Dehn filling and the geometry
of unknotting tunnels

DARYL COOPER
DAVID FUTER
JESSICA S PURCELL

Any one-cusped hyperbolic manifold M with an unknotting tunnel τ is obtained by Dehn filling a cusp of a two-cusped hyperbolic manifold. In the case where M is obtained by "generic" Dehn filling, we prove that τ is isotopic to a geodesic, and characterize whether τ is isotopic to an edge in the canonical decomposition of M. We also give explicit estimates (with additive error only) on the length of τ relative to a maximal cusp. These results give generic answers to three long-standing questions posed by Adams, Sakuma and Weeks.

We also construct an explicit sequence of one-tunnel knots in S^3, all of whose unknotting tunnels have length approaching infinity.

1 Introduction

Let M be a compact orientable 3–manifold whose boundary consists of tori. An *unknotting tunnel* for M is a properly embedded arc τ from ∂M to ∂M, such that $M \smallsetminus \tau$ is a genus-2 handlebody. Not all 3–manifolds with torus boundary admit an unknotting tunnel; those that do are said to be *one-tunnel* or *tunnel number one*. The above definition immediately implies that every one-tunnel manifold M has one or two boundary components, and has Heegaard genus two (unless it is a solid torus).

For this paper, we investigate one-tunnel manifolds M such that the interior of M carries a complete hyperbolic metric. The geometric study of unknotting tunnels in this setting begins with two foundational papers published in 1995 by Adams [1] and Sakuma and Weeks [39]. These papers posed three open questions about unknotting tunnels of one-cusped manifolds:

(1) Is τ always isotopic to a geodesic?

(2) Is there a universal bound B, such that outside a maximal cusp neighborhood in M, the geodesic in the homotopy class of τ is always shorter than B?

(3)   Is τ isotopic to an edge in the canonical polyhedral decomposition of M?

One motivation behind these questions is that for complements of two-bridge knots in S^3, the answer to all three questions is "yes" (Adams and Reid [4], and Akiyoshi, Sakuma, Wada and Yamashita [6]). However, apart from the special family of two-bridge knots, the only progress to date has consisted of selected examples for question (1), and selected counterexamples to questions (2) and (3).

In this paper, we give detailed answers to all three questions, under the hypothesis that M is obtained by "generic" Dehn filling on one cusp of a two-cusped hyperbolic manifold X. In this generic setting, the tunnel τ is indeed isotopic to a geodesic. Generically, this geodesic is quite long, and we provide explicit estimates on the length, with additive error only. Whether or not τ is isotopic to an edge of the canonical decomposition turns out to depend on the length of an associated tunnel $\sigma \subset X$ (see Figure 1).

In addition, we construct an explicit sequence of one-tunnel knots $K_n \subset S^3$, such that each K_n has two unknotting tunnels, whose length approaches infinity as $n \to \infty$.

1.1   Generic Dehn fillings and generic unknotting tunnels

Let X be a compact orientable 3–manifold whose boundary consists of one or more tori, and whose interior is hyperbolic. Let T be one of the tori of ∂M. A *slope* on T is an isotopy class of simple closed curves. The *Dehn filling of X along a slope μ*, denoted $X(\mu)$, is the manifold obtained by attaching a solid torus $D^2 \times S^1$ to T, so that ∂D^2 is glued to μ. The slope μ is called the *meridian* of the filling.

Definition 1.1   An embedded horoball neighborhood of a boundary torus $T \subset X$ is called a *horocusp*, and denoted H_T. If we fix such a horocusp H_T, the horospherical torus ∂H_T inherits a Euclidean metric that allows us to measure the length of slopes. In particular, a slope μ chosen as a meridian for Dehn filling has a well-defined length $\ell(\mu)$, namely the length of a Euclidean geodesic representing μ on ∂H_T. In a similar way, we define a *longitude* of the Dehn filling to be a simple closed curve on ∂H_T that intersects μ once. We will typically be interested in the shortest longitude, denoted λ. In highly symmetric cases, there can be two shortest longitudes (up to isotopy), and we may choose either one. Note that once the horocusp is fixed, the lengths $\ell(\mu)$ and $\ell(\lambda)$ are always well-defined.

Definition 1.2   We say that the Dehn filling along μ is *generic* if both μ and λ are sufficiently long. Equivalently, the filling is generic if both μ and λ avoid finitely many prohibited slopes on the torus T.

If we fix a basis $\langle \alpha, \beta \rangle$ for $H_1(T) \cong \mathbb{Z}^2$, then all the possible choices of Dehn filling slope are parametrized by primitive pairs of integers $(p, q) \in \mathbb{Z}^2$. In this setting, choosing a generic slope amounts to avoiding finitely many points and finitely many lines in $\mathbb{R}^2$.

The term *generic* can be justified as follows. Let $\mathcal{F}$ be the Farey graph, whose vertices are slopes on the torus T, and whose edges correspond to slopes that intersect once. For each prohibited value of λ, the values of μ that have λ as a longitude lie on a circle of radius 1 in $\mathcal{F}$, centered at λ. Thus prohibiting finitely many values of μ and λ amounts to prohibiting μ from lying in finitely many closed balls of radius 1. Since the Farey graph $\mathcal{F}$ has infinite diameter, almost all choices of μ will avoid the prohibited sets, and are indeed generic. In particular, a random walk in $\mathcal{F}$ will land on a generic slope with probability approaching 1.

We would like to argue that the unknotting tunnels created by generic Dehn filling (as in Definition 1.2) are also "generic," in an appropriate sense. This must be done with some care, as there are multiple reasonable notions of genericity (for instance Dunfield and Thurston [16], and Lustig and Moriah [30]).

Suppose X is a manifold with cusps T and T', and an unknotting tunnel σ. Then the Heegaard surface associated to σ (namely, the boundary Σ of a regular neighborhood of $T \cup T' \cup \sigma$) cuts X into a genus-2 handlebody C and a compression body C'. (See Definition 2.1 for details.) One way to obtain a "random" two-cusped 3–manifold of this type is to glue C to C' via a random walk in the generators of the mapping class group $\mathrm{Mod}(\Sigma)$; see [16]. In this context, Maher has shown that with probability approaching 1, a random walk in $\mathrm{Mod}(\Sigma)$ gives a Heegaard splitting of high distance [32]. Then, the work of Scharlemann and Tomova implies that Σ will be the only genus-2 Heegaard surface of X [41]. The same conclusion holds under more measure-theoretic notions of genericity: see Lustig and Moriah [30]. Thus, in two reasonable senses, one can say that generic one-tunnel manifolds have a unique unknotting tunnel and a unique minimal-genus Heegaard surface.

This generic uniqueness is preserved after Dehn filling. Results of Moriah and Rubinstein [34], and Rieck and Sedgwick [37] imply that for a generic Dehn filling slope μ on T, every genus-2 Heegaard surface Σ of $X(\mu)$ comes from a Heegaard surface of X. See Theorem 2.3 for details, including quantified hypotheses. Thus, by the previous paragraph, a generic Dehn filling of a generic one-tunnel manifold X will have exactly one unknotting tunnel.

The unknotting tunnels created by Dehn filling have a natural visual description, summarized in Figure 1. If X is a manifold with cusps T and T', any unknotting tunnel σ of X must connect T with T'. Then there will be a new tunnel $\tau \subset M = X(\mu)$
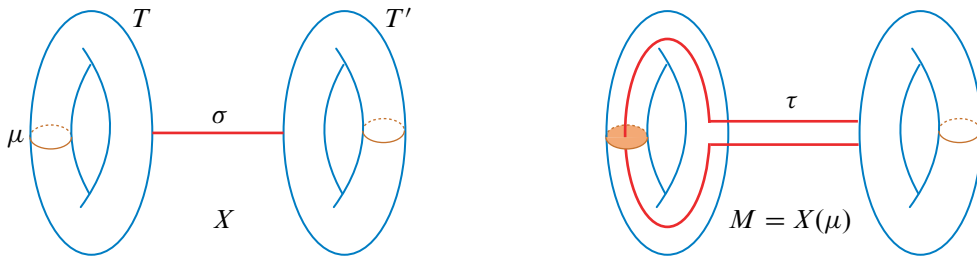
Figure 1: A schematic picture of unknotting tunnels under Dehn filling. Left: σ is a tunnel for a 2–cusped manifold X. Right: the associated tunnel τ of the 1–cusped manifold $M = X(\mu)$.

that starts at T' and runs along σ, followed by a longitude of the filling, followed by backtracking along σ. It is easy to verify that $M \smallsetminus \tau \cong X \smallsetminus \sigma$, hence is a handlebody. We call τ the tunnel of $M = X(\mu)$ that is *associated to* σ. See Definition 2.10 and Theorem 2.11 for a much more detailed description of associated tunnels.

The theorems stated below describe the geometry of all associated tunnels created by generic Dehn filling. In particular, we answer questions (1), (2) and (3) for these tunnels.

1.2 Tunnels isotopic to geodesics

Question (1) has the following history. Adams showed, using a symmetry argument, that every unknotting tunnel of a two-cusped hyperbolic manifold is isotopic to a geodesic [1]; see also Lemma 2.9. Shortly after, Adams and Reid extended these symmetry arguments to prove that the upper and lower tunnels of a 2–bridge knot are isotopic to geodesics [4]. However, since the late 1990s, there has been only minimal progress on question (1). As a negative result, Futer showed that the symmetry arguments of Adams and Reid do not apply to any knots in S^3 besides 2–bridge knots [18].

For generic Dehn fillings, we provide a positive answer to question (1):

Theorem 1.3 *Let X be an orientable hyperbolic 3–manifold that has two cusps and tunnel number one. Choose a generic filling slope μ on one cusp of X, and let $\tau \subset X(\mu)$ be an unknotting tunnel associated to a tunnel $\sigma \subset X$. Then τ is isotopic to a geodesic in the hyperbolic metric on $X(\mu)$.*

Note that by the discussion above (more precisely, by Theorem 2.11 and Remark 2.13), a two-cusped manifold X constructed by a random Heegaard splitting will have a unique unknotting tunnel σ, and its generic Dehn filling $M = X(\mu)$ will have a unique tunnel τ associated to σ. Thus, generically, M has exactly one unknotting tunnel, which is isotopic to a geodesic.

1.3 The length of unknotting tunnels

To measure the length of an unknotting tunnel τ, one first needs to choose horospherical cusp neighborhoods in the ambient manifold M. If M has one boundary torus, there is a canonical choice of *maximal cusp*, namely the closure of the largest embedded horocusp about ∂M. If M has two (or more) cusps, then the maximal cusp neighborhood will depend on the order in which the cusps are expanded. Once the cusp neighborhoods are fixed, we define the *length* of an unknotting tunnel τ to be the length of the geodesic in the homotopy class of τ, outside the given horocusps in M.

If a hyperbolic one-tunnel manifold M has two boundary tori, Adams [1] showed that there always exist disjoint horocusps about these tori such that every unknotting tunnel of M has length at most $\ln(4)$. In a subsequent preprint [3], he improved the upper bound to $\frac{7}{4}\ln(2)$. These universal upper bounds for two-cusped manifolds prompted a wide belief that the unknotting tunnels of one-cusped manifolds also have universally bounded length.

In a recent paper [13], Cooper, Lackenby and Purcell showed that in fact, the answer to question (2) is "no": there exist one-cusped hyperbolic manifolds whose unknotting tunnels are arbitrarily long outside a maximal cusp. However, the examples in [13] either were non-constructive, or could not be complements of knots in S^3. The authors asked whether there exist knots in S^3 with arbitrarily long unknotting tunnels, and whether such examples can be explicitly described.

In this paper, we show that generically, unknotting tunnels are very long. In fact, we compute the length of τ, up to additive error only.

Theorem 1.4 *Let X be an orientable hyperbolic 3–manifold that has two cusps and an unknotting tunnel σ. Let T be one boundary torus of X. Then, for all but finitely many choices of a Dehn filling slope μ, the unknotting tunnel τ of $X(\mu)$ associated to σ satisfies*

$$2\ln \ell(\lambda) - 6 < \ell(\tau) < 2\ln \ell(\lambda) + 5,$$

where λ is the shortest longitude of the Dehn filling, $\ell(\lambda)$ is the length of λ on a maximal cusp corresponding to T, and $\ell(\tau)$ is the length of the geodesic in the homotopy class of τ.

We also apply this result to knots in S^3: see Theorem 1.7 below.

1.4 Canonical geodesics

In a one-cusped hyperbolic manifold M, let H be a closed, embedded horocusp. Then the Ford–Voronoi domain F is defined to be the set of all points in M that have a unique shortest path to H. This is an open set in M, canonically determined by the geometry of M and, in particular, independent of the chosen size of H. The complement $L = M \smallsetminus F$ is a compact 2–complex, called the *cut locus*. The combinatorial dual to L is an ideal polyhedral decomposition $\mathcal{P}$ of M; the n–cells of $\mathcal{P}$ are in bijective correspondence with the $(3 - n)$–cells of L. This is called the *Epstein–Penner decomposition* or *canonical polyhedral decomposition* of M.

For one-cusped manifolds, the canonical decomposition $\mathcal{P}$ is a complete invariant of the homeomorphism type of M. For multi-cusped manifolds, one may perform the same construction, although the combinatorics of the resulting polyhedral decomposition may depend on the relative volumes of the horocusps $H_1, \ldots, H_k$.

Definition 1.5 Let M be a one-cusped hyperbolic manifold. We say that an arc τ from cusp to cusp (in practice, an unknotting tunnel) is *canonical* if τ is isotopic to an edge of the canonical polyhedral decomposition $\mathcal{P}$.

Sakuma and Weeks performed an extensive study of the triangulations of 2–bridge knot complements [39]. Using experimental evidence from SnapPea [14], they conjectured that these triangulations are canonical; a conjecture subsequently proved by Akiyoshi, Sakuma, Wada, and Yamashita [6]. In addition, Sakuma and Weeks observed that the unknotting tunnels of 2–bridge knots are always isotopic to edges of this triangulation, which led them to conjecture that all unknotting tunnels of hyperbolic manifolds are canonical [39].

This conjecture was disproved in 2005, with a single counterexample constructed by Heath and Song [26]. For the $(-2, 3, 7)$ pretzel knot K, they showed that $S^3 \smallsetminus K$ has four unknotting tunnels but only three edges in its canonical triangulation. Although this example settled question (3) in the negative, it did not shed light on the broader question of what properties of an unknotting tunnel imply that it is, or is not, canonical.

In the context of generic Dehn filling, this broader question has the following answer.

Theorem 1.6 *Let X be a two-cusped, orientable hyperbolic 3–manifold in which there is a unique shortest geodesic arc between the two cusps. Choose a generic Dehn filling slope μ on a cusp of X. Then, for each unknotting tunnel $\sigma \subset X$, the tunnel $\tau \subset X(\mu)$ associated to σ will be canonical if and only if σ is the shortest geodesic between the two cusps of X.*

If there are several shortest geodesics between the two cusps of X, then Theorem 1.6 does not give any information. However, the existence of a unique shortest geodesic can also be regarded as a "generic" property of hyperbolic manifolds.

In practice, both alternatives of Theorem 1.6 are quite common. Using this theorem, one may easily construct infinite families of manifolds whose tunnels are canonical, as well as infinite families that have non-canonical tunnels. See Theorem 5.3 for one such construction.

1.5 Knots with long tunnels in S^3

Theorems 1.3, 1.4 and 1.6 can be applied to construct explicit families of knots in S^3, whose unknotting tunnels have interesting properties.

Theorem 1.7 *There is a sequence K_n of hyperbolic knots in S^3, such that each K_n has exactly two unknotting tunnels. Each unknotting tunnel τ_n of K_n is isotopic to a canonical geodesic, whose length is*

$$2n \ln\left(\frac{1+\sqrt{5}}{2}\right) - 5 < \ell(\tau_n) < 2n \ln\left(\frac{1+\sqrt{5}}{2}\right) + 6.$$

The sequence of knots K_n is explicitly described in Section 7. See Figure 11 for a preview.

1.6 Organization of the paper

This paper is organized as follows. In Section 2, we fill in the details of a number of definitions and theorems that were mentioned above. In Theorem 2.3, we describe the effect of Dehn filling on Heegaard surfaces, adding quantified hypotheses to a theorem of Moriah and Rubinstein [34], and Rieck and Sedgwick [37]. In Theorem 2.6, we recall the drilling and filling theorems of Hodgson and Kerckhoff [27] and Brock and Bromberg [11], which allow precise bilipschitz estimates on the change in geometry during Dehn filling. This bilipschitz control will be used in all the geometric estimates that follow.

In Section 3, we use the work of Adams on unknotting tunnels of two-cusped manifolds [1], combined with Theorem 2.6, to prove Theorem 1.4. More precisely, we prove two-sided estimates on the length of the geodesic g_τ in the homotopy class of a tunnel τ, without yet knowing that τ is isotopic to g_τ. Several quantitative estimates from Section 3 will be used in Section 4 to show that the tunnel τ is isotopic to a geodesic, establishing Theorem 1.3.

In Section 5, we prove Theorem 1.6, which relates the canonicity of a tunnel $\tau \subset X(\mu)$ to the length of its associated tunnel $\sigma \subset X$. The argument in this section relies on the recent work of Guéritaud and Schleimer [23], and also uses the length estimates of Section 3. As an application, we construct an infinite family of one-cusped manifolds, each of which has one canonical and one non-canonical tunnel.

In Sections 6 and 7, we construct knots in S^3 whose unknotting tunnels are arbitrarily long. This construction has two flavors. The argument in Section 6 is quick and direct, but requires making non-explicit "generic" choices. The argument in Section 7 is completely explicit, and gives the precise quantitative estimate of Theorem 1.7. The cost of this entirely explicit construction is that the argument of Section 7 is longer, and requires rigorous computer assistance from the programs Regina [12] and SnapPy [14].

Acknowledgements

2 Geometric setup

The goal of this section is to review and synthesize several past results. We recall the work of Moriah and Rubinstein [34], and Rieck and Sedgwick [37] on Heegaard splittings under Dehn filling, the work of Hodgson and Kerckhoff [27], and Brock and Bromberg [11] on the change in geometry under Dehn filling, and the work of Adams on unknotting tunnels of two-cusped manifolds [1]. Then, we synthesize these results in Theorem 2.11, which explains the one-to-one correspondence between genus-2 Heegaard splittings of a two-cusped manifold X and the unknotting tunnels of any generic Dehn filling $X(\mu)$.

2.1 Heegaard splittings under Dehn filling

Although the goal of this paper is to study unknotting tunnels of one-cusped hyperbolic manifolds, this study will require a slightly more general setup.

Definition 2.1 A *compression body* C is a 3–manifold with boundary, constructed as follows. Start with a genus-g surface Σ. Thicken Σ to $\Sigma \times [0, 1]$, and attach some number (at least one, at most g) of non-parallel 2–handles to $\Sigma \times \{0\}$. If, after attaching 2–handles, any component of the boundary becomes a 2–sphere, cap it off with a 3–ball.

The *positive boundary* of C is the boundary component $\partial_+ C = \Sigma \times \{1\}$, untouched during the construction. The *negative boundary* is $\partial_- C = \partial C \smallsetminus \partial_+ C$. When the negative boundary is empty, the compression body C is a genus-g handlebody.

Given a compact orientable 3–manifold M and a genus-g surface $\Sigma \subset M$, we say that Σ is a *Heegaard splitting surface* of M if Σ cuts M into compression bodies C_1 and C_2, such that $\Sigma = \partial_+ C_1 = \partial_+ C_2$.

Definition 2.2 Let C be a compression body whose positive boundary $\partial_+ C$ has genus 2. Then $\partial_- C$ is a disjoint union of at most two tori. If $\partial_- C = \varnothing$, then C is a handlebody. If $\partial_- C \neq \varnothing$, then C can be constructed by adding exactly one 2–handle to $\Sigma \times \{0\}$. Define the *core tunnel* of C to be an arc σ dual to this 2–handle. It is well-known that the core tunnel is unique up to isotopy.

If a genus-2 surface Σ is a Heegaard surface for X, where ∂X consists of tori, then there are at most two such tori on each side of Σ. If one component of $M \smallsetminus \Sigma$ is a handlebody while the other has non-empty negative boundary, the core tunnel σ is an unknotting tunnel for X.

Suppose that, as above, X is a 3–manifold with toroidal boundary and a genus-2 Heegaard splitting surface Σ. If we perform a Dehn filling along one of the boundary tori of X, it is easy to check that Σ remains a Heegaard surface for the filled manifold $X(\mu)$. (The meridian disk of a solid torus can be thought of as a 2–handle added to the negative boundary of a compression body C. This creates a 2–sphere boundary component, and filling it in amounts to adding the rest of the solid torus.) In this setting, a core tunnel σ of a pre-filling compression body gives rise to a core tunnel τ of the after-filling compression body, exactly as in Figure 1.

