal{P})$ which is δ–hyperbolic. A parabolic subgroup Q of G is *uniquely parabolic* if there is a unique conjugate of an element of $\mathcal{P}$ which contains Q.

It follows from [12, Lemma 3.3] that there is a constant C so that any parabolic subgroup of size more than C, and in particular any infinite parabolic subgroup, is uniquely parabolic.

It is an immediate consequence of the definition that a uniquely parabolic subgroup stabilizes a unique horoball in the cusped space.

In order to fix notation, we recall a definition from [3, Section 3], and slightly adapt the notation from there. Let $H \leq G$. Suppose that $\mathcal{D}$ is a collection of representatives of H–conjugacy classes of maximal uniquely parabolic subgroups of H. Given $D \in \mathcal{D}$, there exists $P_D \in \mathcal{P}$ and $c_D \in G$ so that $D \leq c_D P_D c_D^{-1}$. We fix such c_D, and suppose that c_D is a shortest possible choice. We abuse notation slightly and write $(H, \mathcal{D}) \leq (G, \mathcal{P})$. Let Y be a combinatorial cusped space for the pair $(H, \mathcal{D})$. The inclusion $\iota\colon H \hookrightarrow G$ extends to an H–equivariant Lipschitz map $\check{\iota}\colon Y^{(0)} \to X$ as follows:

A vertex in a horoball of Y is determined by a triple (sD, h, n) where $s \in H$, $D \in \mathcal{D}$ and $n \in \mathbb{N}$. We define

$$\check{\iota}(sD, h, n) = (sc_D P_D, hc_D, n).$$

It follows from [3, Lemma 3.1] that $\check{\iota}$ is H–equivariant and α–Lipschitz for some α. We refer to $\check{\iota}$ as the *induced map on cusped spaces*. Whenever we have a pair $(H, \mathcal{D}) \leq (G, \mathcal{P})$ as above, we fix the subgroups $P_D \in \mathcal{P}$ and the (shortest) elements c_D as above.

Definition 2.9. Suppose that $(G, \mathcal{P})$ is relatively hyperbolic and that $H \leq G$ is a subgroup. Suppose that $\mathcal{D}$ consists of a set of representatives of H–conjugacy classes of maximal uniquely parabolic subgroups of H. Then $(H, \mathcal{D})$ is *relatively quasiconvex* in $(G, \mathcal{P})$ if the image of the 0–skeleton of the cusped space of $(H, \mathcal{D})$ in the cusped space of $(G, \mathcal{P})$ is λ–quasiconvex for some λ. In this case we say that λ is a *quasiconvexity constant for $(H, \mathcal{D})$ in $(G, \mathcal{P})$*.

This definition is slightly different than the one in [3], since we do not assume that $(H, \mathcal{D})$ is relatively hyperbolic. However, we do assume that $\mathcal{D}$ consists of maximal uniquely parabolic subgroups of H. If the image of the cusped space of $(H, \mathcal{D})$ in the cusped space of $(G, \mathcal{P})$ is quasiconvex, then it follows from the proof of [18, Theorem A.10] that H is relatively quasiconvex in the sense of Hruska [15], and hence that $(H, \mathcal{D})$ is relatively hyperbolic. Therefore, this definition is equivalent to others in the literature, by the results in [18, Appendix A].

3. H-wide fillings

Definition 3.1. Let P be a group, $B \leq P$ a subgroup, and S a finite set. A normal subgroup $N \trianglelefteq P$ is (B, S)–*wide in P* if whenever there are $b \in B$ and $s \in S$ so that $bs \in N$ we have $s \in B$.

Definition 3.2. Let $(G, \mathcal{P})$ be relatively hyperbolic and let $(H, \mathcal{D}) \leq (G, \mathcal{P})$ be relatively quasiconvex. Let $S \subseteq \bigcup \mathcal{P} \smallsetminus \{1\}$. A filling

$$G \to \overline{G} = G(\mathcal{N})$$

is (H, S)–*wide* if for any $D \in \mathcal{D}$ (with $D \leq P_D^{c_D}$ as above) the normal subgroup N_D is $\left(D^{c_D^{-1}}, S \cap P_D \right)$–wide in P_D. (To simplify notation, for $D \in \mathcal{D}$, we write N_D for N_{P_D}.)

Since it is possible that $P_{D_1} = P_{D_2}$ for $D_1 \neq D_2$, it is also possible that $N_{D_1} = N_{D_2}$. We also remark that N_D need not be a subgroup of H.

In place of the statement in Definition 3.2 above, we sometimes use the equivalent formulation that for any $D \in \mathcal{D}$ (with $D \leq P_D^{c_D}$ as above), any $d \in D$ and any $w \in S \cap P_D$, if $dc_D w c_D^{-1} \in N_D^{c_D}$, then $c_D w c_D^{-1} \in D$.

Definition 3.3. We say that a property P holds *for all sufficiently long and H-wide fillings* if there is a finite set $S \subseteq \bigcup \mathcal{P} \smallsetminus \{1\}$ so that P holds for any (H, S)–wide filling $G \to G(\mathcal{N})$ for which $N \cap S = \emptyset$ for each $N \in \mathcal{N}$.

Remark 3.4. In the definition of (H, S)–wide, one should think of S containing all nontrivial elements of $\bigcup \mathcal{P}$ in a large ball around the identity. This ensures that, for each $D \in \mathcal{D}$, a "big neighborhood" of $\overline{D} = D/(N_D^{c_D} \cap D)$ embeds in $\overline{P}_D^{c_D} = (P_D/N_D)^{c_D}$, ruling out behavior like that pictured in Figure 1.
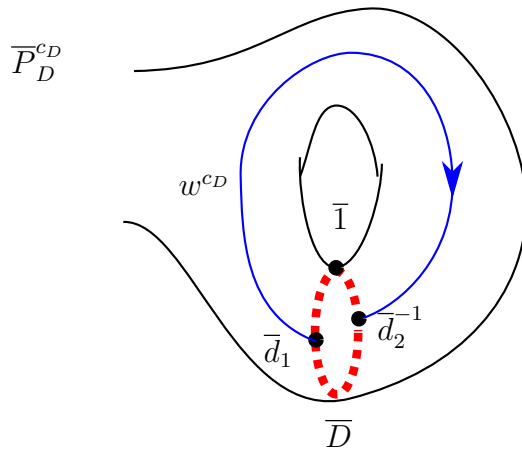

FIGURE 1. A cartoon of the coset graph for $N_D^{c_D}$ in $P_D^{c_D}$ and the kind of loop forbidden by (H, S)–wideness with $w \in S$.

Previous quasiconvex Dehn filling results [2, 3, 18] have been in terms of "H–fillings", whose definition we now recall.

Definition 3.5. Let $(G, \mathcal{P})$ be relatively hyperbolic, let $H < G$ be relatively quasiconvex, and let $\mathcal{N} = \{N_P\}_{P \in \mathcal{P}}$ be a collection of filling kernels. The Dehn filling $G \to \overline{G} = G(\mathcal{N})$ is said to be an H–*filling* if, whenever $\#(P^g \cap H) = \infty$, the kernel N_P^g lies entirely in H.

Remark 3.6. In [3] the condition '$P^g \cap H \neq \{1\}$' was used instead of '$\#(P^g \cap H) = \infty$.' As explained in [18], the formulation in Definition 3.5 is the correct one if there is torsion, and this is the definition that is used in [2, 18].

The following result shows that, at least for long fillings, the notion of H–wide filling generalizes that of H–filling.

Lemma 3.7. *Let $(G, \mathcal{P})$ be relatively hyperbolic, and let $H < G$ be relatively quasiconvex. For any finite $S \subset G$ any sufficiently long H–filling is (H, S)–wide.*

Proof. Let R be the constant from Proposition 2.5, as applied to H. Let $\mathcal{D}$ be the peripheral structure on H consisting of maximal uniquely parabolic subgroups, and $\{P_D \in \mathcal{P}\}$ and $\{c_D \in G\}$ the elements described before, so that $D < P_D^{c_D}$ for each $D \in \mathcal{D}$. Let $M = \max\{d_X(1, c_D)\} + \max\{d_X(1, w) \mid w \in S\}$. Choose filling kernels $\{N_j \lhd P_j\}$ determining a sufficiently long H–filling so that any geodesic joining 1 to $n \in N_j \setminus \{1\}$ must $(R + M + 2\delta + 2)$–penetrate the horoball stabilized by P_j.

Now suppose that for some $w \in S$ and some $d \in D \in \mathcal{D}$ we have $dw^{c_D} \in N_D^{c_D}$. We must show that $w^{c_D} \in D$. Since $(dw^{c_D})^{c_D^{-1}} \in N_D$, the geodesic from 1 to $(dw^{c_D})^{c_D^{-1}}$ must $(R + M + 2\delta + 2)$–penetrate the horoball stabilized by P_D. In particular, $d_X(1, (dw^{c_D})^{c_D^{-1}})$ must be at least $2R + 2M + 4\delta + 4$. Consider the quadrilateral with vertices $1, c_D, dc_D w, d$. The segment $[c_D, dc_D w]$ is the translate of a geodesic $[1, (dw^{c_D})^{c_D^{-1}}]$ by c_D, so it has length at least $2R + 2M + 4\delta + 4$ and $(R + M + 2\delta + 2)$–penetrates the horoball based on $c_D P_D$. Since the sides $[1, c_D]$ and $[dc_D w, d]$ have length at most M, the side $[1, d]$ must pass within 2δ of $[c_D, dc_D w]$ at its midpoint, which is also its deepest point in the horoball. In particular $[1, d]$ must R–penetrate the horoball based on $c_D P_D$. Since $d \in H$, Proposition 2.5 implies that $H \cap P_D^{c_D} = D$ is infinite. Since the filling kernels $\mathcal{N}$ determine an H–filling, this implies that $N_D < D$, and in particular, the element $dw^{c_D} \in D$. It immediately follows that $w^{c_D} \in D$, so we have established that the filling is (H, S)–wide. $\square$

In any case, if $(H, \mathcal{D})$ is a relatively quasiconvex subgroup of the relatively hyperbolic pair $(G, \mathcal{P})$, any Dehn filling of $(G, \mathcal{P})$ induces a Dehn filling of $(H, \mathcal{D})$, which may or may not inject into the filling of G.