Moriah and Rubinstein [34], and Rieck and Sedgwick [37], showed that generically, there is a one-to-one correspondence between minimum-genus Heegaard splittings of X and those of $X(\mu)$. More recently, Futer and Purcell found a way to quantify the hypotheses in their theorem [20]. Here is what the result says in genus 2.

Theorem 2.3 *Let X be an orientable hyperbolic 3–manifold with one or more cusps and Heegaard genus 2. Let T be one boundary torus of X. Choose a Dehn filling slope μ on T, such that $\ell(\mu) > 6\pi$ and the shortest longitude λ for μ has length $\ell(\lambda) > 6$. Then:*

(a) $X(\mu)$ *is a hyperbolic manifold of Heegaard genus* 2.

(b) *For every genus-2 Heegaard surface* Σ *of* $X(\mu)$, *the core curve* γ *of the Dehn filling solid torus is isotopic into* Σ.

(c) *Once* γ *is isotoped into one of the compression bodies separated by* Σ, *the surface* Σ *becomes a Heegaard surface of* $X = X(\mu) \diagdown \gamma$.

Proof If $\ell(\mu) > 2\pi$, the 2π–Theorem of Gromov and Thurston implies that the filled manifold $X(\mu)$ admits a negatively curved metric (Bleiler and Hodgson [9]). (See also Futer, Kalfagianni and Purcell [19, Theorem 2.1] for an explicit construction, with curvature estimates.) Since $X(\mu)$ is negatively curved, its Heegaard genus must be at least 2. But a Heegaard surface Σ of X is also a Heegaard surface for all fillings, hence $X(\mu)$ must have Heegaard genus exactly 2. By geometrization, $X(\mu)$ must also admit a hyperbolic metric, proving conclusion (a).

Conclusions (b) and (c) are a restatement of [20, Theorem 1.1]. $\square$

2.2 Geometric estimates

We will repeatedly need to bound the amount of change of geometry under Dehn filling. To do so, we use a version of the drilling theorem of Brock and Bromberg [11]. Before stating the theorem, we recall several definitions.

Definition 2.4 Given $\epsilon > 0$ and a hyperbolic 3–manifold M, the ϵ–*thin part* $M_{<\epsilon}$ of M is the set of all points in M whose injectivity radius is less than $\epsilon/2$. Equivalently, $M_{<\epsilon}$ is the set of all points that lie on a non-trivial closed curve of length less than ϵ.

A given $\epsilon > 0$ is called a *Margulis number* for M if each component of $M_{<\epsilon}$ has abelian fundamental group. In this case, the ϵ–thin part $M_{<\epsilon}$ is a disjoint union of horocusps and tubular neighborhoods about geodesics. For a particular cusp T, we let $\mathbb{T}_\epsilon(T)$ denote the component of $M_{<\epsilon}$ corresponding to T. Similarly, if γ is a geodesic of length less than ϵ, a tubular neighborhood $\mathbb{T}_\epsilon(\gamma)$ is a component of $M_{<\epsilon}$.

The Margulis Lemma states that there is a positive number ϵ that serves as a Margulis number for every hyperbolic 3–manifold (Benedetti and Petronio [8, Chapter D]). The greatest such ϵ, denoted ϵ_3, is called the (3–dimensional) *Margulis constant*.

The best available estimate on the Margulis constant is $\epsilon_3 \geq 0.104$, due to Meyerhoff [33]. However, under additional hypotheses there are stronger estimates on Margulis numbers. For example, Culler and Shalen recently showed [15] that every cusped hyperbolic 3–manifold has a Margulis number of at least 0.292. See also Shalen [43, Proposition 2.3].

Definition 2.5 Let T be a Euclidean torus, and let g be a closed geodesic on T. The *normalized length* of g is defined to be

$$(2\text{-}1) \qquad L(g) = \ell(g)/\sqrt{\text{area}(T)}.$$

The normalized length $L(\mu)$ of a slope μ on T is defined in the same way, via the normalized length of a geodesic representative of μ.

Note that equation (2-1) is scaling-invariant. Hence, if T is a cusp torus in M, the normalized length of a slope on T does not depend on the choice of horospherical torus.

We can now state a version of Brock and Bromberg's drilling theorem.

Theorem 2.6 (Drilling theorem) *Let X be a hyperbolic 3–manifold with one or more cusps, and let T be a cusp torus of X. Choose any $J > 1$ and any $\epsilon > 0$ that is a Margulis number for each hyperbolic filling along T. Then there is some $K = K(J, \epsilon) \geq 4\sqrt{2} \cdot \pi$ such that every slope μ on T with normalized length $L(\mu) \geq K$ satisfies the following:*

(a) *$X(\mu)$ is a hyperbolic 3–manifold, obtainable from X by a cone deformation.*

(b) *The core curve γ of the added solid torus is a geodesic satisfying*

$$\ell(\gamma) \leq \frac{2\pi}{L(\mu)^2 - 4(2\pi)^2}.$$

(c) *There is a J–bilipschitz diffeomorphism*

$$\phi\colon X \smallsetminus \mathbb{T}_\epsilon(T) \to X(\mu) \smallsetminus \mathbb{T}_\epsilon(\gamma).$$

(d) *ϕ is level-preserving on any remaining cusps of X, mapping horospherical tori to horospherical tori. In particular, if $T' \neq T$ is a different cusp, then*

$$\phi(\partial \mathbb{T}_\epsilon(T')) = \partial \mathbb{T}_\epsilon(\phi(T')).$$

In the setting of finite-volume manifolds, conclusions (a) and (b) are due to Hodgson and Kerckhoff [27]. Conclusions (c) and (d) are due to Brock and Bromberg [11, Theorem 6.2 and Lemma 6.17], who construct the reverse diffeomorphism ϕ^{-1} under the hypothesis that the core curve γ is sufficiently short. (When μ is sufficiently long, this hypothesis will be satisfied by (b).) See Magid [31, Section 4] for a unified treatment of all four statements in this version of the theorem.

Conclusions (c) and (d) can be fruitfully combined, as follows. Let $\mathbb{T}_\epsilon(X)$ denote the union of all the ϵ–thin cusp neighborhoods of X, ie, the cusp components of $X_{<\epsilon}$. Similarly, let $\mathbb{T}_\epsilon(X(\mu))$ denote the union of $\mathbb{T}_\epsilon(\gamma)$ and all the ϵ–thin cusp neighborhoods

in $X(\mu)$. Then, by the drilling theorem, we have

(2-2) $$\phi\colon X \smallsetminus \mathbb{T}_\epsilon(X) \to X(\mu) \smallsetminus \mathbb{T}_\epsilon(X(\mu)).$$

a J–bilipschitz diffeomorphism between compact manifolds.

Remark 2.7 In the forthcoming arguments, particularly in Sections 3 and 4, we will refer to $X \smallsetminus \mathbb{T}_\epsilon(X)$ and $X(\mu) \smallsetminus \mathbb{T}_\epsilon(X(\mu))$ as the *thick parts* of X and $X(\mu)$, respectively. This usage is somewhat abusive, for instance since the manifold $X(\mu)$ may contain other ϵ–short geodesics besides γ. However, any "extra" ϵ–thin regions of X or $X(\mu)$ will not affect the arguments in any way, rendering the abuse relatively harmless.

In practice, we will always use Theorem 2.6 in the setting where X is a finite-volume hyperbolic manifold with two or more cusps. Thus, since every filled manifold $M = X(\mu)$ has one or more cusps remaining, Culler and Shalen's recent theorem [15] implies that $\epsilon = 0.292$ is a Margulis number for both X and every hyperbolic $X(\mu)$. Unless stated otherwise (eg, in the proof of Theorem 1.6), we will always work with the value $\epsilon = 0.29$.

One immediate consequence of Theorem 2.6 is the following fact, which we will use repeatedly.

Lemma 2.8 Let α be a homotopically essential closed curve in X, or an essential arc whose endpoints are on $\partial \mathbb{T}_\epsilon(X)$. Let g_α be a shortest geodesic in the free homotopy class of α in $X \smallsetminus \mathbb{T}_\epsilon(X)$, where the endpoints of α are allowed to slide along $\partial \mathbb{T}_\epsilon(X)$ if α is an arc.

Choose $J > 1$ and a slope μ that satisfies Theorem 2.6. Let $\bar{\alpha} = \phi(\alpha)$ be a curve or arc in $X(\mu)$, where ϕ is the J–bilipschitz diffeomorphism guaranteed by Theorem 2.6. Let $\bar{g}_\alpha$ be a shortest geodesic in the free homotopy class of $\bar{\alpha}$ in $X(\mu) \smallsetminus \mathbb{T}_\epsilon X(\mu)$. Then,

(2-3) $$\frac{1}{J} \cdot \ell(g_\alpha) \le \ell(\bar{g}_\alpha) \le J \cdot \ell(g_\alpha).$$

When the inequality (2-3) holds, we will say that the lengths of g_α and $\bar{g}_\alpha$ are J–*related*.

The reason for the non-unique terminology "a shortest geodesic" is that α can be, for instance, a peripheral curve in $\partial \mathbb{T}_\epsilon$. In this case, g_α is a Euclidean geodesic.

Note that there is no reason to expect that the J–bilipschitz diffeomorphism ϕ maps the geodesic g_α to the geodesic $\bar{g}_\alpha$. Nevertheless, the estimate on the geodesic lengths still holds.

Proof of Lemma 2.8   To prove the upper bound on $\ell(\bar{g}_\alpha)$, suppose that $\alpha = g_\alpha$ is already geodesic. Then, by Theorem 2.6, the arc $\bar{\alpha} = \phi(g_\alpha)$ has length at most $J \cdot \ell(g_\alpha)$. Since the geodesic $\bar{g}_\alpha$ can be no longer than $\bar{\alpha}$, the same upper bound applies: $\ell(\bar{g}_\alpha) \leq J \cdot \ell(g_\alpha)$. By the same argument, starting with the geodesic $\bar{g}_\alpha$ and applying the J–bilipschitz diffeomorphism ϕ^{-1}, we obtain $\ell(g_\alpha) \leq J \cdot \ell(\bar{g}_\alpha)$, which is exactly what is needed to complete the proof.   □

2.3   Geometric estimates and core tunnels

In the remaining sections of the paper, we will apply Theorems 2.3 and 2.6 to unknotting tunnels in cusped hyperbolic 3–manifolds, and more generally, to the core tunnels as in Definition 2.2. In order to do this, we need information about the core tunnels before filling.

Lemma 2.9   *Suppose X is a finite-volume hyperbolic 3–manifold, with a genus-2 Heegaard surface Σ. Suppose that $\sigma \subset X$ is the core tunnel for a compression body of $X \smallsetminus \Sigma$, whose endpoints are on distinct cusp tori T and T'. Then:*

(a)   *X admits a hyper-elliptic involution ψ, which preserves Σ up to isotopy.*

(b)   *The hyperbolic isometry isotopic to ψ fixes a hyperbolic geodesic isotopic to σ.*

Proof   When σ is an unknotting tunnel, this statement is due to Adams [1, Lemma 4.6], and his proof carries through verbatim to core tunnels that connect distinct cusps. We recall the argument briefly. Each compression body C_i in the complement of Σ admits a hyper-elliptic involution, and the restriction of these involutions to Σ is unique up to isotopy (Bleiler and Moriah [10]). Thus the involutions of C_1 and C_2 can be glued together to obtain an involution ψ on X preserving Σ setwise.

The hyper-elliptic involution ψ, restricted to Σ, preserves the isotopy class of every simple closed curve; separating curves on Σ are preserved with orientation. The compression disk of C_1 dual to σ separates T from T', hence its boundary is preserved with orientation (up to isotopy). As a result, ψ can be chosen to fix σ pointwise.

By Mostow–Prasad rigidity, ψ is homotopic to a hyperbolic isometry. This order-2 isometry of X lifts to an elliptic isometry of $\mathbb{H}^3$ that preserves the endpoints of a lift of σ, hence preserves the geodesic $\widetilde{g}_\sigma$ connecting these endpoints. Now, by the work of Waldhausen [47] and Tollefson [46], two homotopic involutions of X are connected by a continuous path of involutions. Thus the fixed-point set of ψ is isotopic to the fixed-point set of the isometry, hence σ is isotopic to the geodesic g_σ in its homotopy class.   □

In the remainder of the paper, we will assume that every core tunnel σ connecting distinct cusps of X is already a geodesic. We will be studying the behavior of this geodesic in the compact, thick part $X \smallsetminus \mathbb{T}_\epsilon(X)$. See Figure 2 for two lifts of this geodesic to $\mathbb{H}^3$.

We may use the geodesic σ to carefully construct an arc τ that will become an associated tunnel in a Dehn filling of X. The point of the following construction is to make Figure 1 precise. In the introduction, we stated that an associated tunnel in the filled manifold runs from the cusp, along σ, then once around a longitude, then back along σ to the cusp. However, this arc as described is not embedded. In the following definition, we push the new tunnel off σ carefully to ensure the result is embedded. We then prove the claim from the introduction that this arc becomes an unknotting tunnel under Dehn filling.

Definition 2.10 Let X be a be an orientable hyperbolic 3–manifold that has two cusps (denoted T and K), and tunnel number one. Let σ be an unknotting tunnel of X, isotoped to be a geodesic. Let μ be a Dehn filling slope on T, and λ be a longitude for μ. Choose any $\epsilon > 0$ that is a Margulis number for X (for example, $\epsilon = 0.29$). Let λ_ϵ be a closed curve representing λ on the horospherical torus $\partial \mathbb{T}_\epsilon(T)$, which passes through the endpoint of σ on $\partial \mathbb{T}_\epsilon(T)$.

Let Q be an embedded quadrilateral contained in a tubular neighborhood of σ, whose top side is on λ_ϵ and whose bottom side is on $\partial \mathbb{T}_\epsilon(K)$, and whose remaining sides, call them s_1 and s_2, run parallel to σ on the boundary of the tubular neighborhood.

We define the *tunnel arc associated to σ and μ*, denoted $\tau(\sigma, \mu)$, to be the embedded arc $s_1 \cup (\lambda_\epsilon \smallsetminus Q) \cup s_2$. This three-part arc is sketched in the right panel of Figure 1. If σ is oriented from K to T, then $\tau(\sigma, \mu)$ is homotopic to $\sigma \cdot \lambda_\epsilon \cdot \sigma^{-1}$.

The hyperbolic geodesic $\sigma \subset X$, the Euclidean geodesic $\lambda_\epsilon \subset X$, and the corresponding geodesics $\overline{\sigma}, \overline{\lambda}_\epsilon \subset X(\mu)$ are depicted in Figures 2 and 3.

As the name suggests, the tunnel arc $\tau(\sigma, \mu)$ will become an unknotting tunnel in $X(\mu)$.

Theorem 2.11 *Let X be an orientable hyperbolic 3–manifold that has two cusps (denoted T and K), and tunnel number one. Let σ be an unknotting tunnel for X. Choose $\epsilon = 0.29$ and $J > 1$, and let μ be any Dehn filling slope on T that is sufficiently long for Theorem 2.6 to ensure a J–bilipschitz diffeomorphism $\phi \colon X \smallsetminus \mathbb{T}_\epsilon(X) \to X(\mu) \smallsetminus \mathbb{T}_\epsilon(X(\mu))$. Then:*

   (a) *If $\tau(\sigma, \mu)$ is the tunnel arc associated to σ and μ, as in Definition 2.10, then $\overline{\tau}(\sigma, \mu) = \phi(\tau(\sigma, \mu))$ is an unknotting tunnel of $X(\mu)$.*

(b) *Suppose, in addition, that $\ell(\mu) > 6\pi$ and $\ell(\lambda) > 6$ on a maximal cusp about T. Then every unknotting tunnel of $X(\mu)$ corresponds to a genus-2 Heegaard surface $\Sigma \subset X$.*

(c) *Suppose that $\ell(\mu) > 6\pi$, that $\ell(\lambda) > 6$, and that both cusps of X lie on the same side of every genus-2 Heegaard surface $\Sigma \subset X$. Then every unknotting tunnel of $X(\mu)$ is isotopic to $\overline{\tau}(\sigma, \mu) = \phi(\tau(\sigma, \mu))$ for some unknotting tunnel σ of X.*
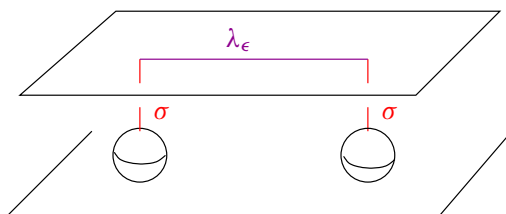

Figure 2: σ is the geodesic unknotting tunnel in the unfilled manifold, X. λ_ϵ is a geodesic representative of the longitude λ in the boundary of the ϵ–thin cusp neighborhood. We take $\epsilon = 0.29$ throughout. Picture in the universal cover.
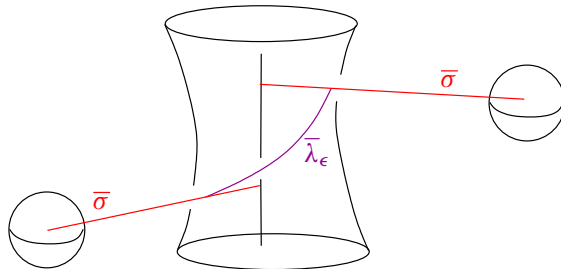

Figure 3: $\overline{\sigma}$ is the geodesic in the homotopy class of the image of σ in the filled manifold $X(\mu)$. The curve $\overline{\lambda}_\epsilon$ is the shortest curve along the ϵ–Margulis tube between points where $\overline{\sigma}$ meets the tube on its boundary. Picture in the universal cover.

Proof Let $\tau = \phi(\tau(\sigma, \mu)) \subset X(\mu)$. We will prove that τ is an unknotting tunnel for $X(\mu)$ by showing that $X(\mu) \smallsetminus \tau$ is homeomorphic to $X \smallsetminus \sigma$, which is a handlebody by hypothesis.

Let λ_ϵ, Q, s_1 and s_2 be as in Definition 2.10. Then observe that $\phi(\lambda_\epsilon)$ is a longitude for the solid torus V added during Dehn filling, hence is isotopic to the core curve γ of V. As a result, $X(\mu) \smallsetminus \phi(\lambda_\epsilon) \cong X(\mu) \smallsetminus \gamma \cong X$. Similarly, s_1 is isotopic to σ. Thus

$$X(\mu) \smallsetminus \phi(s_1 \cup \lambda_\epsilon) \cong \phi(X \smallsetminus s_1) \cong X \smallsetminus s_1 \cong X \smallsetminus \sigma,$$

and $X \smallsetminus \sigma$ is a genus-2 handlebody. Finally, note that the arc on top of Q, namely $(\lambda_\epsilon \cap Q)$, can be replaced with s_2 without altering the complement. This replacement can be accomplished by continuously sliding one endpoint of $(\lambda_\epsilon \cap Q)$ along s_1, turning the "eyeglass" $\lambda_\epsilon \cup \sigma$ into the embedded arc $\tau(\sigma, \mu)$. Thus

$$X(\mu) \smallsetminus \phi(s_1 \cup (\lambda_\epsilon \smallsetminus Q) \cup s_2) \cong X(\mu) \smallsetminus \phi(s_1 \cup \lambda_\epsilon) \cong X \smallsetminus \sigma,$$

proving (a).

Statement (b) follows from Theorem 2.3. Let τ be an unknotting tunnel of $X(\mu)$, and let $\Sigma \subset X(\mu)$ be the Heegaard surface associated to τ. By Theorem 2.3, the core curve γ is isotopic into Σ. Furthermore, isotoping γ off Σ into one of the pieces separated by Σ turns Σ into a Heegaard surface for $X = X(\mu) \smallsetminus \gamma$.

To prove (c), let $\Sigma \subset X$ be the Heegaard surface guaranteed by (b). By hypothesis, both cusps of X must lie on the same side of Σ. Thus $\Sigma \subset X$ has a handlebody on one side and a compression body on the other side. Hence, the core tunnel σ of the compression body in $X \smallsetminus \Sigma$ is an unknotting tunnel for X.

It remains to check that τ is isotopic to $\tau(\sigma, \mu)$ as in Definition 2.10. This is true because the Heegaard surface defined by $\tau(\sigma, \mu)$ is the boundary of a regular neighborhood of $\lambda_\epsilon \cup Q \cup \partial \mathbb{T}_\epsilon(K)$, which is the same Heegaard surface Σ defined by τ. Thus, since τ and $\tau(\sigma, \mu)$ are core tunnels for the same compression body in $X(\mu) \smallsetminus \Sigma$, they must be isotopic. $\qquad\square$

Remark 2.12 The construction in Definition 2.10 involved numerous choices. There are many longitudes for μ, many representatives of λ on $\partial \mathbb{T}_\epsilon(T)$, and many choices for the quadrilateral Q (some of which are twisted). The argument above implies that all of these choices are immaterial: up to isotopy in $X(\mu)$, they all produce the same unknotting tunnel.

Remark 2.13 As we mentioned in Section 1.1, the work of Lustig and Moriah [30], Maher [32], and Scharlemann and Tomova [41] severely restricts the "generic" possibilities for Σ. More precisely, suppose that the two-cusped manifold X is constructed by gluing a genus-2 handlebody C to a compression body C' via some mapping class $\varphi \in \operatorname{Mod}(\Sigma)$. If φ is chosen by a random walk in the generators of $\operatorname{Mod}(\Sigma)$, Maher showed that with probability approaching 1, the Heegaard splitting has curve complex distance $d(\Sigma) \geq 5$: see [32, Theorem 1.1]. Similarly, Lustig and Moriah showed that Heegaard splittings satisfying $d(\Sigma) \geq 5$ are generic in the sense of Lebesgue measure on the projective measured lamination space $\mathcal{PML}(\Sigma)$: see [30]. In either case, once we know that $d(\Sigma) \geq 5$, a result of Scharlemann and Tomova implies that Σ is the unique minimal-genus Heegaard surface of X [41, Corollary on p 594].

Thus, for a generic Dehn filling, Theorem 2.11 implies the filled manifold $X(\mu)$ has a unique unknotting tunnel τ, associated to the tunnel σ of X.

3 The length of unknotting tunnels

The main goal of this section is to write down a proof of Theorem 1.4, which estimates the length of an unknotting tunnel $\tau \subset X(\mu)$ up to additive error. Unfortunately, the clean statement of Theorem 1.4 relies on a number of technical estimates about various related lengths in X and $X(\mu)$. We collect these technical estimates in Sections 3.1 and 3.2. Then, in Section 3.3, we complete the proof of Theorem 1.4.

3.1 Length and waist size

As above, let σ be an unknotting tunnel of a two-cusped manifold X. We know that σ is a geodesic arc that runs between the cusps about K and T. The first step toward estimating the length of an unknotting tunnel $\tau(\mu, \sigma)$ of $X(\mu)$ is estimating the length of σ itself. The length of σ turns out to be closely related to the notion of waist size, defined and explored by Adams [2; 3].

Definition 3.1 Let H be a horocusp in a hyperbolic 3–manifold M (see Definition 1.1). Then the *waist size* $w(H)$ is defined to be the length of the shortest non-trivial curve on ∂H. This shortest curve is necessarily a Euclidean geodesic on ∂H.

Adams proved the following statements about the waist size of a two-cusped manifold:

(A) Given any choice of disjointly embedded horocusps H_K and H_T, such that the smaller of the two waist sizes is w, the length of an unknotting tunnel σ relative to H_K and H_T is $\ell(\sigma) < \ln(4) - 2\ln(w)$. This is [1, Theorem 4.4].

(B) If this choice of cusp neighborhoods is *maximal*, in the sense that neither of H_K or H_T can be expanded while keeping them disjointly embedded, then each of H_K and H_T, has waist size at least 1. This universal estimate is [2, Lemma 2.4].

Facts (A) and (B) have the following consequence.

Lemma 3.2 *In the two-cusped, tunnel number one manifold X, let N_K be a maximal neighborhood about cusp K, expanded until it bumps into itself. Then the waist size of N_K is $1 \le w(N_K) < 4$. The same estimate holds for the other cusp T of X.*

Proof Let H_K^1 be a cusp neighborhood about K whose waist size is exactly 1. By fact (B), this neighborhood is contained in N_K, therefore embedded. Similarly, let H_T^1 be a horocusp about T whose waist size is exactly 1.

We claim that N_K is disjoint from H_T^1. This is because a maximal choice of neighborhoods can be obtained as follows: expand K until it bumps into itself, obtaining N_K. Then, expand T until it bumps into either itself or K; in either case, the resulting horocusp about T will have waist size at least 1, hence contains H_T^1. Therefore, H_T^1 is disjoint from N_K.

Next, we claim that the horospherical tori ∂N_K and ∂H_K^1 are at hyperbolic distance

$$(3\text{-}1) \qquad\qquad\qquad d(\partial N_K, \partial H_K^1) < \ln 4.$$

Here is why. On the one hand, the length of σ relative to N_K and H_T^1 is at least 0, since these cusp neighborhoods are disjoint. On the other hand, the length of σ relative to H_K^1 and H_T^1 is less than $\ln 4$, by fact (A). The difference between these lengths is exactly the hyperbolic distance $d(\partial N_K, \partial H_K^1)$, which must be less than $\ln 4$. Similarly, $d(\partial N_T, \partial H_T^1) < \ln 4$.

Consider the waist size of N_K. This is at least 1 by fact (B). Also, since $w(H_K^1) = 1$ and waist size grows exponentially with hyperbolic distance, (3-1) implies that $w(N_K) < 4$. $\qquad\square$

Facts (A) and (B) also allow us to estimate the length of σ in the thick part of X.

Lemma 3.3 *In the two-cusped manifold X, let N_K be a maximal horocusp about K, expanded until it bumps into itself. Let N_T be a maximal horocusp about T, expanded until it bumps into itself. For $\epsilon = 0.29$, let $\mathbb{T}_\epsilon(K)$ and $\mathbb{T}_\epsilon(T)$, respectively, be the ϵ–thin neighborhoods of those cusps. (See Definition 2.4.)*

Then, in the thick portion of X, the length of σ relative to $\mathbb{T}_\epsilon(K)$ and $\mathbb{T}_\epsilon(T)$ satisfies

$$(3\text{-}2) \qquad\qquad\qquad 2.46 < \ell(\sigma_\epsilon) < 3.86.$$

Relative to the (possibly overlapping) maximal cusps N_K and N_T, the length of σ is

$$(3\text{-}3) \qquad\qquad\qquad -\ln 4 < \ell(\sigma_{\max}) < \ln 4.$$

Here, we are using the convention that the length of σ in $X \smallsetminus (N_K \cup N_T)$ counts positively, and the length of σ in $N_K \cap N_T$ counts negatively.

The length convention in (3-3) is natural, in the following sense. If a horoball is expanded by distance d, the length of a geodesic running perpendicularly into that horoball decreases by distance d. This natural convention requires negative lengths for overlapping horoballs.

Proof   As in Lemma 3.2, let H_K^1 and H_T^1 be cusp neighborhoods about K and T, respectively, whose waist sizes are exactly 1. By Adams' fact (B), these horocusps are disjointly embedded in X. Thus the length of an unknotting tunnel σ relative to these horocusps is at least 0. On the other hand, by (A), the length σ relative to these horocusps is less than $\ln 4$.

Now, consider what happens when we replace H_K^1 by $\mathbb{T}_\epsilon(K)$ and H_T^1 by $\mathbb{T}_\epsilon(T)$. By Lemma A.2 in the appendix, the waist size of $\partial \mathbb{T}_\epsilon$ is

$$(3\text{-}4) \qquad w(\mathbb{T}_\epsilon(K)) = w(\mathbb{T}_\epsilon(T)) = 2 \sinh(0.145) = 0.29101\ldots$$

Because waist size grows exponentially with length, the length of σ will increase by a distance of $-\ln(2 \sinh 0.145)$ as H_K^1 is replaced by $\mathbb{T}_\epsilon(K)$. Replacing H_T^1 by $\mathbb{T}_\epsilon(T)$ has the same effect. Thus the length of σ relative to $\mathbb{T}_\epsilon(K)$ and $\mathbb{T}_\epsilon(T)$ satisfies

$$2.468\ldots = -2 \ln(2 \sinh 0.145) < \ell(\sigma_\epsilon) < \ln(4) - 2 \ln(2 \sinh 0.145) = 3.855\ldots.$$

To prove (3-3), we begin with disjoint cusp neighborhoods N_T and H_K^1. By facts (A) and (B), the length of σ relative to these disjoint horocusps is at least 0 and less than $\ln 4$. As we replace H_K^1 by the larger cusp neighborhood N_K, the length of σ can only become smaller, hence is still bounded above by $\ln 4$. In fact, as we replace H_K^1 by N_K, the length of σ will decrease by precisely $d(\partial N_K, \partial H_K^1)$, which is less than $\ln 4$ by equation (3-1). Thus $\ell(\sigma)$ is bounded below by $-\ln 4$.   □

3.2   Estimating a few related quantities

The next several lemmas involve comparisons between certain geometric measurements in X and those of $X(\mu)$.

Condition 3.4   For the remainder of this section, we set $\epsilon = 0.29$ and $J = 1.1$. We also assume throughout that the Dehn filling slope μ on T is long enough for Theorem 2.6 to guarantee a J–bilipschitz diffeomorphism $\phi: X \smallsetminus \mathbb{T}_\epsilon \to X(\mu) \smallsetminus \mathbb{T}_\epsilon$.

Lemma 3.5   *Assume that μ satisfies Condition 3.4. In the two-cusped manifold X, let $x := d(\partial N(T), \partial \mathbb{T}_\epsilon(T))$, where $N(T)$ is the maximal horocusp about T and $\mathbb{T}_\epsilon(T)$ is the ϵ–thin horocusp about T. In the filled manifold $X(\mu)$, let $s := d(\partial N_\mu(K), \partial \mathbb{T}_\epsilon(\gamma))$, where $N_\mu(K) \subset X(\mu)$ is the maximal horocusp about the remaining cusp K, and $\mathbb{T}_\epsilon(\gamma)$ is the ϵ–thin Margulis tube. Then*

$$(3\text{-}5) \qquad\qquad\qquad\qquad x < 2.621.$$

Furthermore, s and x are equal up to additive error:

$$(3\text{-}6) \qquad\qquad\qquad\qquad -2 < s - x < 2.02.$$

Proof Let us label several more lengths in X and $X(\mu)$. In the unfilled manifold X, define $y := d(\partial N(K), \partial \mathbb{T}_\epsilon(K))$. In the filled manifold $X(\mu)$, define $t := d(\partial N_\mu(K), \partial \mathbb{T}_\epsilon(K))$, the distance between an ϵ–sized cusp and a maximal cusp about K. These definitions are depicted in Figure 4.
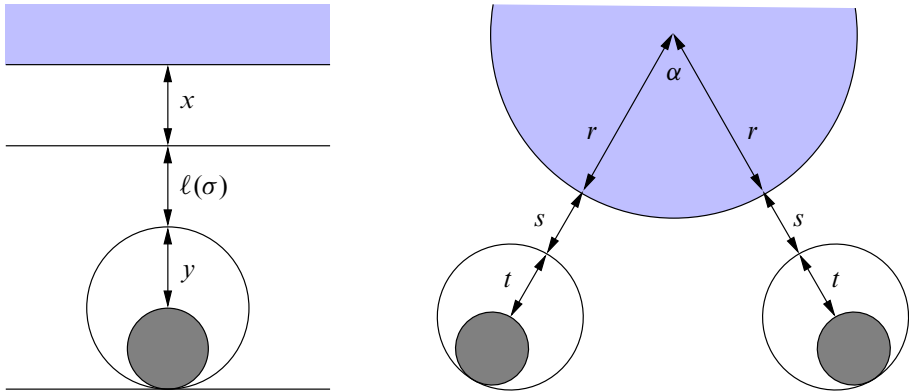

Figure 4: Notation for Section 3. The ϵ–thin parts of the two manifolds are shaded. The J–bilipschitz diffeomorphism of Theorem 2.6 maps the unshaded area on the left to the unshaded area on the right.

Note that the waist sizes of $N(K)$ and $\mathbb{T}_\epsilon(K))$, as well as of $N(T)$ and $\mathbb{T}_\epsilon(T))$, are bounded by Lemma 3.2 and equation (3-4). Thus

$$(3\text{-}7) \qquad y = d(\partial N(K), \partial \mathbb{T}_\epsilon(K)) < \ln\left(\frac{4}{2 \sinh 0.145}\right) = 2.6206\ldots,$$

and similarly for x. This proves (3-5).

In the terminology of Lemma 3.3, we now have

$$\ell(\sigma_\epsilon) = x + y + \ell(\sigma_{\max}).$$

Similarly, the geodesic $\bar{\sigma}_\epsilon$ in the homotopy class of $\phi(\sigma_\epsilon)$ has length

$$\ell(\bar{\sigma}_\epsilon) = s + t.$$

By Lemma 2.8, the lengths of σ_ϵ and $\bar{\sigma}_\epsilon$ are J–related, for $J = 1.1$:

$$(3\text{-}8) \qquad \tfrac{10}{11}(\ell(\sigma_{\max}) + x + y) \le s + t \le \tfrac{11}{10}(\ell(\sigma_{\max}) + x + y).$$

Observe that the shortest geodesic $h \subset X$ from $\mathbb{T}_\epsilon(K)$ back to $\mathbb{T}_\epsilon(K)$ has length exactly $2y$. This is because an expanding horocusp about K will become maximal, and bump into itself, precisely at the midpoint of a shortest geodesic. Similarly, the

shortest geodesic $\bar{h} \subset X(\mu)$ from $\mathbb{T}_\epsilon(K)$ back to $\mathbb{T}_\epsilon(K)$ has length exactly $2t$. Thus the lengths of h and $\bar{h}$ are also J–related:

$$\tag{3-9} \tfrac{10}{11} y \leq t \leq \tfrac{11}{10} y.$$

(Note estimate (3-9) will be true even if $\bar{h}$ is in a different homotopy class from $\phi(h)$, by applying the assumptions that h and $\bar{h}$ are both *shortest*, as in the proof of Lemma 2.8.)

We are now ready to prove the upper and lower bounds of (3-6). By equation (3-8):

$$(\ell(\sigma_{\max}) + x + y) - \tfrac{1}{11}\ell(\sigma_\epsilon) \leq s + t \leq (\ell(\sigma_{\max}) + x + y) + \tfrac{1}{10}\ell(\sigma_\epsilon)$$

$$\underbrace{(y - t)}_{\text{use (3-9)}} + \ell(\sigma_{\max}) - \tfrac{1}{11}\ell(\sigma_\epsilon) \leq s - x \leq \underbrace{(y - t)}_{\text{use (3-9)}} + \ell(\sigma_{\max}) + \tfrac{1}{10}\ell(\sigma_\epsilon)$$

$$\underbrace{(y - \tfrac{11}{10}y)}_{\text{use (3-7)}} + \underbrace{\ell(\sigma_{\max})}_{\text{use (3-3)}} - \underbrace{\tfrac{1}{11}\ell(\sigma_\epsilon)}_{\text{use (3-2)}} \leq s - x \leq \underbrace{(y - \tfrac{10}{11}y)}_{\text{use (3-7)}} + \underbrace{\ell(\sigma_{\max})}_{\text{use (3-3)}} + \underbrace{\tfrac{1}{10}\ell(\sigma_\epsilon)}_{\text{use (3-2)}}$$

$$\underbrace{-\tfrac{1}{10}(2.621) - \ln 4 - \tfrac{1}{11}(3.86)}_{= -1.9993\ldots} < s - x < \underbrace{\tfrac{1}{11}(2.621) + \ln 4 + \tfrac{1}{10}(3.86)}_{= 2.0105\ldots}$$

Therefore, $-2 < s - x < 2.02$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Let $\gamma \subset X(\mu)$ be the geodesic core of the solid torus V added during Dehn filling. Let $\bar{\sigma}$ be the geodesic from $\mathbb{T}_\epsilon(K)$ to γ that contains $\bar{\sigma}_\epsilon$ and extends into the thin part of $X(\mu)$ all the way to γ. There is an arc ρ_0 that follows $\bar{\sigma}$ to the core γ and runs along γ for half the length of γ, and a similar arc ρ_0' that follows $\bar{\sigma}$ and runs halfway along γ in the other direction. Let ρ and ρ' be the geodesics in the homotopy classes of ρ_0 and ρ_0', respectively.

Recall that if $\bar{\sigma}$ is oriented toward γ, then τ is homotopic to $\bar{\sigma} \cdot \gamma \cdot \bar{\sigma}^{-1}$. Equivalently, if ρ and ρ' are oriented toward γ, then τ is homotopic to $\rho' \cdot \rho^{-1}$. The geodesic in this homotopy class is denoted g_τ. Figure 5 depicts lifts of $\bar{\sigma}$, γ, ρ, ρ' and g_τ to the universal cover $\mathbb{H}^3$.

The next lemma estimates the radius r of the Margulis tube $\mathbb{T}_\epsilon(\gamma)$, as well as the distances between $\bar{\sigma}$ and ρ (similarly, $\bar{\sigma}$ and ρ') along $\partial\mathbb{T}_\epsilon(\gamma)$.

Lemma 3.6 *Assume that the Dehn filling slope μ is long enough that its representative μ_ϵ satisfies $\ell(\mu_\epsilon) \geq 10$, and also that μ is long enough to satisfy Theorem 2.6; in particular, the normalized length of μ is $L(\mu) \geq 4\sqrt{2}\pi$. Then the radius of the Margulis tube $\mathbb{T}_\epsilon(\gamma)$ is*

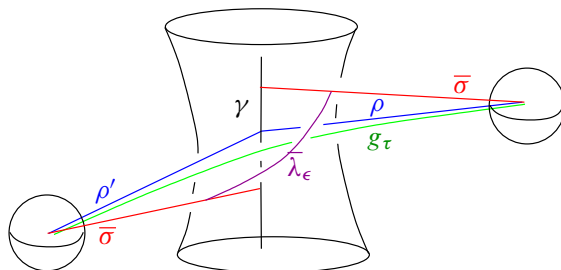$$\tag{3-10} r \geq \sinh^{-1}\left(\frac{\ell(\mu_\epsilon)}{2.2\pi}\right) > 1.16.$$

Figure 5: A schematic picture of the lifts of γ, $\bar{\sigma}$, ρ and ρ' to the universal cover. The arc g_τ is the geodesic in the homotopy class of the tunnel τ.

The distance along ρ between the vertex $v = \rho \cap \rho'$ and $\rho \cap \partial \mathbb{T}_\epsilon(\gamma)$ satisfies

$$(3\text{-}11) \qquad\qquad r \leq d(\rho \cap \gamma, \rho \cap \partial \mathbb{T}_\epsilon(\gamma)) \leq r + h,$$

where

$$h < 1.4 \times 10^{-6} \quad \text{and} \quad h \to 0 \quad \text{as} \quad L(\mu) \to \infty.$$

Finally, the distance on $\partial \mathbb{T}_\epsilon(\gamma)$ from $\bar{\sigma} \cap \partial \mathbb{T}_\epsilon(\gamma)$ to $\rho \cap \partial \mathbb{T}_\epsilon(\gamma)$ satisfies

$$(3\text{-}12) \qquad d(\bar{\sigma} \cap \partial \mathbb{T}_\epsilon(\gamma), \rho \cap \partial \mathbb{T}_\epsilon(\gamma)) < 0.02 e^{-r} + h < 0.0063,$$

and similarly for the distance from $\bar{\sigma} \cap \partial \mathbb{T}_\epsilon(\gamma)$ to $\rho' \cap \partial \mathbb{T}_\epsilon(\gamma)$.

Proof For (3-10), choose μ so that its length on $\partial \mathbb{T}_\epsilon(T)$ is $\ell(\mu_\epsilon) \geq 10$. Then the corresponding curve $\bar{\mu}_\epsilon \subset X(\mu)$ is the circumference of a meridian disk of the Margulis tube $\mathbb{T}_\epsilon(\gamma)$, and has length $\ell(\bar{\mu}_\epsilon) = 2\pi \sinh(r)$. By Lemma 2.8, the lengths of μ_ϵ and $\bar{\mu}_\epsilon$ are J–related. Hence,

$$2\pi \sinh(r) = \ell(\bar{\mu}_\epsilon) \geq \ell(\mu_\epsilon)/1.1.$$

Now, inequality (3-10) follows by solving for r:

$$r \geq \sinh^{-1}\Big(\frac{\ell(\mu_\epsilon)}{2.2\pi}\Big) \geq \sinh^{-1}\Big(\frac{10}{2.2\pi}\Big) = 1.1649\ldots$$

For (3-11), observe that the geodesics γ, ρ, and ρ' form an isosceles $1/3$ ideal triangle Δ, whose axis of symmetry is $\bar{\sigma}$. Lift this triangle to $\mathbb{H}^3$, and label the two material vertices v and v' (these vertices project to the same point in $X(\mu)$, but are distinct in $\mathbb{H}^3$). There is a single horocycle C about the ideal vertex of Δ that passes through v and v'. See Figure 6.

Note that by Theorem 2.6(b), the distance from v to v' along the lift $\tilde{\gamma}$ of γ is

$$\ell(\gamma) \leq \tfrac{1}{8\pi}.$$

By Lemma A.2, the distance from v to v' along the horocycle C is:

(3-13) $$p = 2\sinh\frac{\ell(\gamma)}{2} \leq 2\sinh\frac{1}{16\pi} = 0.03978\ldots$$

Thus, by the triangle inequality, the maximum distance by which $\tilde{\gamma}$ deviates from C is

(3-14) $$0 < h \leq (\sinh\tfrac{1}{16\pi}) - \tfrac{1}{16\pi} = 1.312\ldots \times 10^{-6},$$

where this tiny deviation approaches 0 as $L(\mu) \to \infty$ and $\ell(\gamma) \to 0$.