Definition 3.8. Let $(G, \mathcal{P})$ be relatively hyperbolic, and let $H < G$ be relatively quasiconvex. Let $\mathcal{D}$ be the canonical (uniquely parabolic) peripheral structure on H, so each $D \in \mathcal{D}$ is contained in some $P_D^{c_D}$ for a unique $P_D \in \mathcal{P}$, and some shortest c_D. Let $\mathcal{N} = \{N_P\}_{P \in \mathcal{P}}$ be a collection of filling kernels for $(G, \mathcal{P})$. The *induced filling kernels* for $(H, \mathcal{D})$ are the collection $\mathcal{N}_H = \{N_{P_D}^{c_D} \cap D\}_{D \in \mathcal{D}}$. These define the *induced filling*

$$(H, \mathcal{D}) \longrightarrow (H(\mathcal{N}_H), \overline{\mathcal{D}}),$$

where $\overline{\mathcal{D}}$ consists of the images of the elements of $\mathcal{D}$ in $H(\mathcal{N}_H) = H/\langle\!\langle \bigcup \mathcal{N}_H \rangle\!\rangle_H$.

There is a natural map from $H(\mathcal{N}_H)$ to the filling $G(\mathcal{N})$.

4. Properties of H–wide fillings

In this section we prove various results which imply that a relatively quasiconvex subgroup H can be controlled in H–wide fillings. These results should be compared to those in [3, Section 4], where analogous results are proved for the behavior of a full relatively quasiconvex subgroup H under sufficiently long H–fillings.

Let $(G, \mathcal{P})$ be relatively hyperbolic. According to Proposition 2.2, there exists a constant δ so that the cusped space for G is δ–hyperbolic, and moreover the induced cusped spaces for sufficiently long fillings of $(G, \mathcal{P})$ are also δ–hyperbolic. In this section, we assume that δ is such a constant, and that all fillings we perform are long enough so that the cusped spaces of the filled groups are δ–hyperbolic.

The following lemma is a reformulation of [3, Lemma 4.1].

Lemma 4.1. *Suppose $(G, \mathcal{P})$ is relatively hyperbolic, and that $L_1, L_2 \geq 10\delta$. For sufficiently long fillings $\pi \colon G \to \overline{G} = G/K$ with induced map between cusped spaces $\pi \colon X \to \overline{X}$, and any geodesic γ in X either:*

(1) *There is a 10δ–local geodesic in $\overline{X}$ between the endpoints of $\pi(\gamma)$ which lies in a 2–neighborhood of $\pi(\gamma)$ and agrees with $\pi(\gamma)$ in the L_1–neighborhood of the Cayley graph in $\overline{X}$; or*

(2) *There is a horoball A in X so that γ L_2–penetrates A in a segment $[x, y]$ with $x, y \in G$, and there is some $k \in K$ stabilizing A so that $d_X(x, k.y) \leq 2L_1 + 3$.*

The following result is very similar to [3, Lemma 4.2] but for H–wide fillings rather than H–fillings. The *induced filling* is defined above in Definition 3.8.

Lemma 4.2. *Suppose that $(G, \mathcal{P})$ is relatively hyperbolic and that $H \leq G$ is relatively quasiconvex. Let R be the constant from Proposition 2.5. For any $L_1 \geq 10\delta$, $L_2 > \max\{2L_1 + 3, R\}$, and all sufficiently long and H–wide fillings $\pi \colon G \to \overline{G}$, the following holds: suppose that $K_H \leq \ker(\pi) \cap H$ is the kernel of the induced filling of H, that $h \in H$ and that γ is a geodesic from 1 to h. If conclusion (2) of Lemma 4.1 holds then there exists $k \in K_H$ so that $d_X(1, kh) < d_X(1, h)$.*

Proof. Let $\mathcal{D}$ be a collection of representatives of H–conjugacy classes of maximal uniquely parabolic subgroups of H, so that $(H, \mathcal{D})$ is relatively quasiconvex in $(G, \mathcal{P})$.

Let γ be a geodesic as in the statement of the lemma, and suppose that conclusion (2) of Lemma 4.1 holds. Accordingly there is some horoball A which is L_2–penetrated by γ. Let gP be the coset on which A is based. According to Proposition 2.5, $H \cap P^g$ is infinite. This implies that there are $r \in H$, and $D \in \mathcal{D}$, so that $P = P_D$ and $gP = rc_D P_D$. The intersection of γ with A is the segment $[x, y]$, and there is an element $k \in N_D^{rc_D}$ so that $d_X(x, k.y) \leq 2L_1 + 3$.

Now, by quasiconvexity of $(H, \mathcal{D})$, there exists some $d_1, d_2 \in D$ so that $d_X(x, rd_1c_D), d_X(y, rd_2c_D)$ are both bounded by some constant L depending only on the quasiconvexity constant for $(H, \mathcal{D})$.

Let $w_1 = (rd_1c_D)^{-1}ky$ and $w_2 = y^{-1}rd_2c_D$. Both $d_X(1, w_1)$ and $d_X(1, w_2)$ are at most $L + 2L_1 + 3$, and both w_1 and w_2 lie in P_D. Let S be the set of words in the parabolic subgroups of X–length at most $2(L + 2L_1 + 3)$. Since $k \in N_D^{rc_D}$, we can find $n \in N_D$ so that $k = rc_D n c_D^{-1} r^{-1}$.

We have
$$
\begin{aligned}
k &= ky.y^{-1} \\
&= rd_1c_D w_1 w_2 c_D^{-1} d_2^{-1} r^{-1}.
\end{aligned}
$$

Therefore,
$$
c_D n c_D^{-1} = d_1 c_D w_1 w_2 c_D^{-1} d_2^{-1},
$$

and
$$
(d_2^{-1}c_D)n(d_2^{-1}c_D)^{-1} = (d_2^{-1}d_1)c_D w_1 w_2 c_D^{-1}.
$$

However, for an (H, S)–wide filling, there can only be an element of N_D of this form if $c_D w_1 w_2 c_D^{-1} \in D$. This implies that $c_D n c_D^{-1} \in D$, which implies that $k \in D^r \cap N_D^{c_D} \leq K_H$.

Since $d_X(x, ky) \leq 2L_1 + 3$, but $d_X(x, y) \geq 2L_2 > 2L_1 + 3$, it is clear that $d_X(1, k.h) < d_X(1, h)$, as required. □

The following result is an immediate consequence.

Corollary 4.3. *Suppose that $(G, \mathcal{P})$ is relatively hyperbolic and that $H \leq G$ is relatively quasiconvex. For any $L \geq 10\delta$ and for all sufficiently long and H–wide fillings $\pi\colon G \to \overline{G}$, if $h \in H$ is the shortest element of $H \cap \pi^{-1}(\pi(h))$ and γ is a geodesic from 1 to h then there is a 10δ–local geodesic in $\overline{X}$ with the same endpoints as $\pi(\gamma)$ which lies in a 2–neighborhood of $\pi(\gamma)$ and agrees with $\pi(\gamma)$ in an L–neighborhood of the Cayley graph of $\overline{G}$ in $\overline{X}$.*

Recall that in a δ–hyperbolic space, 10δ–local geodesics are quite close to geodesics. In particular, we have the following (see [5, III.H.1.13] for a more general and precise statement):

Lemma 4.4. *Let γ be a 10δ–local geodesic in a δ–hyperbolic space. Then γ is a $(7/3, 2\delta)$–quasigeodesic, and is Hausdorff distance at most 3δ from any geodesic with the same endpoints.*

Lemma 4.2, and its interpretation in the form of Corollary 4.3 are the key results needed to generalize many results about H–fillings to sufficiently long and H–wide fillings, as we now explain.

Proposition 4.5 (cf. Proposition 4.3, [3])**.** *Let $(G, \mathcal{P})$ be relatively hyperbolic and suppose that H is a relatively quasiconvex subgroup of $(G, \mathcal{P})$, with relative quasiconvexity constant λ. There exists $\lambda' = \lambda'(\lambda, \delta)$ so that for all sufficiently long and H–wide fillings $\pi : G \to \overline{G}$ the subgroup $\pi(H)$ is λ'–relatively quasiconvex in $\overline{G}$.*

Proof. Recall that at the beginning of the section we fixed a constant δ so that the cusped space X for $(G, \mathcal{P})$ is δ–hyperbolic and that for sufficiently long fillings $\pi\colon G \to \overline{G}$ the cusped space $\overline{X}$ for $(\overline{G}, \overline{P})$ is also δ–hyperbolic. Suppose that H is λ–relatively quasiconvex.

Let $\overline{h} \in \pi(H)$ and suppose that $h \in H$ is the shortest element of H so that $\pi(h) = \overline{h}$. Let γ be a geodesic from 1 to h in X. By Corollary 4.3 with $L = 10\delta$, for sufficiently long and H–wide fillings there is a 10δ–local geodesic in $\overline{X}$ from $\overline{1}$ to $\overline{h}$ which lies in a 2–neighborhood of $\pi(\gamma)$ and agrees with $\pi(\gamma)$ in a 10δ–neighborhood of the Cayley graph.

By Lemma 4.4, any geodesic from $\overline{1}$ to $\overline{h}$ is contained in an $(3\delta + 2)$–neighborhood of $\pi(\gamma)$, and thus within a $(\lambda + 3\delta + 2)$–neighborhood of the image of the cusped space of H in $\overline{X}$. This suffices to prove the result, as in the proof of [3, Proposition 4.3]. (All that remains is to consider geodesics between points in the image of the cusped space of

H which do not lie at depth 0, and it is straightforward to deal with these points given what has already been proved.) □

Proposition 4.6 (cf. Proposition 4.4, [3]). *Let $H \leq G$ be relatively quasiconvex. For sufficiently long and H–wide fillings $\pi \colon G \to \overline{G}$ the map from the induced filling of H to $\overline{G}$ is injective.*

Proof. Let X be the cusped space for G and $\overline{X}$ the cusped space for $\overline{G}$. Suppose that $h \in H \cap \ker(\pi)$ is nontrivial. Let K_H be the kernel of the induced filling on H. We must show that $h \in K_H$.

Let γ be a geodesic in X from 1 to h, and note that $\pi(\gamma)$ is a loop. Suppose that condition (1) from Lemma 4.1 holds. Then there is a nontrivial 10δ–local geodesic loop based at 1 in $\overline{X}$ agreeing with $\pi(\gamma)$ in a 10δ–neighborhood of $1 \in \overline{X}$. This is impossible.

Therefore, Lemma 4.2 applies, and there is an element $k \in K_H$ so that $d_X(1, kh) < d_X(1, h)$. Induction on the length of h shows that $h \in K_H$, as required. □

Proposition 4.7 (cf. Proposition 4.5, [3]). *Let $H \leq G$ be relatively quasiconvex and suppose that $g \in G \setminus H$. For sufficiently long and H–wide fillings $\pi \colon G \to \overline{G}$ we have $\pi(g) \notin \pi(H)$.*

Proof. Choose $L_1 = 3d_X(1, g) + 10\delta$ and any $L_2 > \max\{2L_1 + 3, R\}$, and suppose that π is sufficiently long and H–wide that Lemmas 4.1 and 4.2 hold for π, and also so that π induces a bijection between the ball of radius L_1 about 1 in X and the ball of radius L_1 about the image of 1 in the cusped space of $\pi(G)$.