To derive the first inequality of (3-11), observe that the point $\rho \cap \partial\mathbb{T}_\epsilon(\gamma)$ is by definition at distance r from the geodesic γ. Thus the closest point of γ is at distance r, and the vertex $\rho \cap \gamma$ is at distance at $\ell > r$. For the second inequality of (3-11), observe that $\rho \cap \partial\mathbb{T}_\epsilon(\gamma)$ is closer to C than $\bar{\sigma} \cap \partial\mathbb{T}_\epsilon(\gamma)$, which is at distance $r + h$ from C.
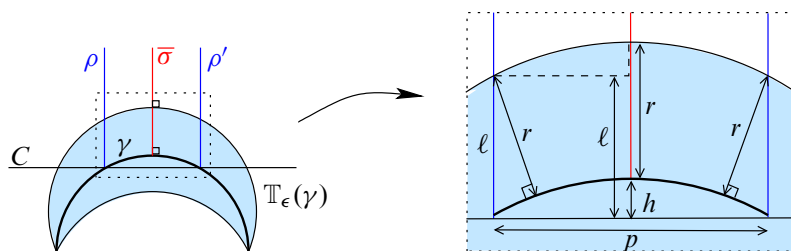

Figure 6: (In universal cover) The triangle $\Delta(\gamma\rho\rho')$ intersects the Margulis tube $\mathbb{T}_\epsilon(\gamma)$ as shown. Right: zoomed in to show lengths.

Finally, to prove (3-12), note that the path β from $\bar{\sigma} \cap \partial\mathbb{T}_\epsilon$ to $\rho \cap \partial\mathbb{T}_\epsilon$ along $\partial\mathbb{T}_\epsilon(\gamma)$ has a length that can be computed as an arclength integral in the upper half-space model:

$$\ell(\beta) = \int_\beta ds = \int_\beta \frac{\sqrt{dx^2 + dy^2}}{y} < \int_\beta \frac{|dy|}{y} + \int_\beta \frac{|dx|}{y}.$$

In words, β is shorter than the union of a geodesic segment along $\bar{\sigma}$ (vertical in Figure 6, dashed) followed by a horocyclic segment (horizontal in Figure 6, dashed). The vertical segment has length $(r + h) - \ell$, where ℓ is the length of the arc of ρ from $\partial\mathbb{T}_\epsilon(\gamma)$ to γ, hence $\ell > r$. Thus the vertical segment has length at most h. The horizontal segment has length $e^{-\ell}p/2$, which is at most $e^{-r}p/2$. Thus,

$$\ell(\beta) = d_{\partial\mathbb{T}_\epsilon}(\bar{\sigma} \cap \partial\mathbb{T}_\epsilon, \rho \cap \partial\mathbb{T}_\epsilon) \leq p/2e^{-r} + h$$

$$< 0.02e^{-r} + h, \quad \text{by (3-13)}$$

$$< 0.02e^{-1.16} + 2 \times 10^{-6} \quad \text{by (3-10), (3-14)}$$

$$= 0.00627\ldots,$$

and similarly for the distance from $\bar{\sigma} \cap \partial\mathbb{T}_\epsilon$ to $\rho' \cap \partial\mathbb{T}_\epsilon$. $\qquad\square$

3.3 The triangle of g_τ

To complete the proof of Theorem 1.4, we need to carefully study the triangle $\Delta \subset X(\mu)$ whose sides are ρ, ρ' and g_τ. (See Figures 5 and 7.) The quantity that we seek is the length of g_τ relative to the maximal horocusp $N(K) \subset X(\mu)$.

Lemma 3.7 *Assume that the Dehn filling slope μ is long enough to satisfy Condition 3.4 and Lemma 3.6. As in Figure 4, let $r + s$ be the length of $\bar\sigma$ from $\partial \mathbb{T}_\epsilon(\gamma)$ to the maximal cusp $\partial N_\mu(K)$. Then,*

$$(3\text{-}15) \qquad\qquad r + s = \frac{\ell(g_\tau)}{2} - \frac{1}{2} \ln\left(\frac{1 - \cos\alpha}{2}\right) - h,$$

where α is the angle at the material vertex $v = \rho \cap \rho'$ of triangle $\Delta(\rho\rho'g_\tau)$, where the length $\ell(g_\tau)$ is measured relative to the maximal cusp $N_\mu(K)$, and where $0 < h < 2 \times 10^{-6}$ is the error term of Lemma 3.6.

Proof The triangle $\Delta(\rho\rho'g_\tau)$ is an isosceles $2/3$ ideal triangle. Thus, by Lemma A.3,

$$(3\text{-}16) \qquad\qquad \ell(\rho) + \ell(\rho') = \ell(g_\tau) - \ln\left(\frac{1 - \cos\alpha}{2}\right),$$

where the lengths $\ell(\rho) = \ell(\rho')$ are measured from v to the torus $\partial N_\mu(K)$.

Next, observe from Figure 6 that the geodesics ρ and $\bar\sigma$ fellow-travel from γ to the torus $\partial N_\mu(K)$, and that their lengths differ by exactly h:

$$(3\text{-}17) \qquad\qquad \ell(\rho) = \ell(\rho') = \ell(\bar\sigma) + h = (r + s) + h.$$

Plugging (3-17) into (3-16) completes the proof. $\square$

We will be able to estimate the length of g_τ once we obtain bounds on the length of a circle arc opposite g_τ.

Lemma 3.8 *Assume that the Dehn filling slope μ is long enough to satisfy Condition 3.4 and Lemma 3.6. Let $\Delta \subset X(\mu)$ be the triangle whose sides are ρ, ρ' and g_τ. Let $v = \rho \cap \rho'$ be the material vertex of Δ. Then the circle arc of radius r and angle α about v has length*

$$(3\text{-}18) \qquad\qquad e^{-x-0.31} \ell(\lambda) < \alpha \sinh r < e^{-x+0.1} \ell(\lambda),$$

where $x = d(\partial N(T), \partial \mathbb{T}_\epsilon(T))$ in X, as in Lemma 3.5, and λ is the shortest longitude for μ.

Proof The length of a circle arc of angle α and radius r is always equal to $\alpha \sinh r$. The main goal of the lemma is to estimate this quantity in terms of x and $\ell(\lambda)$.

By Lemma 3.5, the distance between the maximal cusp and the ϵ–sized cusp in X is

$$x = d(\partial N(T), \partial \mathbb{T}_\epsilon(T)) < 2.621.$$

Since the shortest longitude λ for μ has length $\ell(\lambda) \geq 1$ on the maximal cusp $N(T)$, the corresponding geodesic $\lambda_\epsilon \subset \partial \mathbb{T}_\epsilon(T)$ has length

$$\ell(\lambda_\epsilon) = e^{-x}\ell(\lambda) > e^{-2.621} = 0.07273\ldots.$$

Since the lengths of λ_ϵ and the corresponding curve $\bar{\lambda}_\epsilon \subset \mathbb{T}_\epsilon(\gamma)$ are J–related by Lemma 2.8, it follows that

(3-19) $$0.0661\ldots < \tfrac{10}{11}\ell(\lambda_\epsilon) \leq \ell(\bar{\lambda}_\epsilon) \leq \tfrac{11}{10}\ell(\lambda_\epsilon).$$

Let $C \subset \mathbb{T}_\epsilon(\gamma)$ be a curve in the homotopy class of $\bar{\lambda}_\epsilon$, constructed as follows. We take $C = C_1 \cup C_2$, where C_1 is the shortest segment on $\partial \mathbb{T}_\epsilon(\gamma)$ connecting points $\rho \cap \mathbb{T}_\epsilon(\gamma)$ and $\rho' \cap \mathbb{T}_\epsilon(\gamma)$, hence lying in the plane of the triangle $\Delta(\gamma \rho \rho')$. We take C_2 to be the shortest arc of intersection of $\partial \mathbb{T}_\epsilon(\gamma)$ and the plane containing the triangle $\Delta(\rho \rho' g_\tau)$. These arcs are shown schematically in Figure 7.
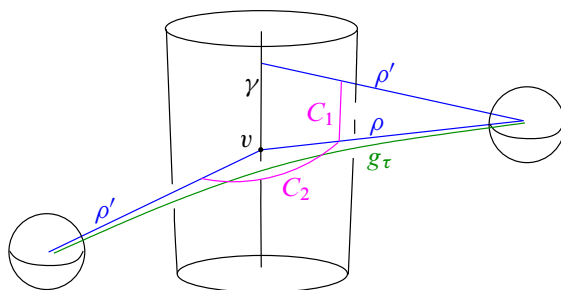

Figure 7: Schematic picture of the arcs C_1 and C_2 in the universal cover of $X(\mu)$

Note that by equation (3-12),

$$\ell(C_1) = d(\rho \cap \partial \mathbb{T}_\epsilon(\gamma), \rho' \cap \partial \mathbb{T}_\epsilon(\gamma)) < 2 \times 0.0063 = 0.0126.$$

Thus, by the triangle inequality and (3-19), the longer segment $C_2 = C \smallsetminus C_1$ has length

(3-20) $$\ell(C_2) > \ell(\bar{\lambda}_\epsilon) - 0.0126 > 0.8094\ell(\bar{\lambda}_\epsilon) > 0.7358\ell(\lambda_\epsilon) = 0.7358\, e^{-x}\ell(\lambda),$$

where the last inequality used the fact that $\ell(\bar{\lambda}_\epsilon)$ and $\ell(\lambda_\epsilon)$ are J–related.

Meanwhile, the points $\rho \cap \partial \mathbb{T}_\epsilon$ and $\rho' \cap \partial \mathbb{T}_\epsilon$ are closer to each other along C_2 than the full length of the longitude $\bar{\lambda}_\epsilon$. This is because a full longitude runs from one endpoint

of C_2 to an endpoint of C_1 in Figure 7, and the arc C_1 is vertical in the cylindrical coordinates on the Euclidean torus $\partial \mathbb{T}_\epsilon(\gamma)$. Thus,

$$(3\text{-}21) \qquad\qquad \ell(C_2) < \ell(\overline{\lambda}_\epsilon) < 1.1\ell(\lambda_\epsilon) = 1.1e^{-x}\ell(\lambda).$$

It remains to relate the length of C_2 to the circle arc of length $\alpha \sinh r$. By equation (3-11), the distance between the vertex $v = \rho \cap \rho'$ and the endpoints of C_2 is somewhere between r and $r + h$, where $h < 1.4 \times 10^{-6}$. Meanwhile, the midpoint of C_2 is at distance exactly r from v. Thus all of C_2 lies between circles of radius r and $r + h$. Hence, up to a multiplicative error of less than

$$e^h < 1.000002,$$

the length of C_2 is the same as the length $\alpha \sinh r$ of the circle arc of radius r. Combining this with (3-20) and (3-21), we conclude that

$$e^{-x-0.31}\ell(\lambda) < e^{-h} \cdot 0.7358e^{-x}\ell(\lambda) < \alpha \sinh r < e^h \cdot 1.1e^{-x}\ell(\lambda) < e^{-x+0.1}\ell(\lambda),$$

as desired. □

At this point, we are ready to complete the proof of Theorem 1.4. In fact, we have the following version of the theorem, which (mostly) quantifies how long μ needs to be in order to ensure that the estimates hold.

Theorem 3.9 *Let X be an orientable hyperbolic 3–manifold that has two cusps and an unknotting tunnel σ. Let T be one boundary torus of X. Let μ be a Dehn filling slope on T such that $\ell(\mu) > 138$, and such that μ is also long enough to satisfy Condition 3.4. Then the unknotting tunnel τ of $M = X(\mu)$ associated to σ satisfies*

$$2\ln \ell(\lambda) - 5.6 < \ell(g_\tau) < 2\ln \ell(\lambda) + 4.5,$$

where λ is the shortest longitude for μ. Here $\ell(\mu)$ and $\ell(\lambda)$ are lengths on a maximal cusp about T in X, and $\ell(g_\tau)$ is the length of the geodesic in the homotopy class of τ, relative to a maximal cusp in $M = X(\mu)$.

Proof Let τ be an unknotting tunnel of $M = X(\mu)$ associated to an unknotting tunnel σ for X. Our assumptions about the length of μ ensure that the estimates of Lemma 3.5 apply. In particular, by Lemma 3.5, $x < 2.621$. Thus the length of μ on $\partial \mathbb{T}_\epsilon(T)$ is

$$\ell(\mu_\epsilon) = e^{-x}\ell(\mu) > e^{-2.621} \cdot 138 > 10,$$

ensuring that Lemmas 3.6, 3.7 and 3.8 apply as well. The proof of the theorem will involve combining the results of Lemmas 3.5 through 3.8.

The quantity we seek is the length $\ell(g_\tau)$. Recall that equation (3-15) expresses $\ell(g_\tau)$ in terms of $r + s$, the angle α, and the (tiny) error term h. We can now use the previous lemmas to express $r + s$ in terms of $\ell(\lambda)$. One immediate consequence of equation (3-10) is that

$$e^{-r} < e^{-2.32}e^r,$$

which implies that

$$(3\text{-}22) \qquad e^{-0.11}e^r < (1 - e^{-2.32})e^r < e^r - e^{-r} = 2\sinh r < e^r.$$

Thus, by combining equation (3-22) with (3-18), we obtain

$$(3\text{-}23) \qquad \frac{2}{\alpha}e^{-x-0.31}\ell(\lambda) < 2\sinh r < e^r < e^{0.11} \cdot 2\sinh r < \frac{2}{\alpha}e^{-x+0.21}\ell(\lambda).$$

Now, equation (3-23) bounds e^r, hence r, in terms of $\ell(\lambda)$, x and the angle α. Inserting the estimate of (3-23) into (3-15), as we will do in the calculation below, will produce a term of the form $f(\alpha) = 2(1 - \cos\alpha)/\alpha^2$. By a derivative calculation, one observes that $f(\alpha)$ is strictly decreasing on the interval $(0, \pi]$. In our case, the angle α must indeed be positive, and is at most π. Thus,

$$(3\text{-}24) \qquad \frac{4}{\pi^2} = f(\pi) \leq \frac{2(1 - \cos\alpha)}{\alpha^2} < \lim_{\alpha \to 0} f(\alpha) = 1.$$

With these preliminaries out of the way, we can perform the final calculation. Taking the log of the first, middle and last terms of (3-23) yields:

$$\ln\ell(\lambda) - x - 0.31 + \ln\left(\tfrac{2}{\alpha}\right) < \quad r \quad < \ln\ell(\lambda) - x + 0.21 + \ln\left(\tfrac{2}{\alpha}\right)$$

$$\ln\ell(\lambda) + \underbrace{s - x}_{\text{use (3-6)}} - 0.31 + \ln\left(\tfrac{2}{\alpha}\right) < r + s < \ln\ell(\lambda) + \underbrace{s - x}_{\text{use (3-6)}} + 0.21 + \ln\left(\tfrac{2}{\alpha}\right)$$

$$\ln\ell(\lambda) - 2 - 0.31 + \ln\left(\tfrac{2}{\alpha}\right) < \underbrace{r + s}_{\text{use (3-15)}} < \ln\ell(\lambda) + 2.02 + 0.21 + \ln\left(\tfrac{2}{\alpha}\right)$$

$$\ln\ell(\lambda) - 2.31 + \underbrace{\tfrac{1}{2}\ln\left(\tfrac{2(1-\cos\alpha)}{\alpha^2}\right)}_{\text{use (3-24)}} < \frac{\ell(g_\tau)}{2} < \ln\ell(\lambda) + \underbrace{2.23 + h}_{\text{use (3-14)}} + \underbrace{\tfrac{1}{2}\ln\left(\tfrac{2(1-\cos\alpha)}{\alpha^2}\right)}_{\text{use (3-24)}}$$

$$\ln\ell(\lambda) + \underbrace{(-2.31) + \ln\left(\tfrac{2}{\pi}\right)}_{=-2.7615\ldots} < \frac{\ell(g_\tau)}{2} < \ln\ell(\lambda) + 2.231$$

$$2\ln\ell(\lambda) - 5.6 < \ell(g_\tau) < 2\ln\ell(\lambda) + 4.5. \qquad \square$$

Remark 3.10 By assuming that μ is extremely long, one can force most of error in the preceding sequence of inequalities to become arbitrarily small. The following

exceptions are the only sources of error in the argument which do not disappear as $\ell(\mu) \to \infty$.

First, the length of the unknotting tunnel $\sigma \subset X$, bounded in equation (3-3), inevitably contributes to additive error in estimating the length of g_τ. This can be seen most clearly near the end of the proof of Lemma 3.5, where the factors of $1/10$ and $1/11$ depend on our choice of J, and the additive error of $\pm \ln 4$, which came from the bound on $\ell(\sigma_{\max})$, is all that remains if $J \to 1$.

The only other source of error that will not vanish as $\mu \to \infty$ is the term $\ln(2/\pi)$, which comes from equation (3-24). This error term has a natural geometric meaning, which can be seen by comparing Figures 2 and 3. The horocycle (and Euclidean geodesic) λ_ϵ in Figure 2 becomes $\bar{\lambda}_\epsilon$ in Figure 3, which is essentially a circle arc when μ is extremely long. This arc can cover anywhere up to half of the circle, corresponding to the angle $\alpha \in (0, \pi]$. In the Euclidean setting, the maximum ratio between an arc of a circle and a chord through the middle is exactly $\pi/2$. In our hyperbolic setting, the logarithmic savings achieved by a geodesic through the thin part of a horoball (or the thin part of a Margulis tube) turns this multiplicative error of $\pi/2$ into an additive error of $\ln(\pi/2)$.

All in all, the sharpest possible version of the above argument, in which $\ell(\mu) \to \infty$, would give the asymptotic estimate

$$(3\text{-}25) \qquad \ln \ell(\lambda) - \ln 4 - \ln(\tfrac{\pi}{2}) \le \ell(g_\tau)/2 \le \ln \ell(\lambda) + \ln 4.$$

4 Geodesic unknotting tunnels

In this section, we prove Theorem 1.3, showing that unknotting tunnels created by generic Dehn filling are isotopic to geodesics. Here are the ingredients of the proof.

First, we will make extensive use of Theorem 2.6. As in Sections 2 and 3, this will allow us to compare the geometry of the unfilled manifold X to that of its Dehn filling $X(\mu)$. In particular, we will need the geometric control of Theorem 2.6 to construct an embedded collar about a geodesic $\bar{\sigma} \subset X(\mu)$. This is done in Section 4.1.

Second, we will build on several results from Section 3 – specifically, Lemmas 3.5 and 3.6 and Theorem 3.9 – to show that the triangle $\Delta(\rho\rho'g_\tau)$ visible in Figure 5 is embedded in $X(\mu)$. The length estimates from Section 3 are only needed in a soft way. Mainly, we need to know that for generic Dehn fillings, the geodesic g_τ is arbitrarily long, which will imply that it stays within the thin part of $X(\mu)$, or else within an embedded tube about $\bar{\sigma}$. This argument is carried out in Section 4.2.

Finally, in Section 4.3, we combine these ingredients to prove Theorem 1.3. Using Theorem 2.11, we can locate an arc $\bar{\tau}(\sigma, \mu) \subset X(\mu)$ that is in the isotopy class of the unknotting tunnel τ. By carefully sliding this arc through the embedded triangle $\Delta(\rho\rho'g_\tau)$, we perform an isotopy of the tunnel to the geodesic g_τ.

4.1 Embedded collars about geodesics

Following the notation of Sections 2 and 3, X will denote a two-cusped, hyperbolic manifold that has tunnel number one. For $\epsilon = 0.29$, and for any $J > 1$, the drilling theorem, Theorem 2.6, implies that if we exclude finitely many slopes μ, there is a J–bilipschitz diffeomorphism ϕ between thick parts of X and $X(\mu)$. We will assume throughout that $1 < J \leq 1.1$; this means that all the estimates of Section 3 that hold true for a 1.1–bilipschitz diffeomorphism will also apply here.

Let $\sigma_\epsilon := \sigma \smallsetminus \mathbb{T}_\epsilon(X)$ denote the portion of the unknotting tunnel $\sigma \subset X$ that lies in the thick part of X. (Recall Remark 2.7.) We begin the argument by studying the diffeomorphic image $\phi(\sigma_\epsilon) \subset X(\mu)$.

Lemma 4.1 *The image of the unknotting tunnel $\sigma \subset X$ under the embedding of X into $X(\mu)$ is homotopic to a geodesic arc $\bar{\sigma}$. Let $\bar{\sigma}_\epsilon \subset X(\mu)$ be the portion of this geodesic that lies in the thick part of $X(\mu)$. Then there exists $J \in (1, 1.1]$, depending only on X and σ, and $\delta > 0$, depending on (X, σ, J), such that the following hold:*

(a) *$N_\delta(\bar{\sigma}_\epsilon)$ is an embedded solid cylinder $D^2 \times I$ in $X(\mu)$.*

(b) *If $\phi \colon X \smallsetminus \mathbb{T}_\epsilon(X) \to X(\mu) \smallsetminus \mathbb{T}_\epsilon(X(\mu))$ is a J–bilipschitz diffeomorphism, then $\phi(\sigma_\epsilon)$ is contained in $N_\delta(\bar{\sigma}_\epsilon)$.*

In particular, J and δ do not depend on the slope μ.

Proof Let $\sigma_\epsilon = \sigma \smallsetminus \mathbb{T}_\epsilon(X)$ denote the portion of the tunnel σ that lies in the thick part of X. Choose $r > 0$ so that $N_r(\sigma_\epsilon)$ is an embedded tube in X. Choose $J \in (1, 1.1]$ so that every arc homotopic to σ_ϵ of length less than $J^3 \ell(\sigma_\epsilon)$ lies in $N_r(\sigma_\epsilon)$. For this J, and for all but finitely many slopes μ, Theorem 2.6 gives a J–bilipschitz diffeomorphism ϕ between the thick parts of X and $X(\mu)$. (Here and below, we are using the term "thick" in the sense of Remark 2.7.)

For any μ such that $X(\mu)$ is hyperbolic, choose the smallest $\delta_\mu > 0$ so that every arc homotopic to $\bar{\sigma}_\epsilon$ in $X(\mu) \smallsetminus T(\epsilon)$ of length strictly less than $J^2 \ell(\sigma_\epsilon)$ must be contained in $N_{\delta_\mu}(\bar{\sigma}_\epsilon)$. Note that when the J–bilipschitz diffeomorphism ϕ exists, $\phi(\sigma_\epsilon)$ has length bounded by $J \ell(\sigma_\epsilon)$, hence $\phi(\sigma_\epsilon)$ lies in $N_{\delta_\mu}(\bar{\sigma}_\epsilon)$.

Let $\delta = \inf_\mu \delta_\mu$. We claim δ is strictly greater than 0. For, suppose not. Then we have a sequence of slopes μ_j such that $\delta_{\mu_j} =: \delta_j$ approaches 0. Note that the normalized lengths of these slopes must necessarily approach infinity. Hence by the Drilling Theorem, for any sequence $J_i > 1$, with $J_i < J$ and $J_i \to 1$, we may find a subsequence $\{\mu_i\}$ such that there is a J_i–bilipschitz diffeomorphism ϕ_i from the thick parts of X to those of $X(\mu_i)$. Let $d > 0$ be such that the neighborhood $N_d(\sigma_\epsilon) \subset X \smallsetminus T(\epsilon)$ is foliated by arcs parallel to σ_ϵ of length less than $J\ell(\sigma_\epsilon)$. Then $\phi_i(N_d(\sigma_\epsilon))$ is contained in $N_{\delta_i}(\overline{\sigma}_\epsilon)$ for all i, and as $i \to \infty$, the radius of $\phi_i(N_d(\sigma_\epsilon))$ approaches $d > 0$. Hence δ_i cannot approach 0, and thus $\delta > 0$.

Now, $N_\delta(\overline{\sigma}_\epsilon)$ is foliated by arcs α of length $\ell(\alpha) \leq J^2 \ell(\sigma_\epsilon)$. Then $\phi^{-1}(N_\delta(\overline{\sigma}_\epsilon))$ is foliated by arcs $\phi^{-1}(\alpha)$, each of length bounded by $J^3 \ell(\sigma_\epsilon)$. By choice of r, this implies $\phi^{-1}(N_\delta(\overline{\sigma}_\epsilon))$ is contained in $N_r(\sigma_\epsilon)$, and hence is embedded. Then $N_\delta(\overline{\sigma}_\epsilon)$ must also be embedded. $\qquad\square$

We may bootstrap from Lemma 4.1 to the following statement.

Lemma 4.2 *Let σ, J and δ be as in Lemma 4.1. Then, whenever there is a J–bilipschitz diffeomorphism $\phi\colon X \smallsetminus \mathbb{T}_\epsilon(X) \to X(\mu) \smallsetminus \mathbb{T}_\epsilon(X(\mu))$, the arc $\phi(\sigma_\epsilon)$ is isotopic to the geodesic $\overline{\sigma}_\epsilon$, through the embedded tube $N_\delta(\overline{\sigma}_\epsilon)$.*

The point of the lemma is that the homotopy from $\phi(\sigma_\epsilon)$ to $\overline{\sigma}_\epsilon$ is actually an isotopy. The proof below is inspired by Scharlemann's proof of Norwood's Theorem that tunnel number one knots are prime [40, Theorem 2.2].

Proof of Lemma 4.2 By Lemma 4.1, $\phi(\sigma_\epsilon)$ is contained in $N_\delta(\overline{\sigma}_\epsilon)$. Furthermore, the complement of $\phi(\sigma_\epsilon)$ in the thick part of $X(\mu)$ is homeomorphic to the complement of σ in the thick part of X, which is a genus-2 handlebody. In particular, $X(\mu) \smallsetminus (\mathbb{T}_\epsilon X(\mu) \cup \phi(\sigma_\epsilon))$ has free fundamental group.

Now, suppose that $\phi(\sigma_\epsilon)$ is not isotopic to the geodesic $\overline{\sigma}_\epsilon$. Since $\phi(\sigma_\epsilon)$ is contained in the embedded ball $N = N_\delta(\overline{\sigma}_\epsilon)$, the only way this can happen is if it is knotted inside that ball. Let $A = \partial N \smallsetminus \partial \mathbb{T}_\epsilon X(\mu)$ be the annulus boundary between N and the rest of the thick part of $X(\mu)$. Then, by van Kampen's Theorem,

$$\pi_1(N \smallsetminus \phi(\sigma_\epsilon)) *_{\pi_1(A)} \pi_1\big(X(\mu) \smallsetminus (\mathbb{T}_\epsilon X(\mu) \cup N)\big) \cong \pi_1\big(X(\mu) \smallsetminus (\mathbb{T}_\epsilon X(\mu) \cup \phi(\sigma_\epsilon))\big),$$

which is a free group. This means that $\pi_1(N \smallsetminus \phi(\sigma_\epsilon))$ is a subgroup of a free group, hence itself free. Since $N \smallsetminus \phi(\sigma_\epsilon)$ is homeomorphic to a knot complement in S^3, and the only knot whose complement has free fundamental group is the unknot, the arc $\phi(\sigma_\epsilon)$ must be unknotted in N. But then $\phi(\sigma_\epsilon)$ must be isotopic to the geodesic at the core of N. $\qquad\square$

We also note the following immediate consequence of Definition 2.4.

Lemma 4.3 *The Margulis tube* $V = \mathbb{T}_\epsilon(\gamma)$ *is an embedded solid torus in* $X(\mu)$. *Therefore, any arcs that are isotopic within* V *must be isotopic in* $X(\mu)$. $\qquad\square$

4.2 Embedded triangles in $X(\mu)$

Recall the geodesics $\rho, \rho' \subset X(\mu)$ that are depicted in Figure 5, and played a role in Section 3: ρ is the geodesic in the homotopy class that follows $\bar{\sigma}$ and runs halfway along the core γ, while ρ' is the geodesic in the homotopy class that follows $\bar{\sigma}$ and runs halfway along γ in the other direction.

For our purposes here, an isotopy from an unknotting tunnel τ to the geodesic g_τ will involve deforming arcs through the triangle $\Delta(\rho\rho'g_\tau)$ with sides ρ, ρ' and g_τ. (See Figure 5 for a lift of Δ to $\mathbb{H}^3$.) Thus we need to know more about this triangle.

Lemma 4.4 *Let* $\delta > 0$ *be the tube radius of Lemma 4.1. Then, for a sufficiently long Dehn filling slope* μ *on* T, *the geodesics* $\rho, \rho' \subset X(\mu)$ *are contained in either the thin part of* $X(\mu)$ *or a tube of radius* $\delta/2$ *about* $\bar{\sigma}$:

$$(4\text{-}1) \qquad (\rho \cup \rho') \subset \mathbb{T}_\epsilon X(\mu) \cup N_{\delta/2}(\bar{\sigma}_\epsilon).$$

Proof First, observe that since ρ and $\bar{\sigma}$ share an endpoint on the sphere at infinity, the distance between $\bar{\sigma}$ and a representative point $x \in \rho$ decreases monotonically as x moves toward the shared ideal vertex. Thus, in the thick part of $X(\mu)$, the distance between ρ and $\bar{\sigma}$ is maximized at $x = \rho \cap \partial\mathbb{T}_\epsilon(\gamma)$. Similarly, the distance between ρ' and $\bar{\sigma}$ is maximized at $x' = \rho' \cap \partial\mathbb{T}_\epsilon(\gamma)$.

Now, the result follows immediately from Lemma 3.6. In the statement of that lemma, choosing μ long enough forces the tube radius r to be as large as we like via equation (3-10). Choosing μ long enough also forces the error h to be as small as we like. Thus, in equation (3-12), choosing a long slope μ ensures that

$$(4\text{-}2) \qquad d(\bar{\sigma} \cap \partial\mathbb{T}_\epsilon(\gamma),\ \rho \cap \partial\mathbb{T}_\epsilon(\gamma)) < 0.02\, e^{-r} + h < \delta/2,$$

and similarly for the distance from $\bar{\sigma} \cap \partial\mathbb{T}_\epsilon$ to $\rho' \cap \partial\mathbb{T}_\epsilon$. $\qquad\square$

Lemma 4.5 *For a generic Dehn filling slope* μ, *the triangle* $\Delta = \Delta(\rho\rho'g_\tau)$ *is contained in* $\mathbb{T}_\epsilon X(\mu) \cup N_\delta(\bar{\sigma}_\epsilon)$. *Furthermore,* g_τ *intersects the Margulis tube* $\mathbb{T}_\epsilon(\gamma)$. *The portion of* g_τ *outside* $\mathbb{T}_\epsilon(\gamma)$, *and outside the maximal cusp of* $X(\mu)$, *has length less than* 10.

Proof   Recall that $\mathbb{H}^3$ is a Gromov hyperbolic metric space, with a δ–hyperbolicity constant of $\ln(\sqrt{2}+1)$. This means that any side of a triangle in $\mathbb{H}^3$ is contained within a $\ln(\sqrt{2}+1)$ neighborhood of the other two sides. Since triangle $\Delta = \Delta(\rho\rho'g_\tau)$ has a reflective symmetry, it follows that the midpoint m of g_τ (fixed by the reflective symmetry) is within a $\ln(\sqrt{2}+1)$ neighborhood of both ρ and ρ'.

Next, observe that because ρ and g_τ share an ideal vertex, the distance between side ρ and $x \in g_\tau$ decreases exponentially toward 0 as x moves with unit speed toward the shared ideal vertex. The same statement is true for ρ' and g_τ. Thus, there is a constant R (depending only on δ) such that if m is the midpoint of g_τ,

$$(4\text{-}3) \qquad x \in g_\tau \text{ and } d(x,m) > R \quad \Rightarrow \quad d(x,\rho) < \delta/2 \text{ or } d(x,\rho') < \delta/2.$$

Now, we recall several estimates from Section 3. First, by Lemma 3.5, the distance in $X(\mu)$ between the Margulis tube $V = \mathbb{T}_\epsilon(\gamma)$ and the maximal horocusp $N_\mu(K)$ is

$$(4\text{-}4) \qquad s = d(\partial N_\mu(K), \partial \mathbb{T}_\epsilon(\gamma)) < 2.621 + 2.02 < 5.$$

On the other hand, by Theorem 3.9, choosing μ and λ sufficiently long will ensure that the length of g_τ relative to the maximal cusp $N_\mu(K)$ is arbitrarily long, in particular

$$(4\text{-}5) \qquad \ell(g_\tau) > 2\ln\ell(\lambda) - 5.2 \gg 2R + 10.$$

Combining (4-4) with (4-5), we conclude that the middle $2R$ of the length of g_τ (ie, the portion of g_τ that is within R of the midpoint m) is contained inside the Margulis tube $\mathbb{T}_\epsilon(\gamma)$. Thus, by (4-3), it follows that when μ and λ are sufficiently long,

$$(4\text{-}6) \qquad g_\tau \smallsetminus \mathbb{T}_\epsilon(\gamma) \subset N_{\delta/2}(\rho) \cup N_{\delta/2}(\rho').$$

Combining (4-1) and (4-6) gives the desired result. $\square$

Lemma 4.6   *For a generic slope μ, the triangle $\Delta = \Delta(\rho\rho'g_\tau)$ is embedded in $X(\mu)$.*

Proof   Consider the intersections between Δ and the following three regions of $X(\mu)$: the Margulis tube $\mathbb{T}_\epsilon(\gamma)$, the cusp neighborhood $\mathbb{T}_\epsilon(K)$, and the remaining thick part $X(\mu) \smallsetminus \mathbb{T}_\epsilon$. See Remark 2.7 and Figure 5.

We may lift Δ to a triangle $\widetilde{\Delta} \subset \mathbb{H}^3$. There is a decomposition of $\mathbb{H}^3$ into tubes covering the Margulis tube $\mathbb{T}_\epsilon(\gamma)$, horoballs covering $\mathbb{T}_\epsilon(K)$, and the remaining piece covering the thick part. We will show that Δ is embedded in $X(\mu)$ by showing that none of these pieces of $\mathbb{H}^3$ contains a point of intersection between $\widetilde{\Delta}$ and one of its translates by the deck transformation group.

First, consider a solid tube about a geodesic $\widetilde{\gamma} \subset \mathbb{H}^3$, which covers the Margulis tube $\mathbb{T}_\epsilon(\gamma)$. By Lemma 4.3, the only deck transformations of $\mathbb{H}^3$ that fail to move this

solid tube off itself belong to a $\mathbb{Z}$ subgroup fixing the geodesic axis $\widetilde{\gamma}$. Now, observe that each lift of $\overline{\sigma}$ intersecting $\widetilde{\gamma}$ is contained in a totally geodesic plane perpendicular to $\widetilde{\gamma}$, and that the triangle $\widetilde{\Delta}$ is bounded between two consecutive planes. See Figure 5. Thus every non-trivial deck transformation in the $\mathbb{Z}$ subgroup fixing $\widetilde{\gamma}$ will move $\widetilde{\Delta}$ completely off itself.

Next, consider the piece of Δ in the thick part $X(\mu) \smallsetminus \mathbb{T}_\epsilon X(\mu)$. By Lemma 4.5, this piece of Δ is entirely contained in the embedded tube $N_\delta(\overline{\sigma}_\epsilon)$. Therefore, $\widetilde{\Delta} \smallsetminus \widetilde{\mathbb{T}_\epsilon}$ is entirely contained in one of the disjointly embedded solid cylinders covering $N_\delta(\overline{\sigma}_\epsilon)$, hence is disjoint from all the images of $\widetilde{\Delta}$ by the deck transformation group.

Finally, consider the part of Δ in the horocusp $\mathbb{T}_\epsilon(K)$. By Lemma 4.5, we already know that $\Delta \cap \partial \mathbb{T}_\epsilon(K)$ is contained in an embedded disk of radius δ about $\overline{\sigma} \cap \partial \mathbb{T}_\epsilon(K)$. Thus $\Delta \cap \mathbb{T}_\epsilon(K)$ follows the $T^2 \times [0, \infty)$ product structure of the horocusp $\mathbb{T}_\epsilon(K)$, and the portion of Δ in this cusp is also embedded. $\square$

4.3 Completing the proof

We can now complete the proof of Theorem 1.3.

Theorem 1.3 *Let X be an orientable hyperbolic 3–manifold that has two cusps and tunnel number one. Choose a generic filling slope μ on one cusp of X, and let $\tau \subset X(\mu)$ be an unknotting tunnel associated to a tunnel $\sigma \subset X$. Then τ is isotopic to the geodesic g_τ in its homotopy class.*

Proof The two-cusped manifold X has finitely many Heegaard splittings of genus 2, hence finitely many unknotting tunnels (Hass [25], Li [29]). Choose a value $J \in (1, 1.1]$ that satisfies Lemma 4.1 for every unknotting tunnel of X. Next, choose a generic Dehn filling slope μ, such that both μ and its shortest longitude λ are long enough to satisfy Lemmas 4.2, 4.4 and 4.6.

Choose an unknotting tunnel $\sigma \subset X$, and let $\tau = \overline{\tau}(\sigma, \mu) \subset X(\mu)$ be an unknotting tunnel associated to σ, as in Definition 2.10 and Theorem 2.11(a). Let us unpack what this means.

Using the geodesic $\overline{\sigma} \subset X(\mu)$ and the core curve γ of the Dehn filling solid torus, construct geodesics ρ and ρ' as in Figure 5. Let $\overline{Q} = \Delta(\rho\rho'\gamma) \smallsetminus \mathbb{T}_\epsilon X(\mu)$ be a totally geodesic quadrilateral obtained by removing the thin part $\mathbb{T}_\epsilon X(\mu)$ from the triangle $\Delta(\rho\rho'\gamma)$. The quadrilateral $\overline{Q}$ can be seen schematically in Figure 7, where a lift of $\overline{Q}$ lies on the right of the figure, with one side labeled C_1, two sides on (lifts of) ρ' and ρ, and the last side on the horosphere where these lifts of ρ' and ρ meet.

By Lemma 4.4, $\overline{Q}$ is embedded in the tubular neighborhood $N_{\delta/2}(\overline{\sigma}_\epsilon)$. Let $\overline{C}$ be an arc in $\partial\mathbb{T}_\epsilon(\gamma)$ whose endpoints are $\rho \cap \partial\mathbb{T}_\epsilon(\gamma)$ and $\rho' \cap \partial\mathbb{T}_\epsilon(\gamma)$, and which lies in the same plane as triangle $\Delta(\rho\rho'g_\tau)$. (This is the same arc that was denoted C_2 in the proof of Lemma 3.8, shown in Figure 7.) Notice that the union of $C_1 = \overline{Q} \cap \partial\mathbb{T}_\epsilon(\gamma)$ and $C_2 = \overline{C}$ is a closed curve isotopic to $\overline{\lambda}_\epsilon$. See Figure 7.

Now, $Q = \phi^{-1}(\overline{Q})$ is an embedded quadrilateral in the thick part of X. Furthermore, observe that the two sides of $\overline{Q}$ that run along ρ and ρ' are each isotopic to $\overline{\sigma}_\epsilon$, and $\overline{\sigma}_\epsilon$ is isotopic to $\phi(\sigma_\epsilon)$ by Lemma 4.2. Pulling back these isotopies via ϕ^{-1}, we conclude that two opposite sides of Q, namely $s_1 = \phi^{-1}(\rho)$ and $s_2 = \phi^{-1}(\rho')$, run parallel to σ through the thick part of X. Furthermore, the union of $\phi^{-1}(\overline{C})$ and $Q \cap \partial\mathbb{T}_\epsilon(T)$ is a closed curve isotopic to λ. In other words, the quadrilateral Q satisfies all the criteria of Definition 2.10.

In the language of Definition 2.10, this means that $\tau(\sigma, \mu) = s_1 \cup \phi^{-1}(\overline{C}) \cup s_2$ is a tunnel arc associated to σ and μ. Therefore, by Theorem 2.11,

$$\phi(\tau(\sigma, \mu)) = (\rho \smallsetminus \mathbb{T}_\epsilon) \cup \overline{C} \cup (\rho' \smallsetminus \mathbb{T}_\epsilon)$$

is an arc isotopic to our unknotting tunnel τ.

Now, we may construct an isotopy from $\phi(\tau(\sigma, \mu))$ to g_τ in two stages. First, homotope arc $\overline{C}$ to the piecewise geodesic arc $(\rho \cap \mathbb{T}_\epsilon(\gamma)) \cup (\rho' \cap \mathbb{T}_\epsilon(\gamma))$, through the Margulis tube $\mathbb{T}_\epsilon(\gamma)$. Because the Margulis tube is embedded by Lemma 4.3, and in fact the entire homotopy occurs in the geodesic plane containing triangle $\Delta(\rho\rho'g_\tau)$, this homotopy is an isotopy. At the end of this isotopy, we have shown that tunnel τ is isotopic to $\rho \cup \rho'$, two of the sides of $\Delta(\rho\rho'g_\tau)$.