In order to obtain a contradiction, suppose that $\pi(g) \in \pi(H)$, and choose $h \in H \cap \pi^{-1}(\pi(g))$ with $d_X(1, h)$ minimal. Let γ be a geodesic from 1 to h in X, and let σ be a geodesic from 1 to g in X. Note that $\pi(\sigma)$ is a geodesic.

The minimality of h and Lemma 4.2 ensure that condition (1) from Lemma 4.1 holds for γ.

There are now two cases, depending on whether $\pi(\gamma)$ (equivalently γ) leaves the L_1–neighborhood of the Cayley graph. If γ lies in the L_1–neighborhood of the Cayley graph, it is a 10δ–local geodesic joining $\overline{1}$ to $\overline{g}$. Its length is therefore at most $\frac{7}{3}d_X(1, g) + 2\delta < L_1$, by Lemma 4.4. But since π is injective on the L_1–ball about 1, this implies $g = h$, a contradiction.

The second case is that $\pi(\gamma)$ leaves the L_1–neighborhood of the Cayley graph, in which case there is a 10δ–local geodesic as in Lemma 4.1, joining $\overline{1}$ to $\overline{g}$, which coincides with $\pi(\gamma)$ in the L_1–neighborhood of the Cayley graph, but may differ elsewhere. The length of this 10δ–local

geodesic is at least $L_1 > \frac{7}{3}d_X(1, g) + 2\delta$, again contradicting Lemma 4.4. $\square$

We finish this section with an example which exhibits the necessity of restricting to H–wide fillings (and not just sufficiently long fillings) in Propositions 4.5, 4.6 and 4.7.

Example 4.8. Let Σ be a genus 2 surface, with $\pi_1\Sigma = F = \langle a, b, c, d \mid abcd(dcba)^{-1}\rangle$. Let ϕ be an automorphism of F induced by a pseudo-Anosov homeomorphism of Σ, so that the mapping torus M_ϕ has fundamental group

$$G = \langle F, t \mid txt^{-1} = \phi(x), \text{ for } x \in F\rangle.$$

By Thurston's geometrization of fiber bundles [23], M_ϕ is a hyperbolic 3–manifold; in particular G is a hyperbolic group. Now extend the centralizer of a (attaching a torus to M_ϕ by gluing its longitude to a loop representing a) to get Γ:

$$\Gamma = \langle G, e \mid [e, a]\rangle.$$

Letting $P = \langle e, a \mid [e, a]\rangle$ we have a relatively hyperbolic pair $(\Gamma, \{P\})$, by [8, Theorem 0.1.(2)].

Let $H = \langle b, c, d, e\rangle < \Gamma$. Then H is a free group on the given generators. This can be seen from the induced action of H on the Bass–Serre tree of the defining graph of groups for Γ, which exhibits H as the free product of the free group $\langle b, c, d\rangle < G$ and the infinite cyclic group $\langle e\rangle$. The subgroup $\langle b, c, d\rangle$ is quasiconvex in G, by a result of Scott and Swarup [26]. It then follows from the argument in the proof of [8, Proposition 4.6] that H is relatively quasiconvex in $(\Gamma, \{P\})$.

For an integer $i > 0$, let $N_i = \langle e^i a^{-1}\rangle \trianglelefteq P$. Taking the sequence of fillings

$$\pi_i \colon \Gamma \to \Gamma_i = \Gamma/\langle\langle N_i \rangle\rangle,$$

gives a cofinal sequence of longer and longer Dehn fillings. For each $i > 0$ the group Γ_i has the following graph of groups decomposition:

$$\Gamma_i = G *_{a = e^i} \langle e\rangle.$$

In particular, G embeds in Γ_i (as a quasiconvex subgroup).

The image $H_i := \pi_i(H)$ of H in Γ_i is an amalgam of F with an infinite cyclic subgroup over a maximal cyclic subgroup of F. Since F is distorted in G, the subgroup H_i is distorted in Γ_i, and hence is not quasiconvex.

Moreover, $a \notin H$ but $\pi_i(a) \in H_i$ for all $i > 0$. Finally, we have $N_i \cap H = \{1\}$, so the induced filling of H is the trivial filling. On the other hand, H_i is not a free group, so the map from the induced

filling of H to Γ_i is not injective for any $i > 0$. Explicitly, the element $(e^i)bcd\left(dcb(e^i)\right)^{-1}$ is in the kernel of the map from the induced filling of H to Γ_i.

5. EXISTENCE OF H–WIDE FILLINGS

In this section, we prove Lemma 5.2 which implies that in our applications in Sections 6 and 7 we can find sufficiently long and H–wide fillings. The key observation is that separability allows us to do this.

Lemma 5.1. *Suppose that P is a group and that B is a separable subgroup. For any finite set S there exists a finite-index normal subgroup $K_S \leq P$ so that for any $N \trianglelefteq P$ with $N \leq K_S$, the subgroup N is (B, S)–wide in P.*

Proof. For each $s \in S \smallsetminus B$, choose some $P_s \leq P$ finite index and satisfying $B < P_s$ and $s \notin P_s$. Let $K_S = \bigcap \{P_s \mid s \in S \smallsetminus B\}$, and note that K_S is finite index in P, and contains B.

Suppose $N \triangleleft P$ is contained in K_S. We verify that N is (B, S)–wide. Let $b \in B$ and $s \in S$, and suppose $bs \in N$. If $s \in B$ there is nothing to show, so suppose $s \notin B$. The element s is not contained in K_S, but $bs \in N < K_S$, so we must have $b \notin K_S$. But this contradicts $B \leq K_S$.

The subgroup K_S just constructed may not be normal, but we may replace K_S by the intersection of its conjugates without disturbing the conclusion. $\square$

In Lemma 5.2 we consider a finite collection $\{(H_1, \mathcal{D}_1), \ldots, (H_k, \mathcal{D}_k)\}$ of relatively quasiconvex subgroups of a relatively hyperbolic pair $(G, \mathcal{P})$. For each i and each $D \in \mathcal{D}_i$ we assume that we have fixed $P_D \in \mathcal{P}$ and $c_D \in G$ so that $D \leq P_D^{c_D}$, and that c_D is a shortest such conjugating element.

Lemma 5.2. *Suppose that $(G, \mathcal{P})$ is relatively hyperbolic, and that $\mathcal{H} = \{(H_1, \mathcal{D}_1), \ldots, (H_k, \mathcal{D}_k)\}$ is a collection of relatively quasiconvex subgroups. Suppose that for each $1 \leq j \leq k$ and for each $D \in \mathcal{D}_j$ the subgroup D is separable in $P_D^{c_D}$.*

Then for any finite $S \subset \bigcup \mathcal{P} \smallsetminus \{1\}$ there exist finite index subgroups $\{K_P \trianglelefteq P \mid P \in \mathcal{P}\}$ so that any filling

$$G \to G(\mathcal{N}), \text{ with } \mathcal{N} = \{N_P \leq K_P \mid P \in \mathcal{P}\}$$

is (H_j, S)–wide for each $1 \leq j \leq k$.

Proof. Fix $S \subset \cup \mathcal{P} \smallsetminus \{1\}$ a finite set.

Fix $P \in \mathcal{P}$ and let $S_P = P \cap S$. Suppose, for some i, that $D \in \mathcal{D}_i$ is so that $P = P_D$. By Lemma 5.1 there is a finite-index normal subgroup

$K_D \trianglelefteq P$ so that any $N \trianglelefteq P^{c_D}$ for which $N \leq K_D^{c_D}$ is (D, S_P)–wide in P^{c_D}.

We choose K_P to be the intersection of all K_D for which $P = P_D$. If we now choose $N_P \leq K_P$ then the conclusion of the lemma holds. This completes the proof. □

6. Application to virtual specialness and virtual fibering

In this section, we explain how the ideas and results in the beginning of the paper, together with a recent result of Cooper and Futer [6] give a proof of Theorem A independent of [28].

The following consequence of Theorem A was reproved (without using the results of [28]) by Cooper and Futer.

Theorem 6.1. [6, Corollary 1.3] *Suppose that G is the fundamental group of a non-compact finite-volume hyperbolic 3–manifold. Then G acts freely and cocompactly on a $CAT(0)$ cube complex dual to finitely many immersed quasi-Fuchsian surfaces.*

In this section, our main result is that this cubulation is virtually special.

Theorem A. *Suppose that G is the fundamental group of a non-compact finite-volume hyperbolic 3–manifold M. Then G is virtually compact special.*

If $\mathcal{P}$ is a collection of conjugacy-representatives of maximal parabolic subgroups of G then $(G, \mathcal{P})$ is relatively hyperbolic. After possibly replacing M by an orientable double-cover of M, each element of $\mathcal{P}$ is free abelian of rank 2. As explained in the proof of Theorem A below, proving Theorem A reduces to establishing separability of certain double cosets of relatively quasiconvex subgroups of G. In the closed case, such double cosets are separable by results in [2] and [20]. We reduce to this case by performing orbifold Dehn filling on M and applying the following 'weak separability' criterion for double cosets.

Proposition 6.2. *Suppose that $(G, \mathcal{P})$ is relatively hyperbolic and that each element of $\mathcal{P}$ is free abelian. Suppose further that $\mathcal{H}$ is a finite collection of relatively quasiconvex subgroups of $(G, \mathcal{P})$ and that $S \subseteq (\bigcup \mathcal{P}) \smallsetminus \{1\}$ and $F \subseteq G$ are finite subsets.*

There exist finite-index subgroups $\{K_P \trianglelefteq P \mid P \in \mathcal{P}\}$ so that for any subgroups $N_P \leq K_P$ the filling

$$G \to G/K := G\left(\{N_P \mid P \in \mathcal{P}\}\right),$$

is (H, S)–wide for each $H \in \mathcal{H}$ and furthermore whenever $f \in F$, $\Psi, \Theta \in \mathcal{H}$ satisfy $1 \notin \Psi\Theta f$, there is no element of K in $\Psi\Theta f$.

Proof. By Lemma 5.2 and the fact that all subgroups of finitely generated abelian groups are separable, there exist finite-index subgroups $K_P \trianglelefteq P$ so that if $N_P \leq K_P$ then the filling is (H, S)–wide for each H. Below, we find other finite-index subgroups $\hat{K}_P \trianglelefteq P$ so that if $N_P < \hat{K}_P$ then the condition on double cosets holds. We then choose $N_P \leq K_P \cap \hat{K}_P$. For the remainder of the proof we concentrate on finding the subgroups $\hat{K}_P$.

Let X be the cusped space for $(G, \mathcal{P})$ and suppose that X is δ–hyperbolic. We further assume that δ is chosen so that the cusped spaces of all sufficiently long fillings are δ–hyperbolic. We suppose that $\delta \geq 1$. Let λ be a quasiconvexity constant which works for every element in $\mathcal{H}$. Finally, let $M = \max\{d_X(1, f) \mid f \in F\}$.