Next, homotope $\rho \cup \rho'$ to the third side g_τ of $\Delta(\rho\rho'g_\tau)$, through the triangle $\Delta(\rho\rho'g_\tau)$. Because $\Delta(\rho\rho'g_\tau)$ is embedded in $X(\mu)$ by Lemma 4.6, the homotopy from two sides of the triangle to the third side is again an isotopy. Therefore, τ is isotopic to the geodesic g_τ. $\qquad\square$

It is worth observing that in the proof of Theorem 1.3, we used the hypothesis on the length of λ only at the very end, to apply Lemma 4.6 and construct an isotopy from $\rho \cup \rho'$ to the geodesic g_τ. When we strip away the last paragraph of the proof, what remains is the following corollary, which does not require any hypotheses on the longitude λ.

Corollary 4.7 *Let X be an orientable hyperbolic 3–manifold that has two cusps and an unknotting tunnel σ. Choose a Dehn filling slope μ on one boundary torus T of X, and let $\tau \subset X(\mu)$ be the unknotting tunnel associated to σ, as in Figure 1. Then, for*

all but finitely many choices of filling slope μ, the tunnel τ is isotopic to the piecewise geodesic curve $\rho \cup \rho'$ depicted in Figure 5.

5 Canonical and non-canonical geodesics

Heath and Song showed by a single counterexample that unknotting tunnels are not necessarily isotopic to canonical geodesics [26]. In this section, we give conditions that will guarantee that an unknotting tunnel is a canonical geodesic, or is *not* a canonical geodesic. This results in the proof of Theorem 1.6. In addition, in Theorem 5.3, we construct an infinite family of one-cusped manifolds M_i, such that each M_i in the family has one canonical tunnel and one non-canonical tunnel.

Recall, from Definition 1.5, that a geodesic in $M = X(\mu)$ is called *canonical* when it is an edge of the canonical polyhedral decomposition $\mathcal{P}$, ie, the geometric dual to a face of the Ford domain. Note that when there are multiple cusps, the Ford domain depends on choice of horoball expansion for all cusps. However, when there is only one cusp, canonical geodesics are well defined.

Consider again the manifold X, with cusps corresponding to torus boundary components K and T, and unknotting tunnel σ running between them. Take a horoball expansion as follows. For the cusp corresponding to K, to be left unfilled, expand a horocusp maximally, until it becomes tangent to itself. Now expand a horocusp about T slightly, so that it has very small volume. This expansion determines a Ford domain, hence a canonical polyhedral decomposition. We will use this canonical polyhedral decomposition of X in the results below. (It is a theorem of Akiyoshi [5] that the combinatorics of the Ford domain stabilizes as the volume of a horocusp shrinks toward 0; thus there is a well-defined canonical decomposition corresponding to a "sufficiently small" horocusp about T.)

The proof of Theorem 1.6 will use the recent work of Guéritaud and Schleimer [23]. In that paper, the authors show that if the canonical polyhedral decomposition of a manifold X is a triangulation, and if there is a unique shortest canonical geodesic meeting a cusp to be filled, then the canonical decomposition of the filled manifold $X(\mu)$ can be determined by replacing two tetrahedra of the canonical triangulation of X with a collection of tetrahedra forming a "layered solid torus."

The results we need are contained in the following lemma.

Lemma 5.1 *Let X be a hyperbolic manifold with two cusps, denoted K and T, such that there is a unique shortest geodesic α running from K to T. Let $\mathcal{D}$ be the canonical polyhedral decomposition of X, relative to a cusp neighborhood where the horocusp*

about T is sufficiently small. Then, for a sufficiently long slope μ on T, every edge E of the canonical decomposition $\mathcal{D}_\mu$ of $X(\mu)$ satisfies one of the following:

   (a) *E is isotopic to the image of a canonical edge in $\mathcal{D}$ under the (topological) inclusion $X \hookrightarrow X(\mu)$, or*

   (b) *E is isotopic to the image of an edge created by subdividing some polyhedron of $\mathcal{D}$ into ideal tetrahedra, under the inclusion $X \hookrightarrow X(\mu)$, or*

   (c) *E is isotopic to an arc that follows α into the filled solid torus, runs n times around the core of the solid torus (for some positive integer $n \in \mathbb{N}$), and then follows α back out. There is exactly one such edge for $n = 1$.*

The edges in (c) are exactly the edges of the canonical tetrahedra that make up the layered solid torus $\mathcal{V}$, which will be described in the proof.

Proof Suppose, for the moment, that the polyhedral decomposition $\mathcal{D}$ consists entirely of tetrahedra. (This hypothesis is assumed globally in [23]. See Remark 5.2 below.)

If the horocusp about T is sufficiently small, Guéritaud and Schleimer show that the only edge of $\mathcal{D}$ entering cusp T is the unique shortest geodesic α. In particular, the cusp cellulation of T will contain exactly one vertex, corresponding to the endpoint of α at T. There are two special tetrahedra, denoted Δ and Δ', such that three edges of Δ (sharing an ideal vertex) and three edges of Δ' are identified at the shortest geodesic α. All other tetrahedra of $\mathcal{D}$ are disjoint from the horocusp about T, and have all of their vertices at K. See [23, Section 4.1].

For a sufficiently long Dehn filling slope μ, Guéritaud and Schleimer observe that the tetrahedra of $\mathcal{D} \smallsetminus (\Delta \cup \Delta')$ remain canonical in $X(\mu)$. This is because the canonicity of a tetrahedron can be encoded in finitely many strict inequalities (which express relative distance to various horoballs in $\mathbb{H}^3$, or equivalently convexity in Minkowski space $\mathbb{R}^{3+1}$). Hence, the canonicity of a tetrahedron is an open condition, and remains true as the complete hyperbolic metric on X is perturbed slightly to give the hyperbolic structure on $X(\mu)$.

If $\mathcal{D}$ consists entirely of tetrahedra, Guéritaud and Schleimer show that the canonical decomposition $\mathcal{D}_\mu$ will combinatorially be of the form

$$\mathcal{D}_\mu = (\mathcal{D} \smallsetminus \{\Delta, \Delta'\}) \cup \mathcal{V},$$

where $\mathcal{V} = \Delta_1 \cup \cdots \cup \Delta_N$ is a solid torus with one point on the boundary removed. This removed point corresponds to the endpoint of α on K. The solid torus $\mathcal{V}$ has a layered triangulation by tetrahedra $\Delta_1, \ldots, \Delta_N$, as follows. The tetrahedron Δ_1 is glued along two faces to the punctured torus $\partial \mathcal{V} = \partial(\Delta \cup \Delta')$, then Δ_2 is glued along

two faces to the punctured torus on the other side of Δ_1, and so on. At the core of the solid torus $\mathcal{V}$, two faces of Δ_N are glued by folding onto a Möbius band. See [23, Section 2] for more details, and in particular [23, Figure 3].

For our purposes, the salient points are as follows. First, every edge of $\mathcal{V} = \Delta_1 \cup \cdots \cup \Delta_N$ runs some nonzero number of times about the core of $\mathcal{V}$. Second, all of these edges share the same ideal vertex, namely the endpoint of α at cusp K. Thus every edge of $\mathcal{V}$ is homotopic to a portion of α, followed by $n \in \mathbb{N}$ trips about the core, followed by returning to the ideal vertex along α. The homotopy class of each (unoriented) edge in $\mathcal{V}$ is completely determined by the positive integer n. Thus, since there can only be one hyperbolic geodesic in any homotopy class, there is at most one canonical edge for any value $n \in \mathbb{N}$. Note that there will actually be an edge in $\mathcal{V}$ for $n = 1$: this is the core of the Möbius band onto which the tetrahedra are layered. This edge is marked with an arrow in [23, Figure 3].

This completes the proof in the case where $\mathcal{D}$ consists entirely of tetrahedra. Next, consider what can be said without this restrictive hypothesis.

By the same argument as in Section 4.1 of [23], the cusp cellation of a horospherical torus about T will contain exactly one vertex, corresponding to the endpoint of α at T. Thus the Delaunay decomposition of the torus T is either two triangles or one rectangle. The corresponding canonical 3–cells of $\mathcal{D}$ with ideal vertices at T are either two tetrahedra (which may be labeled Δ, Δ' as before), or a single ideal rectangular pyramid, which can be subdivided into two tetrahedra Δ, Δ' by choosing a diagonal for the rectangle.

When we perform Dehn filling along a sufficiently long slope μ, the same argument as above implies that every canonical tetrahedron of $\mathcal{D} \smallsetminus (\Delta \cup \Delta')$ will remain canonical in $X(\mu)$. The presence of a larger canonical polyhedron in $\mathcal{D}$ is equivalent (by duality) to a vertex v of the Ford–Voronoi domain that is equidistant to five or more horoballs. As the complete hyperbolic metric on X is perturbed slightly to give the metric on $X(\mu)$, this equality of distances may break into inequalities, and a large polyhedron in $\mathcal{D}$ may become subdivided into tetrahedra or other cells. Nonetheless, we observe that any new edges created in this fashion are interior to polyhedra of $\mathcal{D}$ or to faces of $\mathcal{D}$. Thus all edges in $X(\mu)$ that do not come from $(\Delta \cup \Delta')$ satisfy (a) or (b) in the statement of the lemma.

After all the cells of $\mathcal{D} \smallsetminus (\Delta \cup \Delta')$ are mapped to $X(\mu)$, and subdivided as necessary, what remains is again a solid torus $\mathcal{V}$ with one point on the boundary removed. Then, by the same argument of Guéritaud and Schleimer [23, Section 4.4], the canonical subdivision of $\mathcal{V}$ will again be a layered triangulation. As above, every edge of $\mathcal{V}$ satisfies (c) in the statement of the lemma, completing the proof. □

Remark 5.2 The main reason why Guéritaud and Schleimer assumed that the canonical decomposition $\mathcal{D}$ of X consists entirely of tetrahedra is that their goal was to completely describe the canonical decomposition of $X(\mu)$. Typically, large cells in the canonical decomposition $\mathcal{D}$ of X tend to break up into tetrahedra in $\mathcal{D}_\mu$, in a pattern that seems hard to predict from the combinatorial data alone. For our purposes in this paper, it suffices to know that each edge created in this subdivision is contained in the closure of a cell of $\mathcal{D}$.

We may now complete the proof of Theorem 1.6.

Theorem 1.6 *Let X be a two-cusped, orientable hyperbolic 3–manifold in which there is a unique shortest geodesic arc between the two cusps. Choose a generic Dehn filling slope μ on a cusp T of X. Then, for each unknotting tunnel $\sigma \subset X$, the tunnel $\tau \subset X(\mu)$ associated to σ will be canonical if and only if σ is the shortest geodesic between the two cusps of X.*

Proof For the "if" direction of the theorem, suppose that the unknotting tunnel σ of X is the unique shortest canonical geodesic from T to the other cusp K of X. By Definition 2.10, the associated tunnel $\overline{\tau}(\sigma, \mu) \subset X(\mu)$ follows $\overline{\sigma}$ to the added solid torus V, runs once around the longitude λ (ie, once around the core of V), and returns along $\overline{\sigma}$. By Lemma 5.1(c), there is exactly one edge E in the canonical decomposition of $X(\mu)$ that does exactly that. Thus the tunnel $\overline{\tau} = \tau(\sigma, \mu)$ is homotopic to this edge E. In fact, since E and $\overline{\tau}$ are both boundary-parallel arcs in the layered solid torus $\mathcal{V}$, they are isotopic in $\mathcal{V}$, hence in $X(\mu)$.

We remark that the proof of this direction does not need any hypotheses on λ; all that's needed is that the slope μ is long enough to apply the work of Guéritaud and Schleimer.

For the "only if" direction of the theorem, suppose that the unknotting tunnel σ of X is *not* the shortest canonical geodesic connecting cusps T and K. Denote this unique shortest geodesic by α.

Choose a value of $\epsilon \leq 0.29$ small enough so that the ϵ–thin part $\mathbb{T}_\epsilon(T) \subset X$ is contained in a horocusp about T that is *sufficiently small* to satisfy Lemma 5.1. Then Theorem 2.6 implies that for a sufficiently long Dehn filling slope μ, the boundary of the ϵ–thin Margulis tube $\mathbb{T}_\epsilon(\gamma) \subset X(\mu)$ will be contained in $\mathcal{V}$:

$$\partial \mathbb{T}_\epsilon(\gamma) \subset \mathcal{V} \subset X(\mu),$$

where $\mathcal{V}$ is the layered solid torus of Lemma 5.1.

Next, let $\tau = \overline{\tau}(\sigma, \mu)$ be the unknotting tunnel of $X(\mu)$ associated to σ. We claim that when μ and λ are sufficiently long, the geodesic g_τ in the homotopy class of τ must intersect the Margulis tube $\mathbb{T}_\epsilon(\gamma)$. The proof of this claim is essentially identical to the proof of Lemma 4.5, where it is proved for the value $\epsilon = 0.29$. (For a different value of ϵ, there would be different numerical estimates in (4-4), but the argument is otherwise the same.) Thus we may conclude that for a generic slope μ, the geodesic g_τ in the homotopy class of τ must intersect the layered solid torus $\mathcal{V}$.

We are now ready to complete the proof. By Lemma 5.1, if an edge of the canonical decomposition meets $\mathcal{V}$, then it is isotopic to an arc that follows α, runs some number of times around the core curve γ of $\mathcal{V}$, then follows α back out. But we have assumed that $\alpha \neq \sigma$; in particular, these geodesics enter the horocusp K at different points. Thus the geodesic g_τ, which is homotopic to $\overline{\sigma} \cdot \gamma \cdot \overline{\sigma}^{-1}$, cannot be a canonical edge in $X(\mu)$. □
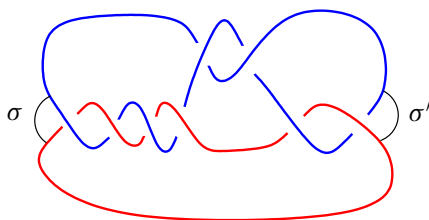

Figure 8: A two-bridge link L of slope $5/22$, with upper tunnel σ and lower tunnel σ'

Theorem 5.3 *There exists an infinite family $\{M_i\}$ of one-cusped hyperbolic tunnel number one manifolds, such that each M_i has two unknotting tunnels, of which one is canonical and the other is not.*

Proof To prove the result, all we need is a two-cusped, one-tunnel manifold X with two unknotting tunnels, one of which is the unique shortest canonical geodesic, and one of which is not. We may then Dehn fill one cusp of X and apply Theorem 1.6 to obtain infinitely many tunnel number one manifolds $M_i = X(\mu_i)$ as in the statement of the theorem.

The example we use is the complement of a 2–bridge link L, shown in Figure 8. By a result of Adams and Reid [4], the only unknotting tunnels of a two-bridge link are the upper and lower tunnels. These are shown in Figure 8 as σ and σ'.

Using SnapPy [14], we compute the hyperbolic structure on $S^3 \smallsetminus L$, shrink the cusp about the lower (red) component of Figure 8, and expand the cusp about the upper (blue)

component maximally to determine the appropriate canonical polyhedral decomposition. Figure 9 shows the cusp neighborhood for this link. Notice that there is a single maximal horoball in a fundamental domain in the figure. This implies that there is a unique shortest canonical geodesic between the two cusps. We will see that this corresponds to the lower tunnel σ'.
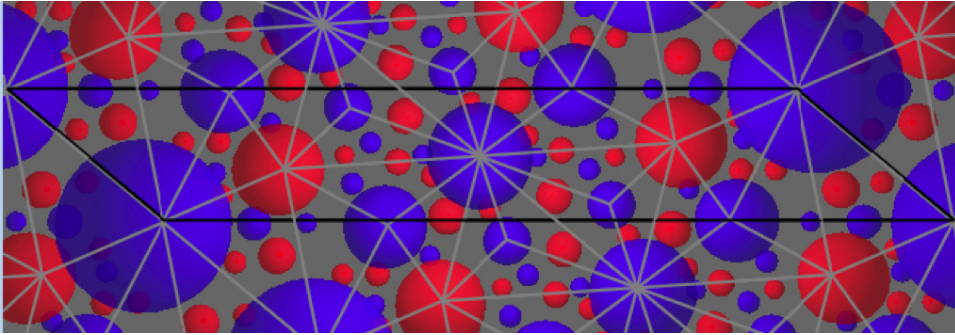

Figure 9: A horoball packing diagram for the two-bridge link of slope $5/22$. The red cusp (shaded lighter in grayscale) lifts to the horoball about infinity.

To do so, we consider the combinatorics of the canonical decomposition of the 2–bridge link complement. By a theorem of Akiyoshi, Sakuma, Wada and Yamashita [6], and independently Guéritaud [21], the canonical polyhedral decomposition of the link complement follows a combinatorial pattern that can be read off the diagram of the link. See Sakuma and Weeks [39] or Futer [22, Appendix] for detailed descriptions of this triangulation.

In particular, the canonical decomposition determines a triangulation of the cusp. This cusp triangulation has the feature that exactly two edges run over an entire meridian of the cusp. Each of these two special edges forms one side of an ideal triangle, while the other two sides of the triangle run along the (same) upper or lower tunnel. In Figure 9, we see the upper and lower tunnels "head on," as vertices in the cusp torus. One of the tunnels must correspond to the maximal horoball, as claimed, and the other corresponds to the smaller horoball shown in the center of Figure 9. By counting valences of the vertices lying over these horoballs, we see that the lower tunnel σ' corresponds to the maximal horoball.

Now, Theorem 2.11 implies that for a generic Dehn filling, the resulting manifold has two unknotting tunnels $\overline{\tau}(\sigma, \mu)$ and $\overline{\tau}(\sigma', \mu)$. By Theorem 1.6, the second of these is canonical, while the first is not. □

6 Knots with long tunnels in S^3

In this section, we prove there exist knots in S^3 with arbitrarily long unknotting tunnel. Our starting point is the alternating chain link in S^3 with four link components. We will denote the link by C, its complement by $M = S^3 \smallsetminus C$. This link is hyperbolic, for example by work of Neumann and Reid [36]. (In fact, the arguments below will apply to any choice of hyperbolic chain link on four strands. We choose the alternating one for concreteness.)

Label the four link components L_1, L_2, L_3 and L_4, with L_1 and L_3 opposite each other. Notice that there is a 3–punctured sphere with boundary components a longitude of L_1 and a meridian of L_2 and L_4, embedded on the plane of projection of the link diagram. This will play a part in the arguments below. See Figure 10.

Our knot complement in S^3 will be obtained by Dehn filling three of the four link components, along carefully chosen slopes.

Lemma 6.1 *The link complement $M = S^3 \smallsetminus C$ has exactly two genus-2 Heegaard splittings.*

Proof One genus-2 Heegaard surface separates L_1 and L_2 from L_3 and L_4. It restricts to a standard genus-2 Heegaard splitting of S^3, and the links L_1 and L_2 run through two distinct 1–handles of a corresponding genus-2 handlebody in S^3. The two core tunnels in the compression bodies separated by this surface correspond to arcs between L_1 and L_2, and between L_3 and L_4. These core tunnels are the unique arcs on the intersection of 3–punctured spheres bounded by longitudes of these link components. See Figure 10.

The other genus-2 Heegaard surface of M is obtained by rotating the right panel of Figure 10 by a quarter-turn. It separates L_1 and L_4 from L_2 and L_3, with core tunnels running between the corresponding boundary components.

By Lemma 2.9, every core tunnel of M can be isotoped so it is fixed by an isometry that is a hyper-elliptic involution. Since M has 4 torus boundary components, such an involution must map each boundary component to itself. It must also reverse the orientation on each longitude and each meridian. We claim there is only one such isometry, and that it preserves the 3–punctured sphere S depicted in Figure 10.

If there are two such involutions, their composition is an isometry, ψ, which sends (a neighborhood of) each component of the link to itself, preserving the orientation of each longitude. The orbit of S under the finite cyclic group F generated by ψ consists of disjoint 3–punctured spheres, one for each element of F. These spheres
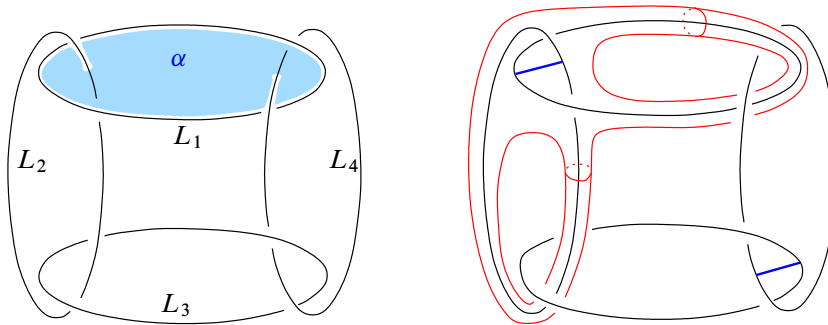
Figure 10: The alternating chain link with four link components. On the left, an embedded 3–punctured sphere S is shown, shaded. On the right, a Heegaard surface is shown, along with corresponding core tunnels.

separate the link complement into $|F|$ isometric pieces. However, the cusp of L_3 lies in exactly one of these complementary components. This contradicts the existence of a second hyper-elliptic isometry.