Fix $f \in F$ and $\Psi, \Theta \in \mathcal{H}$ so that $1 \notin \Psi\Theta f$, and consider the equation $k \in \Psi\Theta f$ for elements k of the kernel of a filling. After finding conditions on the filling which ensure there is no such element, we consider a filling appropriate for all $f \in F$ simultaneously.

Let $\mathcal{D}$ be a collection of representatives of Ψ–conjugacy classes of maximal uniquely parabolic subgroups of Ψ. For $D \in \mathcal{D}$, we have $D \leq P_D^{c_D}$ for some $P_D \in \mathcal{P}$ and some (shortest) $c_D \in G$. Similarly, let $\mathcal{E}$ be a collection of representatives of Θ–conjugacy classes of maximal parabolic subgroups of Θ, and for $E \in \mathcal{E}$ we have $E \leq P_E^{d_E}$ for some $P_E \in \mathcal{P}$ and some (shortest) $d_E \in G$. Let X_Ψ be the cusped space for the pair $(\Psi, \mathcal{D})$ and let X_Θ be the cusped space for $(\Theta, \mathcal{E})$ (both with respect to some choices of generating sets). Let $\check{\iota}_\Psi \colon X_\Psi \to X$ and $\check{\iota}_\Theta \colon X_\Theta \to X$ be the induced maps of cusped spaces, and note that $\check{\iota}_\Psi(X_\Psi^{(0)})$ and $\check{\iota}_\Theta(X_\Theta^{(0)})$ are both λ–quasiconvex subsets of X.

In order to apply the results from Section 4, choose $L_1 = 10\delta$ and $L_2 = \max\{20\delta + M + \lambda + 4, R_\Psi, R_\Theta\}$, where R_Ψ and R_Θ are the constants from Proposition 2.5 applied to Ψ and Θ, respectively.

For $P \in \mathcal{P}$, let

$$S_P \supseteq \{p \in P \mid d_X(1, p) \leq 32\delta + 2M + 4\lambda + 2L_1 + 3\}$$

be a finite set which is large enough so that for all $H \in \mathcal{H}$ the (H, S)–wideness condition of Lemma 4.2 is satisfied with L_1 and L_2 as above. Consider the collection of subgroups of P of the form $P_{B_1, B_2} = \langle B_1, B_2 \rangle$ where $B_1 = P \cap \Psi^{g_1}$ for some $g_1 \in G$ and $B_2 = P \cap \Theta^{g_2}$ for some $g_2 \in G$. There are finitely many such pairs of subgroups of P. For each such pair (B_1, B_2), by Lemma 5.1 there exists a finite-index $\hat{K}_{B_1, B_2} \triangleleft P$ which is (P_{B_1, B_2}, S_P)–wide. We define $\hat{K}_P = \bigcap \hat{K}_{B_1, B_2}$ and check the condition on double cosets.

Choose $N_P \le \hat{K}_P$, and consider the filling

$$G \to G\left(\{N_P \mid P \in \mathcal{P}\}\right) = G/K.$$

In order to obtain a contradiction suppose that there is an element $g \in K$, and elements $f \in F$, $\psi \in \Psi$ and $\theta \in \Theta$ so that

$$g = \psi\theta f.$$

Choose a g so that $d_X(1, g)$ is minimal amongst all choices of g for which there is such an expression.

Consider a geodesic quadrilateral in X with vertices $1, \psi, \psi\theta, g$, and let ξ_1 be the geodesic from 1 to ψ, ξ_2 the geodesic from ψ to $\psi\theta$, η the geodesic from $\psi\theta$ to g and ρ the geodesic from 1 to g, respectively. By assumption, we know that $g \ne 1$.

Let $\pi\colon X \to \overline{X}$ be the map on cusped spaces induced by the filling map $\pi\colon G \to G/K$. Since $g \in K \smallsetminus \{1\}$ the image of ρ in $\overline{X}$ is a loop, so condition (1) from Lemma 4.1 cannot hold. This means that condition (2) from Lemma 4.1 holds. Let A be a horoball L_2–penetrated by ρ, so ρ meets A in a segment $[x, y]$, and let $k \in K \cap \mathrm{Stab}(A)$ be so that $d_X(x, k.y) \le 2L_1 + 3$. It is straightforward to see that $d_X(1, k.g) < d_X(1, g)$, since $d_X(x, k.y) \le 2L_1 + 3$ but $d_X(x, y) \ge 2L_2 > 2L_1 + 4$. We arrive at a contradiction by showing that $k.g \in \Psi\Theta f$, contradicting the choice of g as a shortest element of K with such an expression.

Without loss of generality, the subsegment of ρ between x and y is a geodesic through A which consists of a vertical segment down from x, a horizontal segment of at most 3 edges, and then a vertical segment terminating at y (see [11, Lemma 3.10]). Let x' be the point on this geodesic directly below x at depth $3\delta + M + \lambda$ and let y' be the point directly below y at depth $3\delta + M + \lambda$. The quadrilateral $\xi_1 \cup \xi_2 \cup \eta \cup \rho$ is 2δ–slim, so there are points on $\eta \cup \xi_1 \cup \xi_2$ within 2δ of x' and of y'. Because η is a geodesic (of length at most M) joining two points at depth 0, no point on η can be within 2δ of either x' or y'. Therefore, there are points on $\xi_1 \cup \xi_2$ within 2δ of x' and of y'.

The geodesic ξ_1 travels between two points in Ψ and ξ_2 joins two points in $\psi\Theta$. By quasiconvexity, any point on ξ_1 lies within λ of a point in $\check{\iota}_\Psi(X_\Psi)$ and any point on ξ_2 lies within λ of a point in $\psi.\check{\iota}_\Theta(X_\Theta)$. Let u_0 and v_0 be points in $\check{\iota}_\Psi(X_\Psi) \cup \psi.\check{\iota}_\Theta(X_\Theta)$ lying within distance $2\delta + \lambda$ of x' and y' respectively. Note that u_0 and v_0 lie in the horoball A.

There are points at depth more than 0 in a horoball A which lie in $\check{\iota}_\Psi(X_\Psi)$ exactly when they are of the form $(sc_D P_D, hc_D, n)$ for some $s \in \Psi$, $h \in sD$ and $n \in \mathbb{N}$, where $D \in \mathcal{D}$ is so that $D \le P_D^{c_D}$, as above, and A is the horoball based on the coset $sc_D P_D$.

Similarly, there are points at depth more than 0 in A which lie in $\psi.\check{\imath}_\Theta(X_\Theta)$ exactly when they are of the form $(\psi.td_E P_E, \psi.gc_D, m)$ for $t \in \Theta$, $g \in tE$ and $m \in \mathbb{N}$, where $E \in \mathcal{E}$ and $E \leq P_E^{d_E}$, and A is the horoball based on $\psi.td_E P_E$.

The points u_0, v_0 have one of these forms, and they are at distance at most $2\delta + \lambda$ from x' and y' respectively, which implies that the appropriate n or m is at most $3\delta + M + 2\lambda$. Thus, there are points u, v at depth 0 in A, directly above u_0 and v_0 respectively, so that $d_X(u, x), d_X(v, y) \leq \alpha := 10\delta + 2M + 4\lambda$. All of the points in A directly above u_0 lie in $\check{\imath}_\Psi(X_\Psi)$ or $\psi.\check{\imath}_\Theta(X_\Theta)$, except possibly the point u at depth 0. This point u will not lie in $\check{\imath}_\Psi(X_\Psi)$ unless $c_D = 1$, and similarly for $\check{\imath}_\Theta(X_\Theta)$. However, certainly u lies within distance 1 of $\check{\imath}_\Psi(X_\Psi)$ or $\psi.\check{\imath}_\Theta(X_\Theta)$. Similarly, v lies within distance 1 of $\check{\imath}_\Psi(X_\Psi)$ or $\psi.\check{\imath}_\Theta(X_\Theta)$.

We deal with four cases, depending on whether each of u_0 and v_0 are contained in $\check{\imath}_\Psi(X_\Psi)$ or $\psi.\check{\imath}_\Theta(X_\Theta)$.

Case 1: Both u_0 and v_0 are contained in $\check{\imath}_\Psi(X_\Psi)$.

(The case where they are both contained in $\psi.\check{\imath}_\Theta(X_\Theta)$ is entirely similar and we omit it.)

In this case, if $u_0 = (sc_D P_D, \psi_u c_D, n)$ then $u = \psi_u c_D$, where $\psi_u \in \Psi$ and A is the horoball based on $sc_D P_D$. For ease of notation we write $P = P_D$ and $c = c_D$, and so we have $\psi_u c \in scP$.

Note that $v = \psi_v c \in scP$ also, for some $\psi_v \in \Psi$. We have $u^{-1}v = c^{-1}(\psi_u^{-1}\psi_v)c \in D^{c^{-1}}$.

Now, $d_X(x, k.y) \leq 2L_1 + 3$, and we also have $d_X(x, u), d_X(y, v) \leq \alpha$, from which it follows that $d_X(u, k.v) \leq 2\alpha + 2L_1 + 3$. Write $w = v^{-1}k^{-1}u$, a group element of X–length at most $2\alpha + 2L_1 + 3$ and note that $u^{-1}k^{-1}u$ is in the filling kernel $N_P \lhd P$, and so $cu^{-1}k^{-1}uc^{-1}$ is contained in N_P^c. On the other hand, we also have

$$
\begin{aligned}
cu^{-1}k^{-1}uc^{-1} &= cu^{-1}v(v^{-1}k^{-1}u)c^{-1} \\
&= cu^{-1}vwc^{-1} \\
&= c(c^{-1}\psi_u^{-1}\psi_v c)wc^{-1} \\
&= (\psi_u^{-1}\psi_v)cwc^{-1}.
\end{aligned}
$$

Note that $w \in S_P$, and that the filling is (Ψ, S_P)–wide. Since $(\psi_u^{-1}\psi_v) \in D$, this implies that $cwc^{-1} \in D$, so $cu^{-1}k^{-1}uc^{-1} = (\psi_u^{-1}\psi_v)cwc^{-1} \in D$. Therefore, $\psi_u^{-1}k\psi_u = cu^{-1}kuc^{-1} \in D$, which means that

$$
k.\psi = k\psi_u(\psi_u^{-1}\psi) = \psi_u(\psi_u^{-1}k\psi_u)\psi_u^{-1}\psi \in \Psi.
$$

Therefore,

$$
k.g = (k.\psi)\,\theta f,
$$

gives an expression for $k.g$ as an element of $\Psi\Theta f$, contradicting the fact that g was the shortest element of K with such an expression.

Case 2: u_0 is contained in $\breve{\iota}_\Psi(X_\Psi)$ and v_0 contained in $\psi.\breve{\iota}_\Theta(X_\Theta)$.