Now, suppose for a contradiction that the manifold $M = S^3 \smallsetminus C$ contains another genus-2 Heegaard surface Σ, apart from the two surfaces that we have already exhibited. We know that the hyper-elliptic involution corresponding to Σ rotates the 3–punctured sphere S about its central axis, and that two link components of C lie on the same side of Σ. By the pigeonhole principle, an arc in the axis of S that connects two link components on the same side of Σ must be a core tunnel of Σ.

There are three potential core tunnels contained in the axis of S. Two of these are known core tunnels of existing splittings, described above. The third is an arc α running from L_4 to L_2, shown in Figure 10. We now argue that this arc cannot be a core tunnel in $M \smallsetminus \Sigma$. For, suppose that α is a core tunnel. Then the genus-2 Heegaard surface Σ must separate L_2 and L_4 from L_1 and L_3. This surface must carry the $\mathbb{Z}_2$–homology of $M = S^3 \smallsetminus C$. Note $H_1(M; \mathbb{Z}_2)$ is generated by the four meridians m_1, m_2, m_3 and m_4, hence has rank 4.

Let N be the compression body that is the neighborhood of the cusp tori of L_2 and L_4, along with the arc α joining these tori. The subgroup $G \subset H_1(M; \mathbb{Z}_2)$ corresponding to the inclusion $N \hookrightarrow M$ is generated by meridians and longitudes of the boundary components L_2 and L_4. The longitude of L_2 is homologous to $m_1 + m_3$ (mod 2), as is the longitude of L_4. Hence $G \subset H_1(M; \mathbb{Z}_2)$ has rank only 3, and cannot be the entire group. On the other hand, if the positive boundary $\partial_+ N$ was a Heegaard surface Σ, N would have to carry all of the homology of M. This is a contradiction. Hence there are exactly two Heegaard splittings, described above. $\qquad\square$

Lemma 6.2 *Dehn filling $M = S^3 \smallsetminus C$ along an integer slope $p/1$ on cusp L_3 yields a manifold homeomorphic to $(T^2 \times I) \smallsetminus K$, where K is a knot. The homeomorphism maps the boundary torus of L_2 to $T^2 \times \{0\}$, the boundary torus of L_4 to $T^2 \times \{1\}$, and the link component L_1 to the knot $K \subset T^2 \times I$.*

Proof If we perform meridian Dehn filling on L_1, then we obtain an unlinked chain of three unknots. The complement of this chain of unknots is homeomorphic to the product of a 3–punctured sphere P and the circle, $P \times S^1$, which is Seifert fibered with base orbifold $S^2(\infty, \infty, \infty)$. Thus L_1 is a knot in $P \times S^1$.

Now consider p/q Dehn filling on the component L_3. The longitude of L_3 bounds the 3–punctured sphere fiber, so is horizontal. Dehn filling yields a new Seifert fibered space with base orbifold $S^2(\infty, \infty, q)$. Hence if we perform integral Dehn filling, $q = 1$, then the Seifert fibered space has base orbifold $S^2(\infty, \infty)$. This is an annulus. The Seifert fibered space must be the product of an annulus and a circle, which is homeomorphic to $T^2 \times I$. Note that the two boundary tori $T^2 \times \{0, 1\}$ come from the boundary tori of L_2 and L_4. Thus, after $p/1$ integer filling on L_3, the link component L_1 is a knot $K \subset T^2 \times I$. $\square$

Lemma 6.3 *For large n, the manifold obtained by Dehn filling $M = S^3 \smallsetminus C$ along the slope $2n$ on L_3 is hyperbolic, and has exactly two genus-2 Heegaard splittings, which come from the Heegaard splittings of M.*

Proof Let $X = T^2 \smallsetminus K$ denote the manifold obtained by this Dehn filling. Again by Lemma 2.9, any genus-2 Heegaard splitting of $T^2 \smallsetminus K$ gives rise to an involution of the manifold X that fixes the boundary components. For large n, the core of the Dehn filling solid torus will be shorter than any other closed geodesic. Thus it must be taken to itself by the involution. But then the involution restricts to an involution of $S^3 \smallsetminus C$ that fixes boundary components, preserving slopes on the boundary components L_1, L_2 and L_4.

As in the proof of Lemma 6.1, this fixes the 3–punctured sphere S with boundary components isotopic to the longitude of L_1 and meridian of L_2 and L_4. So again the core tunnel must be one of the three arc components fixed by a reflection of that 3–punctured sphere. Again two of these arcs are known core tunnels, and we must again rule out the third arc α, shown in Figure 10. We do so by a homology argument.

As before, if a core tunnel α connects L_2 and L_4, then the compression body N obtained as a neighborhood of these two components and the tunnel must carry the homology of the manifold M. The homology $H_1(S^3 \smallsetminus C; \mathbb{Z}_2)$ is generated by meridians m_1, m_2, m_3 and m_4. After Dehn filling along $(2n)m_3 + \ell_3$, the longitude ℓ_3

becomes equal to zero mod 2. Hence $\ell_3 = 0 = m_2 + m_4$, and the homology now has rank 3.

On the other hand, the subgroup $G \subset H_1(M; \mathbb{Z}_2)$ induced by the inclusion $N \hookrightarrow M$ is generated by meridians and longitudes m_2, ℓ_2, m_4 and ℓ_4. Since

$$\ell_2 = m_1 + m_3 \pmod 2 = \ell_4,$$

and $m_2 = m_4$ after Dehn filling, G has rank only 2. Thus the positive boundary $\partial_+ N$ cannot be a Heegaard splitting surface. □

By Lemmas 6.2 and 6.3, there exists a knot $K \subset (T^2 \times I)$ with exactly two genus-2 Heegaard splittings, and with hyperbolic complement. Denote the hyperbolic manifold $X = (T^2 \times I) \smallsetminus K$. To obtain a knot in S^3, we will be Dehn filling two more of the boundary components of X: those corresponding to $T^2 \times \{0\}$ and to $T^2 \times \{1\}$.

Let $\iota(\,\cdot\,, \cdot\,)$ denote the geometric intersection number between two slopes on a torus. The following lemma is well-known.

Lemma 6.4 *Consider the manifold $Y = T^2 \times [0, 1]$. Put the same framing on $T^2 \times \{0\}$ and $T^2 \times \{1\}$. Choose slopes μ_i on $T^2 \times \{i\}$, such that $\iota(\mu_0, \mu_1) = 1$. Then the Dehn filled manifold $Y(\mu_0, \mu_1)$ is homeomorphic to S^3.*

Proof There is a mapping class $\varphi \in SL(2, \mathbb{Z})$, mapping $T^2 \to T^2$, which sends the slope μ_0 to $0/1$ and the slope μ_1 to $\pm 1/0$. Now, if we identify the target T^2 with the standard Heegaard torus of S^3, the product homeomorphism $(\varphi \times \mathrm{id})$ maps $Y = T^2 \times I$ to the complement of the Hopf link in S^3, with μ_0 and μ_1 mapped to the meridians of the two link components. Filling these in gives S^3. □

We are now ready to prove the main theorem of this section.

Theorem 6.5 *For any $L > 0$, there exists a tunnel number one knot in S^3 with exactly two unknotting tunnels, each of which has length at least L.*

Proof The knot in S^3 will be obtained by Dehn filling two cusps of the manifold $X = (T^2 \times I) \smallsetminus K$ of Lemma 6.2. In particular, we will fill cusps corresponding to $T_0 := T^2 \times \{0\}$ and $T_1 := T^2 \times \{1\}$. Note one of the two core tunnels of $(T^2 \times I) \smallsetminus K$, call it σ_0, runs from K to T_0, and the other, σ_1, from K to T_1. After appropriate Dehn filling, we will see that Theorem 2.11 applies to give unknotting tunnels τ_i corresponding to σ_i, and that by Theorem 2.3 these are the only unknotting tunnels of the resulting knot.

We need to take care in our choice of slopes, to ensure that the tunnels stay long, and also to ensure the resulting filled manifold is a knot complement in S^3 with no new genus-2 Heegaard splittings.

Put the same framing on T_0 and on T_1, and choose disjoint horocusps about these tori. Set $J = 1.1$, and $\epsilon = 0.29$, as Section 3. Also, choose a length $L > 0$. Now, we will choose slopes μ_0 on T_0 and μ_1 on T_1, such that the following hold:

(A) The normalized lengths $L(\mu_i)$ are long enough to satisfy Theorem 2.6 for $J = 1.1$ and $\epsilon = 0.29$. In fact, each $L(\mu_i)$ is at least J^2 times longer than necessary to apply the drilling theorem. This way, we can apply Theorem 2.6 twice: once to fill T_0 and a second time to fill T_1 (or in the opposite order).

(B) On the disjoint horocusps about T_0 and T_1, the length of each μ_i (on its respective torus) satisfies $\ell(\mu_i) > 152$.

(C) For each slope μ_i, the shortest longitude λ_i satisfies $\ell(\lambda_i) > 7$ and $\ell(\lambda_i) > e^{3+L/2}$. Again, these lengths are measured on the chosen horospherical tori about T_0 and T_1. Note that *every* longitude for μ_i satisfies the same lower bound on length.

(D) The intersection number is $\iota(\mu_0, \mu_1) = 1$.

We claim that in the complement of a bounded region in the Farey graph $\mathcal{F}$, conditions (A)–(D) hold simultaneously. In the language of the Farey graph, conditions (A) and (B) require μ_0 and μ_1 to avoid finitely many vertices of $\mathcal{F}$. Condition (C) requires each of μ_0 and μ_1 to avoid finitely many closed balls of radius 1 in $\mathcal{F}$, where each ball is the set of all Farey neighbors of a longitude too short for (C). Finally, condition (D) requires μ_0 to be a Farey neighbor of μ_1. Thus, by choosing an edge $[\mu_0, \mu_1] \subset \mathcal{F}$ that lies outside the bounded prohibited region, we satisfy all of (A)–(D).

We Dehn fill T_0 along μ_0, and T_1 along μ_1. By Lemma 6.4, the result is a knot in S^3.

Applying Theorem 2.3 twice (once to fill T_0, and again to fill T_1), we conclude that the resulting knot complement $X(\mu_0, \mu_1)$ has Heegaard genus 2, with any genus-2 Heegaard surfaces coming from the Heegaard surfaces of the original manifold X. By Lemma 6.3, there are exactly two Heegaard surfaces in X (with core tunnels σ_0 running from K to T_0, and σ_1 running from K to T_1). Thus there are exactly two unknotting tunnels in $X(\mu_0, \mu_1)$.

We claim that the two unknotting tunnels for the resulting knot are associated to σ_0 and σ_1. To see this, first fill one of the cusps, say T_0. Then σ_1, running from K to T_1, remains a core tunnel in its compression body, and is now an unknotting tunnel for

$X(\mu_0)$. Now, Theorem 2.11 implies that $\tau_1 = \overline{\tau}(\sigma_1, \mu_1)$, is an unknotting tunnel for the knot complement $X(\mu_0, \mu_1)$. Similarly, if we first fill T_1, then Theorem 2.11 implies that $\tau_0 = \overline{\tau}(\sigma_0, \mu_0)$, is an unknotting tunnel for the knot complement $X(\mu_0, \mu_1)$. Thus we have found exactly two unknotting tunnels, and these must be all the tunnels for the manifold.

It remains to show that the tunnels have length at least L. First, fill X along slope μ_0 on T_0. Then, by Theorem 2.6(d), there is a J–bilipschitz diffeomorphism ϕ that maps the horocusp about T_1 in X to an embedded horocusp about T_1 in $X(\mu_0)$. Since $J = 1.1$, the meridian slope $\phi(\mu_1)$ has length at least

$$\tag{6-1} \ell(\mu_1)/1.1 > 152/1.1 > 138.$$

Similarly, the shortest longitude for $\phi(\mu_1)$ in $X(\mu_0)$ has length at least

$$\tag{6-2} \ell(\lambda_1)/1.1 > e^{3+L/2} \cdot e^{-0.1} = e^{(5.8+L)/2}.$$

Thus, by equation (6-1) and condition (A), Dehn filling $X(\mu_0)$ along slope μ_1 on T_1 will satisfy all the hypotheses of Theorem 3.9. Therefore, by Theorem 3.9 and equation (6-2), the unknotting tunnel τ_1 of $X(\mu_0, \mu_1)$ will have length longer than L.

By the same argument, reversing the order of the fillings, the other unknotting tunnel τ_0 of $X(\mu_0, \mu_1)$ will also have length longer than L. □

7 A concrete example

In this section, we take another look at the construction of Section 6, giving a concrete example. As a result, in Theorem 7.9 we obtain a thoroughly effective version of Theorem 6.5, with an explicit sequence of knots and an explicit bound on the length of their tunnels. The downside of this concrete construction is that rigorous computer assistance will be required at two places in the argument (Lemmas 7.1 and 7.3).

7.1 The manifold X and its Heegaard splittings

For the entirety of this section, we will work with the knot K in $T^2 \times I$ that is depicted in Figure 11. It is a pleasant exercise to show that this knot is obtained by $2/1$ Dehn filling on one component of the alternating chain link C from Section 6. Since we will not need this fact, we omit the derivation.

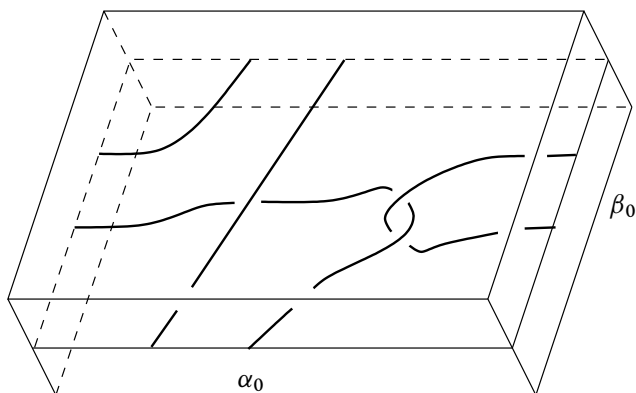The following lemma collects a few useful facts about $X = (T^2 \times I) \smallsetminus K$.

Figure 11: Diagram of a knot K in $T^2 \times I$. Note the framing: α_0 runs to the right, β_0 runs to meet α_0 as shown. Slopes α_1 and β_1 are on the top torus, parallel to α_0 and β_0, respectively.

Lemma 7.1 *Let* $X = (T^2 \times I) \smallsetminus K$ *be the complement of the knot in Figure 11. Then:*

(a) *X has an ideal triangulation consisting of seven tetrahedra, as depicted in Figure 12.*

(b) *X is isometric to the SnapPea census manifold $\mathtt{v}_{3227}$.*

(c) *The basis $\langle \alpha_i, \beta_i \rangle$ for $H_1(T^2 \times \{i\})$, shown in Figure 11, places the same framing on $T_0 := T^2 \times \{0\}$ and $T_1 := T^2 \times \{1\}$.*

(d) *In the hyperbolic metric on X, there is a unique shortest edge from K to T_0 (edge e_0, marked 0 in Figure 12), and a unique shortest edge from K to T_1 (edge e_1, marked 1).*

(e) *On the maximal cusp about T_0, the slopes α_i and β_i are realized by parabolic translations*

(7-1) $$\alpha_0 \colon z \mapsto z + 2.383, \quad \beta_0 \colon z \mapsto z + 4.222 + 2.657\sqrt{-1}.$$

On the maximal cusp about T_1, the slopes α_i and β_i are realized by parabolic translations

(7-2) $$\alpha_1 \colon z \mapsto z + 7.961 + 1.269\sqrt{-1}, \quad \beta_1 \colon z \mapsto z + 4.989.$$

The real and imaginary parts in (7-1) and (7-2) are accurate to within 0.01.

Proof Conclusions (a), (b) and (c) are immediate consequences of rigorous routines in SnapPy [14]. In particular, the construction of an ideal triangulation is rigorous.
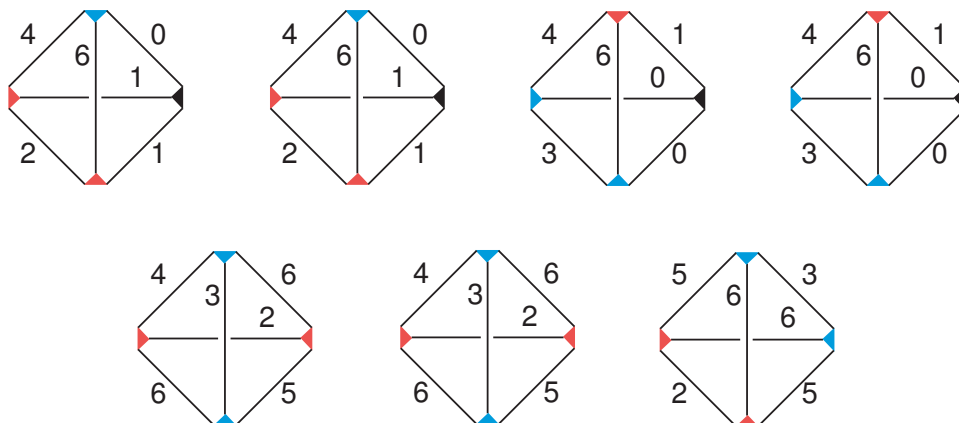
Figure 12: The 7 tetrahedra in an ideal triangulation of $X = (T^2 \times I) \smallsetminus K$. The cusp of T_0 is blue, the cusp of T_1 is red, and the cusp of K is black. For convenience in the ensuing argument, edge labels are permuted from the labels in the SnapPea census, via the permutation $e_0 \leftrightarrow e_5$, $e_1 \leftrightarrow e_6$, $e_2 \leftrightarrow e_3$.

The isometry checker is rigorous, because (in the case of a "yes" answer) it exhibits a simplicial isomorphism between X and v_{3227}. Similarly, the realization of particular peripheral curves as moves in the triangulation is a rigorous combinatorial operation. Note that, once the meridian of K is filled, α_0 becomes isotopic to α_1, and β_0 to β_1.

Next, we turn to the work of Moser [35], which can be summarized as follows. Whenever SnapPy produces an approximate solution to the gluing equations for a triangulation, with sufficiently small error, then an exact solution exists nearby (where "nearby" is explicitly quantified, with Lipschitz estimates on the distance between approximate tetrahedron shapes and true tetrahedron shapes). For every census manifold, including $X \cong v_{3227}$, Moser verifies that the error in SnapPy's approximate solution to the gluing equations is bounded by

$$(7\text{-}3) \qquad\qquad\qquad\qquad |b| < 1.8 \cdot 10^{-26},$$

easily enough to ensure a true solution nearby. Thus X is hyperbolic.

Moser's estimate on the distance between an approximate solution and a true solution also gives error bounds on every geometric quantity computed by SnapPy. It follows that the edge e_0 (marked 0 in Figure 12) is the unique shortest geodesic between K and T_0, because the next shortest edge is significantly longer. Similarly, the edge e_1 (marked 1 in Figure 12) is the unique shortest geodesic between K and T_1, proving (d).

Finally, SnapPy computes the action of α_0 and β_0 on the complex plane covering the maximal cusp torus about T_0. The computed values are as in (7-1). Moser's error

bound on the gluing equations, expressed in (7-3), implies that the error in equation (7-1) is significantly less than 0.01. Similarly, the lengths of α_1 and β_1, as computed in (7-2), are accurate to an error much less than 0.01. □

Next, we turn our attention to Heegaard splittings of X.

Lemma 7.2 *The arcs σ_0 and σ_1, depicted in Figure 13, are core tunnels for genus-2 Heegaard splittings of the manifold $X = (T^2 \times I) \smallsetminus K$.*
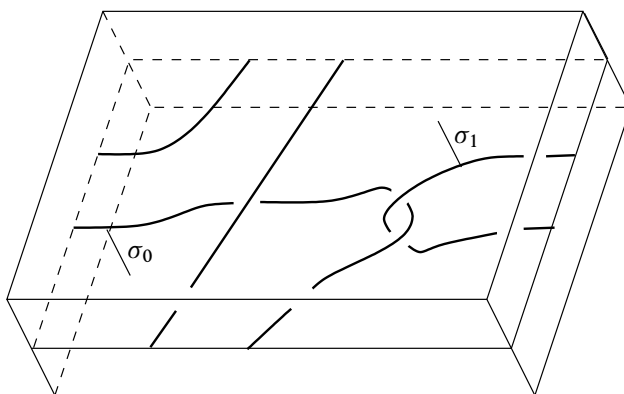

Figure 13: The arcs σ_0 and σ_1 are core tunnels for $X = (T^2 \times I) \smallsetminus K$.
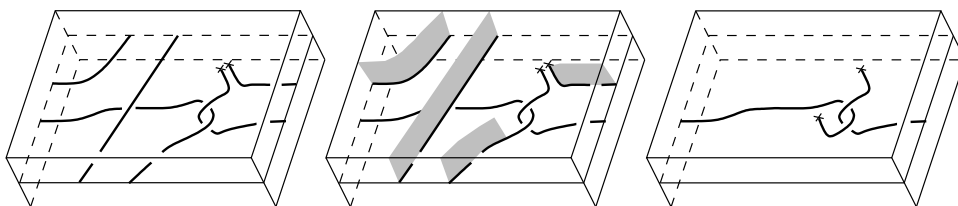

Figure 14: Manifolds homeomorphic to $X \smallsetminus \sigma_1$

Proof We prove this for the arc σ_1, as the proof for σ_0 is similar. Let $V \subset X$ be a regular neighborhood of $T_1 \cup K \cup \sigma_1$. Then $\overline{V}$ is obtained by attaching a 1–handle to the cusp tori of T_1 and K, along arc σ_1. Thus $\overline{V}$ is a compression body whose positive boundary is a genus-2 surface $S_1 = \partial_+ V$, and whose core tunnel is σ_1.