We can write $u_0 = (sc_D P_D, \psi_u c_D, n)$ and $v_0 = (\psi.td_E P_E, \psi.r_v d_E, m)$, so $u = \psi_u c_D$ and $v = \psi.\theta_v d_E$, where $\psi_u \in \Psi$, $\theta_v \in \Theta$, $D \le P_D^{c_D}$ and $E \le P_E^{d_E}$. We clearly have $P_E = P_D$, which we write as P. We write $c = c_D$ and $d = d_E$, and note that A is the horoball based on the coset $scP = \psi td P$.

We still have $d_X(u, x), d_X(v, y) \le \alpha$. The geodesic ξ_1 intersects A in a segment $[g_1, h_1]$, where the entrance point g_1 is within 4δ of x, and within α of u. The exit point h_1, we may similarly argue, is within α of some group element $w = \psi_w c$ in the coset scP, with $\psi_w \in \Psi$. Likewise, the geodesic ξ_2 intersects the horoball A in a segment $[g_2, h_2]$, where $d_X(g_2, h_1) \le 4\delta$, and there is another point $z = \psi\theta_z d$ in scP with $\theta_z \in \Theta$, and satisfying $d_X(z, g_2) \le \alpha$. See Figure 2.
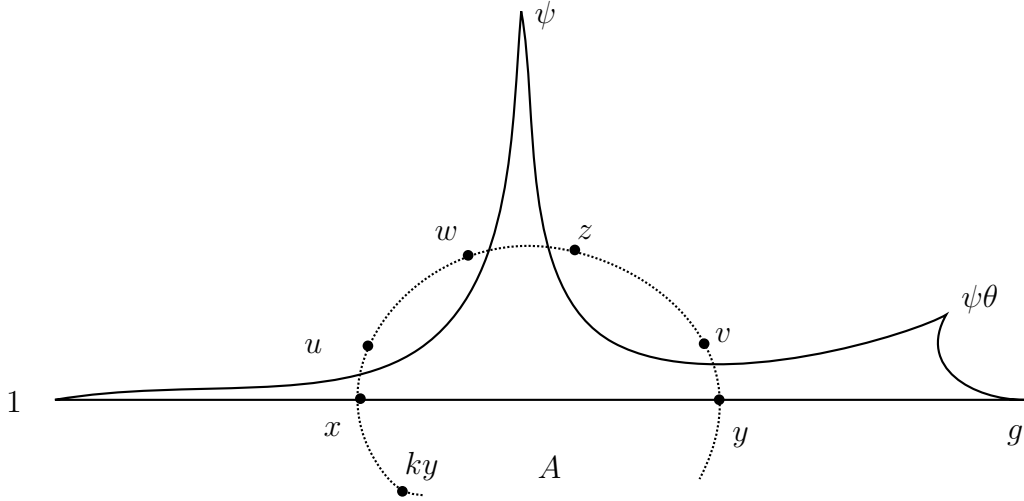

FIGURE 2. Case 2. The dotted line represents the coset $scP = \psi td P$.

We have $u^{-1}w \in D^{c^{-1}}$ and $z^{-1}v \in E^{d^{-1}}$. Both $D^{c^{-1}}$ and $E^{d^{-1}}$ are subgroups of P.

Let $B_1 = D^{c^{-1}}$ and $B_2 = E^{d^{-1}}$ so $u^{-1}wz^{-1}v \in P_{B_1,B_2}$. Now,

$$\begin{aligned} u^{-1}k^{-1}u &= (u^{-1}w)(w^{-1}z)(z^{-1}v)(v^{-1}k^{-1}u) \\ &= (u^{-1}w)(z^{-1}v)\left((w^{-1}z)(v^{-1}k^{-1}u)\right). \end{aligned}$$

(Note that we use here that P is abelian.)

Since $w^{-1}z$ has length at most $2\alpha + 4\delta$ and $v^{-1}k^{-1}u$ has length at most $2\alpha + 2L_1 + 3$, so the last of the three terms above is in S_P and

we can apply the (P_{B_1,B_2}, S_P)–wideness of the kernel to deduce that $u^{-1}ku \in P_{B_1,B_2} = B_1B_2$. Choose elements $b_1 \in B_1$ and $b_2 \in B_2$ so that $u^{-1}ku = b_1b_2$. We now have

$$
\begin{aligned}
k\psi\theta &= u(u^{-1}ku)(u^{-1}w)(w^{-1}z)(z^{-1}v)(v^{-1}\psi\theta) \\
&= u(b_1b_2)(u^{-1}w)(w^{-1}z)(z^{-1}v)(v^{-1}\psi\theta) \\
&= u\left(b_1(u^{-1}w)\right)(w^{-1}z)\left((z^{-1}v)b_2\right)(v^{-1}\psi\theta) \\
&= u\left(b_1(u^{-1}w)\right)(w^{-1}\psi)(\psi^{-1}z)\left((z^{-1}v)b_2\right)(v^{-1}\psi\theta) \\
&= \psi_u\left(c(b_1u^{-1}w)c^{-1}\right)(cw^{-1}\psi)(\psi^{-1}z)\left((z^{-1}v)b_2\right)(v^{-1}\psi\theta) \\
&= \left(\psi_u\left(c(b_1u^{-1}w)c^{-1}\right)(\psi_w^{-1}\psi)\right)\left(\theta_z\left(d(z^{-1}vb_2)d^{-1}\right)(\theta_v^{-1}\theta)\right)
\end{aligned}
$$

The first three terms of this expression are in Ψ and the last three terms are in Θ, which proves that $k\psi\theta \in \Psi\Theta$. Therefore, $k.g = k\psi\theta f \in \Psi\Theta f$. Since we know that $d_X(1, k.g) < d_X(1, g)$, this contradicts the minimality of g, hence proving the result in Case 2.

It remains to note that the case that u_0 is contained in $\psi.\check{\iota}_\Theta(X_\Theta)$ and v_0 is contained in $\check{\iota}_\Psi(X_\Psi)$ essentially becomes Case 1. Indeed, suppose that v_0 is contained in $\check{\iota}_\Psi(X_\Psi)$. Then the geodesic from v to 1 lies near to the geodesic from y to 1, which easily implies (since x lies on the geodesic from y to 1) that x lies near $\check{\iota}_\Psi(X_\Psi)$, as required. This completes the proof of Proposition 6.2. $\qquad\square$

We now turn to the proof of Theorem A.

Proof (of Theorem A). Pass to a finite cover which is orientable, so that all cusps in M have torus cross-sections. It is well known that if $\mathcal{P}$ is a collection of representatives of G–conjugacy classes of maximal cusp subgroups of M then $(G, \mathcal{P})$ is relatively hyperbolic.

By Theorem 6.1, there is a CAT(0) cube complex X upon which G acts freely and cocompactly. By [24, Criterion 2.3] (see also [14, Section 4]), to prove that the action of G on X is virtually special it suffices to prove that for a certain finite list of subgroups Q_i which stabilize hyperplanes in X, the subgroups Q_i and double cosets Q_iQ_j are separable in G. Since G is a Kleinian group, it is LERF by [2, Corollary 9.4], so it remains to prove double coset separability.

The cube complex X built by Cooper and Futer for Theorem 6.1 is built using the Sageev construction [25] (see also [17]). The hyperplane subgroups in G are commensurable to the codimension 1 subgroups, which are quasi-Fuchsian surface subgroups and therefore geometrically finite. It now follows immediately by [15, Corollary 1.3] that the subgroups Q_i are relatively quasiconvex in G.

Suppose now that $h \notin Q_i Q_j$. Equivalently, $1 \notin Q_i Q_j h^{-1}$. By Proposition 4.5, for sufficiently long and Q_i–wide fillings the image of Q_i is relatively quasiconvex, and similarly for Q_j. By Proposition 6.2, there exist finite-index subgroups $\{K_P \trianglelefteq P \mid P \in \mathcal{P}\}$ so that for any choices $\{\gamma_P \in K_P \mid P \in \mathcal{P}\}$, the filling

$$G \to \overline{G} = G\left(\{\gamma_P \mid P \in \mathcal{P}\}\right)$$

is such that the images of Q_i and Q_j are relatively quasiconvex and there is no element of K in $Q_i Q_j h^{-1}$. For such a filling, the image of h is outside the image of $Q_i Q_j$.

Possibly replacing the γ_P by powers (which does not change containment in K_P, and so the above properties continue to hold), the Orbifold Hyperbolic Dehn Surgery Theorem [10, Theorem 5.3], implies that the group $\overline{G}$ is the fundamental group of a compact hyperbolic orbifold, and so is Kleinian and word-hyperbolic. Since it is Kleinian, it is LERF by [2, Corollary 9.4], and since it is also word-hyperbolic [20, Theorem 1.1] implies that all double cosets of quasiconvex subgroups of $\overline{G}$ are separable. Therefore, the image of h can be separated from the image of $Q_i Q_j$ in a finite quotient of $\overline{G}$, which is clearly also a finite quotient of G.

This proves that $Q_i Q_j$ is separable in G, which proves that the G–action on X is virtually special, as required. $\qquad\square$

7. Relative height and relative multiplicity

For quasiconvex subgroups of hyperbolic groups, the *height* is an important invariant. For full relatively quasiconvex subgroups, it remains a useful invariant, but because we cannot control the normalizer in P of an intersection $H \cap P$ when H is (non-full) relatively quasiconvex and P is a maximal parabolic subgroup, height is not always a useful notion as it is too often infinite. Instead, we should consider the relative height, defined as follows.

Definition 7.1. (cf. [16, §1.4]) Suppose that $(G, \mathcal{P})$ is relatively hyperbolic and $H \leq G$. The *relative height* of H in $(G, \mathcal{P})$ is the maximum number $n \geq 0$ so that there are distinct cosets $\{g_1 H, \ldots, g_n H\}$ so that $\bigcap_{i=1}^{n} g_i H g_i^{-1}$ is an infinite non-parabolic subgroup.

In [16], they refer to relative height merely as 'height', but we prefer to keep this term for its traditional meaning.

Remark 7.2. It follows from the classification of groups acting isometrically on δ–hyperbolic spaces that a subgroup of a relatively hyperbolic group is infinite and non-parabolic if and only if it contains a

loxodromic element. We use this equivalent characterization without further mention.

In this section, we prove results for relative height analogous to those proved for height in [2, Appendix A]. Specifically, we define a notion of *relative multiplicity* (see Definition 7.7) and prove in Theorem 7.8 that relative multiplicity is equal to relative height. This gives a new proof of a theorem of Hruska and Wise [16] that the relative height of a relatively quasiconvex subgroup is finite. In Theorem 7.15 we prove that for sufficiently long and H–wide fillings the relative height of a relatively quasiconvex subgroup does not increase under Dehn filling.

The definition of a *weakly geometrically finite* (or *WGF*) action is given in [2, A.27]. We note here that a weakly geometrically finite action differs from the usual notion of a *geometrically finite* action (as in [15, Definition 3.2 (RH-2)]) in allowing horoballs with finite stabilizer.

Fix a relatively quasiconvex subgroup $(H, \mathcal{D})$. Let X_H be a cusped space for $(H, \mathcal{D})$ and $\check{\iota} \colon X_H \to X$ be the extension of the natural inclusion of H into G on the level of cusped spaces, as described in [3, Section 3].

Definition 7.3. Suppose that $(G, \mathcal{P})$ is relatively hyperbolic and that $(H, \mathcal{D})$ is relatively quasiconvex. Let X be a cusped space for $(G, \mathcal{P})$ (considered as containing G as a subset), let $\widetilde{*}$ be the basepoint of X, and let $R \geq 0$. An R–*hull for H acting on X* is a connected H–invariant full sub-graph $\widetilde{Z} \subset X$ so that

(1) $\widetilde{*} \in \widetilde{Z}$;
(2) If γ is a geodesic in X with endpoints in $\Lambda(H)$ then $N_R(\gamma) \cap N_R(G) \subset \widetilde{Z}$; and
(3) If A is any horoball containing a vertex a of depth greater than 0 in the image $\check{\iota}(X_H)$, then $\tilde{Z} \cap A^{(0)}$ contains every vertex of a maximal vertical ray in A containing a.
(4) The action of $(H, \mathcal{D})$ on $\widetilde{Z}$ (with its induced path metric) is WGF.

Remark 7.4. This definition is not the same as [2, Definition A.32] unless H is full relatively quasiconvex (as was assumed in [2]). It is important that we do not include an R–neighborhood of γ, but only that part of the R–neighborhood near the Cayley graph. The third condition in [2, Definition A.32] has similarly been modified. Both of these changes are made so that Lemma 7.12 below is true.

Definition 7.5. Suppose that $(H, \mathcal{D}) \leq (G, \mathcal{P})$ is relatively quasiconvex, and $\check{\iota} \colon X_H \to X$ is the inclusion of cusped spaces as above. For

a positive integer D, the *restricted D–neighborhood of $\breve{\iota}(X_H)$*, denoted $N_D^R(\breve{\iota}(X_H))$, is the full subgraph of X on the vertices of either of the following two types:

(1) Vertices of $N_D(\breve{\iota}(X_H)) \cap N_D(G)$; and

(2) For any horoball A so $\breve{\iota}(X_H)$ contains vertices of arbitrary depth in A, include all vertices $a \in A$ which are connected by a vertical geodesic to a vertex of the first type.

Lemma 7.6 (cf. Lemma A.41, [2]). *Let $R \geq 0$. There exists some D so that the restricted D-neighborhood of $\breve{\iota}(X_H)$ is an R–hull for the action of H on X.*

Proof. We first note that if any of the requirements of an R–hull are satisfied by the restricted D–neighborhood of $\breve{\iota}(X_H)$, then they are satisfied for the restricted D'–neighborhood of $\breve{\iota}(X_H)$, for any $D' \geq D$. It therefore suffices to consider each of the four requirements separately, and take D to be the maximum needed for any of the four.

Condition (1) is satisfied for any D, since $\widetilde{\ast} = 1 \in H$.

Condition (2) is satisfied as soon as $D \geq R + 2\delta + \lambda$, where λ is the quasiconvexity constant for $\breve{\iota}(X_H)$. Indeed, $\Lambda(H) \subset \Lambda(\breve{\iota}(X_H))$, so if γ is a biinfinite geodesic with endpoints in $\Lambda(H)$, it must lie in a $\lambda + 2\delta$–neighborhood of $\breve{\iota}(X_H)$. Suppose $x \in N_R(\gamma) \cap N_R(G)$; we want to show that $x \in N_D^R(\breve{\iota}(X_H))$. Let $z \in \gamma$, $g \in G$ be vertices at distance at most R from x. As we have noted, there is a $q \in \breve{\iota}(X_H)$ satisfying $d_X(z, q) \leq \lambda + 2\delta$. Thus x lies in the $(R + \lambda + 2\delta)$–neighborhood of $\breve{\iota}(X_H)$. If $D \geq R + 2\delta + \lambda$, then $x \in N_D(\breve{\iota}(X_H)) \cap N_D(G)$.

Condition (3) is built in to the definition of restricted D–hull.

Condition (4) (the weak geometric finiteness) follows once we observe that for large enough D, the restricted D–neighborhood is equivariantly quasi-isometric to the D–neighborhood, and either one is quasi-isometrically embedded in X. In particular, the limit set of the restricted D–neighborhood is equivariantly homeomorphic to $\partial X_H = \partial(H, \mathcal{D})$. (Though in general $\breve{\iota}$ is not a quasi-isometric embedding if some $D \in \mathcal{D}$ is very distorted in $P_D^{c_D}$.) $\square$

Let $\widetilde{Z}$ be an R–hull for the action of H on G, and let $Z = {}_H\backslash^{\widetilde{Z}}$. Similarly, let $Y = {}_G\backslash^X$. Then there is a natural map $i\colon Z \to Y$ which induces the inclusion of H into G (in the sense described in [2]).

For $n > 0$, let

$$S_n = \{(z_1, \ldots, z_n) \in Z^n \mid i(z_1) = \cdots = i(z_n)\} \setminus \Delta$$

where $\Delta = \{(z_1, \ldots, z_n) \mid \text{ there exist } i \neq j \text{ so that } z_i = z_j\}$ is the fat diagonal.

Points in S_n have a well-defined *depth* which is the depth of the image in Y We consider components C of S_n which contain a point with depth 0.

As in [2], choosing a maximal tree in Z, and a basepoint p at depth 0, a component C of S_n induces well-defined maps $\tau_{C,i}\colon \pi_1(C,p) \to H$, for $i = 1, \ldots, n$.

Definition 7.7. The *relative multiplicity* of $i\colon Z \to Y$ is the largest n so that S_n contains a component C so that for all $i \in 1, \ldots, n$ the group $\tau_{C,i}(\pi_1(C,p))$ contains a loxodromic element.

Theorem 7.8 (cf. Theorem A.38, [2])**.** *For sufficiently large R, depending only on δ and the quasi-convexity constant of H, if $\widetilde{Z}$ is an R–hull for the action of H on X, and $i : Z \to Y$ is as described above, then the relative height of H in G is equal to the relative multiplicity of $i : Z \to Y$.*

Definition 7.9. A geodesic σ in a combinatorial horoball is *regular* if it has at most three horizontal edges, and these are at the maximum depth for σ. A path in a cusped space $X(G, \mathcal{P})$ is *regular* if every intersection with a horoball is regular.

A path σ in a combinatorial horoball is *super-regular* if it has at most 1 horizontal edge, this edge is at maximum depth for σ, and σ has minimal length among paths with this property. A path in a cusped space $X(G, \mathcal{P})$ is super-regular if every intersection with a horoball is super-regular.

Lemma 7.10. *Let g be a loxodromic element of the relatively hyperbolic group pair $(G, \mathcal{P})$. Then for any $D > 0$, and any sufficiently large $n > 0$, there is a bi-infinite quasigeodesic axis σ for g^n satisfying:*

(1) σ is super-regular;
(2) σ is contained in a $(4\delta + 3)$–neighborhood of any geodesic with the same endpoints.
(3) $d_X(p, g^n p) > D$ for any point $p \in \sigma$.

Proof. Let $g^{\pm\infty}$ be the two limit points in ∂X of the cyclic group $\langle g \rangle$. Since X is proper, there is a bi-infinite geodesic γ joining $g^{\pm\infty}$. Note that $g^n \gamma$ and γ are Hausdorff distance at most 2δ from one another, for any n. Fix n large enough so that $d_X(x, g^n x) > \max\{D, 100\delta\}$ for every point $x \in X$.

Since the endpoints of γ are distinct, γ is not contained in a single horoball. Choose some $h \in \gamma$ in the Cayley graph of G. Choose a regular geodesic α_0 joining h to $g^n h$.

Let σ_0 be the concatenation of the g^n–translates of α_0; namely $\sigma_0 = \bigcup_{i \in \mathbb{Z}} g^{in} \alpha_0$. Let σ be the path obtained my modifying σ_0 to be super-regular. (This means first ensuring that paths within any horoball consist of two vertical segments and a single horizontal segment, and then removing the horizontal subsegments of σ_0 inside horoballs, and replacing them by minimal length super-regular paths with the same endpoints.)

The path σ_0 is a broken geodesic with each breakpoint on $g^k \gamma$ for some γ. The individual geodesics have length at least 100δ, and the Gromov products at the vertices are at most 6δ. In particular, σ is a quasi-geodesic. The local modifications producing σ from σ_0 do not change the fact of quasi-geodesicity (though they do change the constants of quasi-geodesicity).

The path σ_0 lies a 2δ–neighborhood of γ. The path σ thus lies in a $(2\delta + 3)$–neighborhood of γ, and in a $(4\delta + 3)$–neighborhood of any other geodesic with the same endpoints. $\qquad\square$

Proof of Theorem 7.8. The proof from [2] works almost as written. Let λ be the constant of quasiconvexity for $\breve{\iota}(X_H)$, and let $C = 2(\lambda + 2\delta) + \max_i\{d_X(1, c_D)\}$, where the elements c_D are those elements chosen as in Section 2. We suppose $R > C + \lambda + 6\delta + 4$.

We first show the more difficult direction, that relative multiplicity dominates relative height. Suppose that the relative height is at least n, so there is some collection of cosets $\{H, g_2 H, \ldots, g_n H\}$ and loxodromic elements $h_1, \ldots h_n \in H$ so that $h_1 = g_2 h_2 g_2^{-1} = \cdots = g_n h_n g_n^{-1}$. Let σ be the quasi-axis for h_1 given by Lemma 7.10, and let γ be any bi-infinite geodesic with the same endpoints at infinity as σ. Requirement (2) implies that $N_R(\gamma) \cap N_R(G)$ is contained in $J = \widetilde{Z} \cap g_2 \widetilde{Z} \cap \cdots \cap g_n \widetilde{Z}$. In particular, any points of $\sigma \cap N_{R-(4\delta+3)}(G)$ are contained in J. We next need to show that the deeper points of σ are also contained in J.

Choose $g \in G$ on the quasi-axis σ. Fix $1 \leq i \leq n$. We claim that there is an element $\hat{h}_i$ in H so that $d_X(g, g_i \hat{h}_i) \leq C$ (taking $g_1 = 1$). Indeed, $g_i h_i g_i^{-1} = h_1$ leaves σ invariant, so the endpoints of σ lie in $g_i \Lambda_H$. Suppose that ρ is a bi-infinite geodesic with the same endpoints as σ. Then by Lemma 7.10 σ lies in a $(4\delta + 3)$–neighborhood of ρ. On the other hand, quasi-convexity implies that any point on ρ lies within distance $\lambda + 2\delta$ of $\breve{\iota}(X_H)$. Possibly, the point g lies within $\lambda + 2\delta$ of a point in $\breve{\iota}(X_H)$ which lies within a horoball, but then this point has depth at most $\lambda + 2\delta$, and so lies within distance $\lambda + 2\delta + \max\{d_X(1, c_D)\}$ of a point in H. The claim follows.

Now, let A be a horoball $(R - 4\delta + 3)$–penetrated by σ, and note that $R - (4\delta + 3) > C + \lambda + 2\delta + 1$. Any point on σ lies within 2δ of some
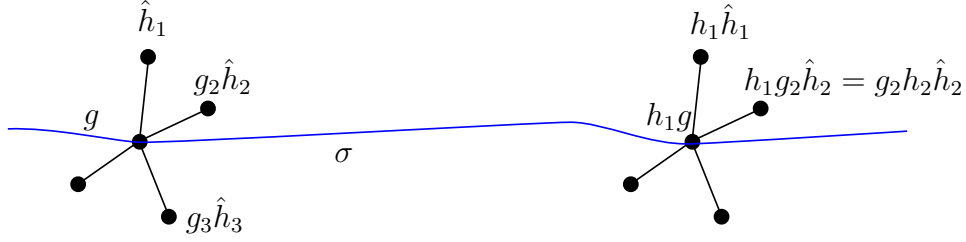
FIGURE 3. Showing that the points of γ lie in J.

point of geodesics $[g, \hat{h}_1], [\hat{h}_1, h\hat{h}_1], [h\hat{h}_1, h_1g]$. However, the first and third of these geodesics are between points in the Cayley graph and have length at most C. Therefore, any points of σ at depth greater than $R - (4\delta + 3)$ in A must be within 2δ of the geodesic between $\hat{h}_1$ and $h\hat{h}_1$. This implies that there is a geodesic with endpoints in H which $(\lambda + 1)$–penetrates A, which by λ-quasiconvexity implies that A contains points at depth greater than 0 in the image $\check{\imath}(X_H)$. Condition (3) from Definition 7.3 (along with the requirement that an R–hull be a full subgraph) ensures that the intersection of $\widetilde{Z}$ with A consists of a collection of vertical lines together with any horizontal edges connecting them. In particular, the (super-regular) subsegment of σ meeting A is contained in $\widetilde{Z}$. An exactly analogous argument shows that this subsegment is contained in $g_i\widetilde{Z}$ for each i, so all of σ is contained in J.

This implies that σ projects to a loop in S_n of the type desired; if C is the component containing the image of σ, then $\tau_{C,i}(\pi_1(C, p))$ contains a conjugate of the loxodromic h_1 for each i.

The other direction, that relative height dominates relatively multiplicity, is almost exactly the same as in [2, Appendix A]. The only difference is that we assume that the intersection is infinite and non-parabolic, so that it contains a loxodromic element. This loxodromic element is then the one required by Definition 7.7. $\square$

Corollary 7.11. [16, Theorem 1.4] *The relative height of a relatively quasiconvex subgroup of a relatively hyperbolic group is finite.*

Proof. If the relative multiplicity is n, then in particular, S_n contains a loop with a vertex at depth 0. Since S_n avoids the fat diagonal, this vertex represents an n–tuple of distinct depth 0 vertices of Z. There are only finitely many such vertices, so the relative height is bounded. $\square$

7.1. **Non-increasing of height under wide fillings.** Suppose that $\widetilde{Z}$ is an R–hull for the action of H on X. The following is an analog of [2, Lemma A.45].

Lemma 7.12 (cf. Lemma A.45, [2]). *For all sufficiently long and H–wide fillings $\phi\colon G \to G(N_1,\ldots,N_m)$, if $K = \ker(\phi)$, $K_H = K \cap H$, and $k \in K \smallsetminus K_H$, then $k.\widetilde{Z} \cap \widetilde{Z} = \emptyset$.*

Proof. By conditions (3) and (4) of an R–hull, there is some R' so that if $g\widetilde{Z} \cap \widetilde{Z} \neq \emptyset$, and if N is the R'–neighborhood of H in the Cayley graph of G, then $gN \cap N \neq \emptyset$. It follows that the set of g for which $g\widetilde{Z} \cap \widetilde{Z} \neq \emptyset$ is contained in a finite union of double cosets

$$A = \bigsqcup_{i=0}^{l} Hg_iH, \text{ with } g_0 = 1.$$

Now let ϕ be long and H–wide enough to apply Proposition 4.7 to all the elements $g_1,\ldots,g_l$. For such a filling we have $\phi(g_i) \notin \phi(H)$ for each i. Equivalently, there is no $k \in K$ of the form g_ih for $h \in H$ and $i > 0$.

Suppose by way of contradiction that $k \in (K \setminus K_H) \cap A$. Then we can write $k = h_1g_ih_2$ for some $i > 0$ and some $h_1, h_2 \in H$. Conjugating we obtain a $k' \in (K \setminus K_H) \cap A$ which lies in g_iH. But this contradicts the last paragraph. $\square$

Remark 7.13. Lemma A.45 in [2] is a special case of Lemma 7.12. The proof given in [2] contains the erroneous assertion that A is a finite union of *left* cosets; otherwise the proof given there is similar to our proof here of Lemma 7.12, but using a theorem about H–fillings [2, A.43] in place of our Proposition 4.7.

Suppose $\pi\colon (G,\mathcal{P}) \to (\overline{G},\overline{\mathcal{P}})$ is a Dehn filling, and $X(G,\mathcal{P})$ is the combinatorial cusped space for $(G,\mathcal{P})$. If K is the kernel of the quotient map $G \to \overline{G}$, then the quotient $\overline{X} = {}_K\backslash X(G,\mathcal{P})$ is very nearly equal to the cusped space for the pair $(\overline{G},\overline{\mathcal{P}})$, differing only in the addition of some self-loops. In particular, their 0–skeleta are isometric, and we can safely ignore the difference.

Putting Lemma 7.12 together with uniformity of hyperbolicity and quasiconvexity after long Dehn fillings, we can prove the following:

Lemma 7.14. *Fix $(G,\mathcal{P})$ relatively hyperbolic, and a relatively quasiconvex subgroup H. For all R, there is an R' satisfying the following: For all sufficiently long and H–wide fillings $\phi\colon G \to G(N_1,\ldots,N_m)$, if $K = \ker(\phi)$, if $\widetilde{Z}$ is an R'–hull for H, then $\overline{\widetilde{Z}} \subset {}_K\backslash X$ is an R–hull for the image of H in $G(N_1,\ldots,N_m)$.*

Proof. Let δ be such that $\overline{X} = {}_K\backslash X$ is δ–hyperbolic whenever K is the kernel of a sufficiently long filling (see Proposition 2.2). As discussed above this quotient is essentially equal to the combinatorial cusped

space for the pair $(\overline{G}, \overline{\mathcal{P}})$, where $\overline{G} = G(N_1, \ldots, N_m)$, and $\overline{\mathcal{P}}$ consists of the images of the elements of $\mathcal{P}$. Let λ' be the constant from Proposition 4.5, so that $\phi(H)$ is λ'–relatively quasiconvex for a sufficiently long and H–wide filling. Let $R_0 = R_0(\lambda', \delta)$ be such that any bi-infinite geodesic with endpoints in the limit set of a λ'–quasiconvex subset of a δ–hyperbolic space is contained in the R_0–neighborhood of that quasiconvex subset. Let $C = \max\{d_X(1, c_D)\}$, where c_D ranges over the elements chosen in Section 2. Finally we fix some $R' > 3R + 2R_0 + C$.

We assume that ϕ is sufficiently long and H–wide so that the results from the last paragraph apply.

We suppose that $\tilde{Z}$ is an R'–hull, and show that the image $\widetilde{\overline{Z}} \subset {}_K\backslash X$ is an R–hull for the image $\overline{H}$ of H in $G(N_1, \ldots, N_m)$.

Conditions (1) and (3) of Definition 7.3 follow easily from the fact that $\tilde{Z}$ is an R'–hull. Condition (4) is a fairly straightforward consequence of the fact that $\overline{H}$ is relatively quasiconvex.

We now establish Condition (2). Suppose that γ is a bi-infinite geodesic with endpoints in $\Lambda(\overline{H})$. Let $p \in N_R(\gamma) \cap N_R(\overline{G})$. Since $\check{\iota}(X_{\overline{H}})$ is λ'–quasiconvex, we have $d_{\overline{X}}(p, x) \leq R_0 + R$ for some $x \in \check{\iota}(X_{\overline{H}})$. Since the depth of p is at most R, the depth of x is at most $R_0 + 2R$. Thus there is some $\overline{h} \in \overline{H}$ with $d_{\overline{X}}(x, \overline{h}) \leq R_0 + 2R + C$. Choose $h \in H$ projecting to $\overline{h}$, and note that there is a bi-infinite geodesic passing through h with endpoints in $\Lambda(H)$.

(We remark that because we assumed that γ exists, $\Lambda(\overline{H})$ contains more than one point. It follows that $\Lambda(H)$ contains more than one point, which implies the existence of such a bi-infinite geodesic.)

In particular, an R'–ball about h is contained in the R'–hull $\tilde{Z}$. Since $R' > d_{\overline{X}}(\overline{h}, p)$, the image of $\tilde{Z}$ in ${}_K\backslash X$ must contain p. $\square$

We now prove that the relative height of H does not increase under sufficiently long and H–wide fillings.

Theorem 7.15. *(cf. [2, A.46]) For sufficiently long and H–wide fillings, the relative height of $(\overline{H}, \overline{\mathcal{D}})$ in $(\overline{G}, \overline{\mathcal{P}})$ is at most the relative height of $(H, \mathcal{D})$ in $(G, \mathcal{P})$.*

Proof. As usual, let δ be a constant so that the cusped space of $(G, \mathcal{P})$ and also those of sufficiently long fillings, are δ–hyperbolic, and let λ be a quasi-convexity constant for H, which we also assume (using Proposition 4.5) is a quasi-convexity constant for the image of H under sufficiently long and H–wide fillings.

Let R be sufficiently large to apply Theorem 7.8 with these values of δ and λ. Let R' the the constant (depending on R) from the conclusion of Lemma 7.14.

Consider the following commutative diagram, which is equivariant with respect to the group actions and the natural maps between the groups (inclusion and quotient maps, as appropriate):

$$
\begin{array}{ccc}
 & H & & G \\
 & \curvearrowright & & \curvearrowleft \\
\widetilde{Z} & \longrightarrow & X \\
\downarrow & & \downarrow \\
H/K_H \curvearrowright \widetilde{\overline{Z}} & \longrightarrow & \overline{X} \curvearrowleft G/K
\end{array}
$$

where X is the cusped space for $(G, \mathcal{P})$, $\overline{X} = {}_K\backslash X$, $\widetilde{Z}$ is an R–hull for H, $K_H = H \cap K$ is the kernel of the induced filling on H, and $\widetilde{\overline{Z}} = {}_{K_H}\backslash\widetilde{Z}$.

It follows immediately from Lemma 7.12 that $\widetilde{\overline{Z}}$ embeds in $\overline{X}$, and it follows from Lemma 7.14 that $\widetilde{\overline{Z}}$ is an R–hull for H/K_H in $\overline{X}$.

Taking quotients by the relevant groups we get the diagram,

(1)
$$
\begin{array}{ccc}
Z & \xrightarrow{\ i\ } & Y \\
\downarrow & & \downarrow \\
\overline{Z} & \xrightarrow{\ \overline{i}\ } & \overline{Y}
\end{array}
$$

where the horizontal maps are immersions inducing the inclusions $H \to G$ and $\overline{H} \to \overline{G}$ on the level of fundamental group. The vertical maps from Y to $\overline{Y}$ and from Z to $\overline{Z}$ are homeomorphisms. Theorem 7.8 implies that the relative multiplicities of Z in Y and of $\overline{Z}$ in $\overline{Y}$ measure the relative heights of H in G and of H/K_H in G/K, respectively.

For $n > 0$, define

$$S_n = \{(z_1, \ldots, z_n) \in Z^n \mid i(z_1) = \cdots = i(z_n)\} \setminus \Delta$$

and

$$\overline{S}_n = \{(z_1, \ldots, z_n) \in \overline{Z}^n \mid \overline{i}(z_1) = \cdots = \overline{i}(z_n)\} \setminus \overline{\Delta},$$

where $\Delta = \{(z_1, \ldots, z_n) \mid \text{ there exist } i \neq j \text{ so that } z_i = z_j\}$ is the fat diagonal in Z^n and $\overline{\Delta}$ is the fat diagonal in $\overline{Z}^n$.

For each $i \in \{1, \ldots, n\}$ and each component C of S_n the projections of Z^n to its factors induce maps

$$\tau_{C,i}\colon \pi_1(C) \to H,$$

and

$$\overline{\tau}_{C,i}\colon \pi_1(C) \to \overline{H}.$$

Since the quotient $Z = {}_H\backslash\widetilde{Z}$ can also be thought of as $(H/K_H)\backslash\widetilde{\overline{Z}}$, the homomorphisms $\overline{\tau}_{C,i}$ all factor as $\overline{\tau}_{C,i} = \phi|_H \circ \tau_{C,i}$, where ϕ is the filling map.

In particular, if γ is a loop in $\overline{S}_n$ so that $\overline{\tau}_{C,i}([\gamma])$ is infinite for each $i \in \{1, \ldots, n\}$ then it must be that $\tau_{C,i}([\gamma])$ is already infinite for each i. $\qquad\square$

In case parabolic subgroups of a relatively hyperbolic group are finite, the relative height is the same as the height. Recall that a filling $G \to G(N_1, \ldots, N_m)$ is *peripherally finite* if for each $1 \le j \le m$ the subgroup N_j has finite-index in P_j. The following is an immediate consequence of Theorem 7.15.

Corollary 7.16. *For sufficiently long and H–wide peripherally finite fillings, the height of $\overline{H}$ in $\overline{G}$ is at most the relative height of $(H, \mathcal{D})$ in $(G, \mathcal{P})$.*

7.2. **A result required by Wilton and Zalesskii.**

Definition 7.17. [27, Definition 4.1] Suppose that $(G, \mathcal{P})$ is relatively hyperbolic and $H \le G$. We say that H is *relatively malnormal* if for any $g \notin H$ the intersection $H^g \cap H$ is conjugate into some element of $\mathcal{P}$.

If $(G, \mathcal{P})$ is relatively hyperbolic and G is torsion-free, then a relatively malnormal subgroup is either parabolic or a subgroup of relative height 1.

For his joint work with Zalesskii [27], Henry Wilton asked us to prove the following result, which appeared as [27, Theorem 4.4]:

Theorem 7.18. *Let G be a toral relatively hyperbolic group with parabolic subgroups $\{P_1, \ldots, P_n\}$ and let H be a subgroup which is relatively quasi-convex and relatively malnormal. There exist subgroups of finite index $K_i' \subset P_i$ (for all i) such that, for all subgroups of finite index $L_i \subset K_i'$, if*

$$\eta\colon G \to Q = G/\langle\!\langle L_1, \ldots, L_n \rangle\!\rangle$$

is the quotient map, the quotient Q is word-hyperbolic and the image $\eta(H)$ in Q is quasi-convex and almost malnormal.

Proof. We restrict to peripherally finite fillings. For sufficiently long peripherally finite fillings $\eta\colon G \to Q$, the quotient Q is hyperbolic (since it is hyperbolic relative to finite groups) by [11, Theorem 7.3.(2)]. Since the peripheral subgroups of Q are finite, there is no difference between quasiconvex and relatively quasiconvex subgroups.

By Proposition 4.5, for sufficiently long and H–wide fillings $\eta\colon G \to Q$ the image $\eta(H)$ is quasi-convex, and by Corollary 7.16, for sufficiently long and H–wide fillings $\eta(H)$ is almost malnormal in Q.

It remains only to note that sufficiently long and H–wide peripherally finite fillings exist by Lemma 5.2, since the peripheral subgroups of G are free abelian, and hence ERF. $\square$

References

[1] I. Agol. Criteria for virtual fibering. *J. Topol.*, 1(2):269–284, 2008.

[2] I. Agol. The Virtual Haken Conjecture. *Doc. Math.*, 18:1045–1087, 2013. With an appendix by Agol, Daniel Groves, and Jason Manning.

[3] I. Agol, D. Groves, and J. F. Manning. Residual finiteness, QCERF and fillings of hyperbolic groups. *Geom. Topol.*, 13(2):1043–1073, 2009.

[4] I. Agol, D. Groves, and J. F. Manning. An alternate proof of Wise's malnormal special quotient theorem. *Forum Math. Pi*, 4:e1, 54, 2016.

[5] M. R. Bridson and A. Haefliger. *Metric Spaces of Non–Positive Curvature*, volume 319 of *Grundlehren der mathematischen Wissenschaften*. Springer–Verlag, Berlin, 1999.

[6] D. Cooper and D. Futer. Ubiquitous quasi-Fuchsian surfaces in cusped hyperbolic 3–manifolds. *Geom. Topol.*, 23(1):241–298, 2019.

[7] R. Coulon. On the geometry of Burnside quotients of torsion free hyperbolic groups. *International Journal of Algebra and Computation*, 24(3):251–345, 2014.

[8] F. Dahmani. Combination of convergence groups. *Geom. Topol.*, 7:933–963 (electronic), 2003.

[9] F. Dahmani and V. Guirardel. Recognizing a relatively hyperbolic group by its Dehn fillings. *Duke Math. J.*, 167(12):2189–2241, 2018.

[10] W. D. Dunbar and G. R. Meyerhoff. Volumes of hyperbolic 3-orbifolds. *Indiana Univ. Math. J.*, 43(2):611–637, 1994.

[11] D. Groves and J. F. Manning. Dehn filling in relatively hyperbolic groups. *Israel Journal of Mathematics*, 168:317–429, 2008.

[12] D. Groves and J. F. Manning. Dehn fillings and elementary splittings. *Trans. Amer. Math. Soc.*, 370(5):3017–3051, 2018.

[13] F. Haglund and D. T. Wise. Special cube complexes. *Geom. Funct. Anal.*, 17(5):1551–1620, 2008.

[14] F. Haglund and D. T. Wise. Coxeter groups are virtually special. *Adv. Math.*, 224(5):1890–1903, 2010.

[15] G. C. Hruska. Relative hyperbolicity and relative quasiconvexity for countable groups. *Algebr. Geom. Topol.*, 10(3):1807–1856, 2010.

[16] G. C. Hruska and D. T. Wise. Packing subgroups in relatively hyperbolic groups. *Geom. Topol.*, 13(4):1945–1988, 2009.

[17] G. C. Hruska and D. T. Wise. Finiteness properties of cubulated groups. *Compos. Math.*, 150(3):453–506, 2014.

[18] J. F. Manning and E. Martínez-Pedroza. Separation of relatively quasiconvex subgroups. *Pacific J. Math.*, 244(2):309–334, 2010.

[19] E. Martínez-Pedroza. Combination of quasiconvex subgroups of relatively hyperbolic groups. *Groups Geom. Dyn.*, 3(2):317–342, 2009.

[20] A. Minasyan. Separable subsets of GFERF negatively curved groups. *J. Algebra*, 304(2):1090–1100, 2006.

[21] I. Mineyev. Straightening and bounded cohomology of hyperbolic groups. *Geom. Funct. Anal.*, 11(4):807–839, 2001.

[22] D. V. Osin. Peripheral fillings of relatively hyperbolic groups. *Invent. Math.*, 167(2):295–326, 2007.

[23] J.-P. Otal. *The hyperbolization theorem for fibered 3-manifolds*, volume 7 of *SMF/AMS Texts and Monographs*. American Mathematical Society, Providence, RI; Société Mathématique de France, Paris, 2001. Translated from the 1996 French original by Leslie D. Kay.

[24] P. Przytycki and D. T. Wise. Mixed 3-manifolds are virtually special. *J. Amer. Math. Soc.*, 31(2):319–347, 2018.

[25] M. Sageev. Ends of group pairs and non-positively curved cube complexes. *Proc. London Math. Soc. (3)*, 71(3):585–617, 1995.

[26] G. P. Scott and G. A. Swarup. Geometric finiteness of certain Kleinian groups. *Proc. Amer. Math. Soc.*, 109(3):765–768, 1990.

[27] H. Wilton and P. Zalesskii. Distinguishing geometries using finite quotients. *Geom. Topol.*, 21(1):345–384, 2017.

[28] D. T. Wise. The structure of groups with a quasiconvex hierarchy. Unpublished manuscript, version dated October 29, 2012.

DEPARTMENT OF MATHEMATICS, STATISTICS, AND COMPUTER SCIENCE, UNIVERSITY OF ILLINOIS AT CHICAGO, 322 SCIENCE AND ENGINEERING OFFICES (M/C 249), 851 S. MORGAN ST., CHICAGO, IL 60607-7045

E-mail address: groves@math.uic.edu

DEPARTMENT OF MATHEMATICS, 310 MALOTT HALL, CORNELL UNIVERSITY, ITHACA, NY 14853

E-mail address: jfmanning@math.cornell.edu