It remains to show that $W = X \smallsetminus V$ is also a compression body. To see this, note that $X \smallsetminus V$ is homeomorphic to $X \smallsetminus \sigma_1 = (T^2 \times I) \smallsetminus (K \cup \sigma_1)$. Furthermore, the complement of $K \cup \sigma_1$ in $T^2 \times I$ is homeomorphic to the manifold shown on the left of Figure 14,

where we now are taking the complement of an arc with endpoints on the top boundary of $T^2 \times I$.

Drag one of these endpoints along the torus $T_1 = T^2 \times \{1\}$, following the disk shown in the middle part of Figure 14. The result is shown on the right of that figure. Now untwist. The result is a manifold homeomorphic to $T^2 \times I$, with a simple boundary-parallel arc removed. Thus $W = X \smallsetminus V$ is a genus-2 compression body. □

In fact, σ_0 and σ_1 lead to the only two genus-2 Heegaard surfaces of X.

Lemma 7.3 *The manifold X of Figure 13 has exactly two genus-2 Heegaard surfaces S_0 and S_1, associated to tunnels σ_0 and σ_1.*

Proof Let $S \subset X$ be a genus-2 Heegaard surface. Since X is hyperbolic, S must be strongly irreducible (every compression disk on one side of S intersects every compression disk on the other side of S). Thus, by a theorem of Rubinstein [38] and Stocking [44], S must be isotopic to an *almost normal* surface in the triangulation of Figure 12. Recall that a *normal* surface intersects every tetrahedron in a disjoint union of normal triangles and quadrilaterals, as in the left two panels of Figure 15. An *almost normal* surface is composed of triangles and quads, as well as exactly one octagon or tube, as in the right two panels of Figure 15.

The program Regina [12] can perform a rigorous combinatorial analysis of normal and almost normal surfaces in X. In particular, Regina verifies that X contains no almost normal genus-2 surfaces with an octagon. Thus S must contain a tube.

Because S has genus 2, compressing S along the tube will produce one or two normal tori. But since the triangulation of Figure 12 supports a positively oriented solution to the gluing equations (by Lemma 7.1), any normal torus must be boundary-parallel, composed entirely of vertex-linking triangles. (See, eg, Lackenby [28, Proposition 4.4].) Thus the almost normal tube in S is obtained by tubing together two non-parallel triangles in one tetrahedron.

Any tube between two normal triangles is isotopic to the neighborhood of an edge in the triangulation. Thus, since there are seven edges in Figure 12, we must consider seven tubed surfaces $S_0, \ldots, S_6$, in one-to-one correspondence with edges $e_0, \ldots, e_6$. For each S_i, let V_i be the closure of the component of $X \smallsetminus S_i$ that contains the 1–handle through the tube. Since V_i is obtained by joining together one or two cusp tori along a 1–handle, it is a compression body.

For each almost normal tubed surface S_i, we begin isotoping S_i toward its associated edge e_i, according to the tightening algorithm of Schleimer [42, Sections 7–9]. In

the case at hand, Schleimer's tightening procedure constructs an embedded isotopy between each S_i and a normal surface in the triangulation.

The tubed surfaces S_0 and S_2 tighten to the same normal surface, from opposite sides. Thus these surfaces are isotopic, and furthermore $V_0 \cong \overline{X \smallsetminus V_2}$. Thus S_0 splits X into compression bodies V_0 and V_2. Similarly, S_1 and S_3 tighten to the same normal surface, from opposite sides. Thus S_1 splits X into compression bodies V_1 and V_3.
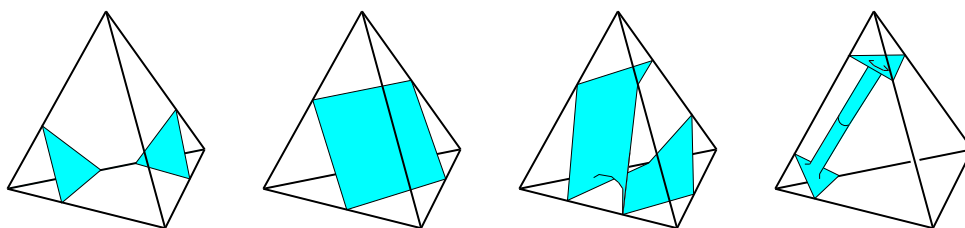

Figure 15: Left to right: Normal triangles. A normal quadrilateral. An almost normal octagon. An almost normal tube between two vertex-linking triangles. (Graphics by Saul Schleimer.)

It remains to check that S_i is not a Heegaard surface for $i = 4, 5, 6$. This follows from a theorem of Bachman [7]. He proves that if a normal surface $T \subset X$ is a Heegaard surface, then there must be two almost normal surfaces isotopic to X. (For example, this is the situation for S_0 and S_2.) In our case, each of S_4, S_5 and S_6 separates the cusp of K from the cusps T_0 and T_1 (see edges $4, 5, 6$ in Figure 12). Thus, if two of these surfaces S_i and S_j were isotopic, then the compression bodies V_i and V_j on the side of T_0 and T_1 would also be isotopic. But the core tunnels of these compression bodies, namely edges e_i and e_j, are distinct edges of a geometric triangulation, hence non-isotopic. This is a contradiction.

Alternately, one may show that S_i is not a Heegaard surface for $i = 4, 5, 6$ by using Regina to cut X along the normal surface isotopic to S_i. Regina can retriangulate the resulting pieces. Then, it verifies that the piece $\overline{X \smallsetminus V_i}$ is not a compression body, by showing that it has incompressible boundary. (If the boundary was compressible, there would have to be a normal compression disk.) Since cutting and retriangulating greatly increases the number of tetrahedra, this verification took six hours of runtime.

We conclude that S_0 is the only Heegaard surface of X that separates T_1 from $T_0 \cup K$. Since the core tunnel σ_0 connects T_0 to K, it must be a core tunnel for S_0, isotopic to edge e_0 from $T_0 \cup K$. Similarly, since σ_1 connects T_1 to K, it must be a core tunnel for S_1, isotopic to edge e_1 in the triangulation. $\qquad\square$

As a corollary of the above argument, we obtain:

Lemma 7.4 *For $i = 0, 1$, the core tunnel σ_i of Figure 13 is isotopic to the unique shortest geodesic in X between the cusps of K and T_i.*

Proof The analysis of almost normal surfaces in Lemma 7.3 shows that σ_i is isotopic to edge e_i, labeled i in the triangulation of Figure 12. By Lemma 7.1, edge e_i is the unique shortest geodesic between K and T_i. $\qquad\square$

7.2 Fibonacci slopes

As in Section 6, we will construct a sequence of knots in S^3 by filling cusp tori $T_0 = T^2 \times \{0\}$ and $T_1 = T^2 \times \{1\}$. We choose the filling slopes as follows.

Definition 7.5 For every integer $n \geq 0$, let f_n be the nth Fibonacci number (with $f_0 = 0$ and $f_1 = 1$). Then, for each $n \geq 1$, define the filling slopes

$$(7\text{-}4) \qquad \mu_0^n = f_n \alpha_0 + f_{n+1} \beta_0, \qquad \mu_1^n = f_{n-1} \alpha_1 + f_n \beta_1,$$

where α_i and β_i are as in Figure 11.

Lemma 7.6 *The slopes μ_0^n and μ_1^n have intersection number $\iota(\mu_0^n, \mu_1^n) = 1$. Therefore, by Lemma 6.4, filling cusp T_0 along μ_0^n and cusp T_1 along μ_1^n produces a knot $K_n \in S^3$.*

Proof It is well-known that Fibonacci numbers can be produced by the explicit formula

$$(7\text{-}5) \qquad \begin{bmatrix} f_{n-1} & f_n \\ f_n & f_{n+1} \end{bmatrix} = A^n, \quad \text{where } A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

Since the left column of A^n expresses the slope μ_1^n, and the right column expresses the slope μ_0^n, we have

$$\iota(\mu_0^n, \mu_1^n) = |\det(A^n)| = |\det A|^n = 1.$$

Thus, by Lemma 6.4, filling cusp T_i along μ_i^n produces a knot $K_n \in S^3$.

In fact, recalling the proof of Lemma 6.4 allows a concrete way to visualize the knot K_n. If we embed the slab $T^2 \times I$ in the complement of the Hopf link in S^3 via the mapping class

$$\begin{bmatrix} f_{n-1} & (-1)^n f_n \\ f_n & (-1)^n f_{n+1} \end{bmatrix} \in SL(2, \mathbb{Z}),$$

the curves μ_0^n and μ_1^n will be mapped to the meridians of the two Hopf components. Then, filling the two Hopf components along their meridians in S^3 (ie, erasing these two link components from the diagram) will leave the knot $K_n \in S^3$. $\qquad\square$

Lemma 7.7 Let $n \geq 5$. Then, on a maximal horocusp about $T_i \subset X$, the slope μ_i^n of Definition 7.5 has length

$$(7\text{-}6) \qquad 4.3\,\varphi^n < \ell(\mu_i^n) < 4.7\,\varphi^n,$$

where $\varphi = (1 + \sqrt{5})/2$ is the golden ratio. Furthermore, the shortest longitude for μ_i^n is $\lambda_i^n = \mu_i^{n-2}$.

Proof The Fibonacci number f_n has the closed form expression

$$(7\text{-}7) \qquad f_n = \frac{\varphi^n - (-\varphi)^{-n}}{\sqrt{5}},$$

which can be derived by diagonalizing the matrix A of equation (7-5). Thus the slope μ_0^n of Definition 7.5 can be written as

$$(7\text{-}8) \qquad \begin{aligned} \mu_0^n &= \left(\frac{\varphi^n - (-\varphi)^{-n}}{\sqrt{5}}\right)\alpha_0 + \left(\frac{\varphi^{n+1} - (-\varphi)^{-(n+1)}}{\sqrt{5}}\right)\beta_0 \\ &= \left(\frac{\varphi^n}{\sqrt{5}}\right)(\alpha_0 + \varphi\beta_0) + \left(\frac{(-\varphi)^{-(n+1)}}{\sqrt{5}}\right)(\varphi\alpha_0 - \beta_0). \end{aligned}$$

Note that when $n \geq 3$, we have

$$(7\text{-}9) \qquad \varphi^n \geq \varphi^7 \cdot \varphi^{-(n+1)} > 29\,\varphi^{-(n+1)}.$$

Thus, substituting (7-9) and the translation lengths of (7-1) into equation (7-8) gives:

$$\frac{\varphi^n}{\sqrt{5}}\ell(\alpha_0 + \varphi\beta_0) - \frac{\varphi^n\,\ell(\varphi\alpha_0 - \beta_0)}{\sqrt{5}\cdot 29} < \ell(\mu_0^n) < \frac{\varphi^n}{\sqrt{5}}\ell(\alpha_0 + \varphi\beta_0) + \frac{\varphi^n\,\ell(\varphi\alpha_0 - \beta_0)}{\sqrt{5}\cdot 29}$$

$$\frac{\varphi^n}{\sqrt{5}}(10.1) - \frac{\varphi^n}{\sqrt{5}}\left(\frac{2.7}{29}\right) < \ell(\mu_0^n) < \frac{\varphi^n}{\sqrt{5}}(10.2) + \frac{\varphi^n}{\sqrt{5}}\left(\frac{2.7}{29}\right)$$

$$4.4\,\varphi^n < \ell(\mu_0^n) < 4.7\,\varphi^n$$

An identical calculation, using

$$(7\text{-}10) \qquad \mu_1^n = \left(\frac{\varphi^{n-1}}{\sqrt{5}}\right)(\alpha_1 + \varphi\beta_1) + \left(\frac{(-\varphi)^{-n}}{\sqrt{5}}\right)(\varphi\alpha_1 - \beta_1),$$

the estimate (7-9), and the translation lengths of (7-2) gives

$$4.3\,\varphi^n < \ell(\mu_1^n) < 4.6\,\varphi^n.$$

Thus both μ_0^n and μ_1^n satisfy the estimates of (7-6).

Now, consider the shortest longitude λ_i^n of μ_i^n. Since μ_1^n is the same slope on T^2 as μ_0^{n-1}, Lemma 7.6 implies that $\iota(\mu_0^{n-1}, \mu_0^n) = 1$. Thus, by definition, μ_0^{n-1} is a longitude for μ_0^n. Furthermore, every longitude for μ_0^n must have the form

$$\mu_0^{n-1} + k\mu_0^n, \quad \text{where } k \in \mathbb{Z}.$$

Now, observe from equations (7-8) and (7-9) that both μ_0^{n-1} and μ_0^n are nearly parallel to the vector $(\alpha_0 + \varphi\beta_0)$, hence nearly parallel to one another. Furthermore, by (7-6), the lengths of μ_0^{n-1} and μ_0^n differ by a ratio of at most $(4.7/4.3)\varphi < 2$. Thus the shortest longitude of the form $\mu_0^{n-1} + k\mu_0^n$ will have $k = -1$, and the shortest longitude is the slope

$$
\begin{aligned}
\lambda_0^n &= \mu_0^{n-1} - \mu_0^n \\
&= (f_{n-1}\alpha_0 + f_n\beta_0) - (f_n\alpha_0 + f_{n+1}\beta_0) \\
&= -(f_{n-2}\alpha_0 + f_{n-1}\beta_0) \\
&= -\mu_0^{n-2}.
\end{aligned}
$$

(The minus sign is immaterial, since we do not need an orientation on λ_0^n.) By an identical argument, the shortest longitude for μ_1^n is $\lambda_1^n = \mu_1^{n-2}$. $\qquad\square$

Remark 7.8 Although Fibonacci numbers are convenient for the proof of Lemma 7.7, in fact the construction has many generalizations. The key fact that makes the entire argument work is that the sequence of slopes comes from powers of a pseudo-Anosov matrix A, as in equation (7-5). If we had defined a sequence of slopes μ^n by iterating some other pseudo-Anosov matrix $B \in \mathrm{SL}(2, \mathbb{Z})$, then it would follow, in analogy to equation (7-8), that for sufficiently large n the slopes μ^n are nearly parallel to the stable foliation of B. The lengths $\ell(\mu^n)$ of slopes defined in this way would be controlled by the dilatation (or largest eigenvalue) of B, and the shortest longitude of μ^n would be immediately visible.

In this way, ideas from hyperbolic geometry play a significant role in a combinatorial construction on the torus.

7.3 Completing the argument

We may now Dehn fill X, along slope μ_i^n on cusp T_i, to produce a knot $K_n \subset S^3$. The following result about the unknotting tunnels of K_n immediately implies Theorem 1.7 in the introduction.

Theorem 7.9 Let μ_0^n and μ_1^n be the Dehn filling slopes of Definition 7.5. Then, for sufficiently large n, the knot $K_n \subset S^3$ obtained by filling $T_i \subset X$ along μ_i^n has exactly

two unknotting tunnels τ_0^n and τ_1^n. Both tunnels are isotopic to canonical geodesics in $S^3 \smallsetminus K_n$, and both tunnels have length satisfying

$$2n \ln(\varphi) - 4.8 < \ell(\tau_i^n) < 2n \ln(\varphi) + 5.9,$$

where $\varphi = (1 + \sqrt{5})/2$ is the golden ratio.

Proof The proof parallels the proof of Theorem 6.5, with more explicit estimates. As in that proof, let $J = 1.1$ and $\epsilon = 0.29$. Then, by Lemma 7.7, choosing a sufficiently large n ensures the following:

(A) For $n \gg 0$, the normalized lengths $L(\mu_i^n)$ are J^2 times longer than necessary to satisfy the Drilling Theorem 2.6, for $J = 1.1$ and $\epsilon = 0.29$. This way, we can apply Theorem 2.6 twice: once to fill T_0 and a second time to fill T_1 (or in the opposite order).

(B) For $n \geq 8$, the length of each μ_i^n (on its respective horospherical torus) satisfies

$$\ell(\mu_i^n) > 4.3 \, \varphi^8 > 152.$$

(C) For $n \geq 5$, the shortest longitude $\lambda_i^n = \mu_i^{n-2}$ satisfies $\ell(\lambda_i) > 4.3 \, \varphi^3 > 7$.

Applying Theorem 2.3 twice (once to fill T_0, and again to fill T_1), we conclude that the resulting knot complement $S^3 \smallsetminus K_n$ has Heegaard genus 2, with any genus-2 Heegaard surfaces coming from the Heegaard surfaces of the original manifold X. Recall that by Lemma 7.3, there are exactly two Heegaard surfaces in X (with core tunnels σ_0 running from K to T_0, and σ_1 running from K to T_1).

When n is sufficiently large and μ_0^n is sufficiently long, filling torus T_0 along μ_0^n will preserve the property (from Lemma 7.4) that σ_1 is the shortest geodesic from K to T_1. Note that in the two-cusped manifold $X(\mu_0^n)$, the arc σ_1 is a core tunnel of a compression body V_1, and the complement of V_1 is a genus-2 handlebody. Thus σ_1 is an unknotting tunnel of $X(\mu_0^n)$, and is the shortest arc between the two cusps of $X(\mu_0^n)$. Thus, by Theorem 1.6, the unknotting tunnel τ_1^n of K_n that is associated to σ_1 must be a canonical geodesic in $S^3 \smallsetminus K_n$. (The filling of T_1 along μ_1^n will be generic because choosing n large ensures that both μ_1^n and λ_1^n are arbitrarily long.)

In a similar way, if we choose n sufficiently large and start by filling T_1 along μ_1^n, the unknotting tunnel σ_0 of $X(\mu_1^n)$ will be the shortest arc between the two cusps of $X(\mu_0^n)$. Thus, by Theorem 1.6, the unknotting tunnel τ_0^n of K_n that is associated to σ_0 must be a canonical geodesic in $S^3 \smallsetminus K_n$.

Note that by Lemma 4.5, the large majority of the length of τ_0^n is contained in the Margulis tube created by filling T_0. Similarly, the large majority of the length of τ_1^n is

contained in the Margulis tube created by filling T_1. Thus the two unknotting tunnels τ_0^n and τ_1^n are distinct, and account for the only two genus-2 Heegaard splittings of $S^3 \smallsetminus K_n$.

It remains to estimate the lengths of τ_0^n and τ_1^n. First, fill X along slope μ_0^n on T_0. Then, by Theorem 2.6(d), there is a J–bilipschitz diffeomorphism ϕ that maps a maximal horocusp about T_1 in X to a self-tangent (hence, maximal) horocusp about T_1 in $X(\mu_0)$. Applying Lemma 7.7 with error bounded by $J = 1.1$, we conclude that in $X(\mu_0^n)$, the meridian $\phi(\mu_1^n)$ and its shortest longitude $\varphi(\lambda_1^n)$ have lengths satisfying

$$(7\text{-}11) \qquad \ell(\mu_1^n) > 152/1.1 > 138 \quad \text{and} \quad \frac{4.3}{1.1}\varphi^{n-2} < \ell(\lambda_1^n) < 4.7 \cdot 1.1\,\varphi^{n-2}.$$

Thus, by equation (7-11) and condition (A), Dehn filling $X(\mu_0^n)$ along slope μ_1^n on T_1 will satisfy all the hypotheses of Theorem 3.9. Therefore, by Theorem 3.9, the unknotting tunnel τ_1^n associated to σ_1 will have length satisfying:

$$2\ln\left(\frac{4.3}{1.1}\varphi^{n-2}\right) - 5.6 < \ell(\tau_1^n) < 2\ln(4.7 \cdot 1.1\,\varphi^{n-2}) + 4.5$$

$$2n\,\ln(\varphi) + 2\ln\left(\frac{4.3}{1.1}\varphi^{-2}\right) - 5.6 < \ell(\tau_1^n) < 2n\,\ln(\varphi) + 2\ln(4.7 \cdot 1.1\,\varphi^{-2}) + 4.5$$

$$2n\,\ln(\varphi) - 4.8 < \ell(\tau_1^n) < 2n\,\ln(\varphi) + 5.9$$

By the same argument, reversing the order of the fillings, the other unknotting tunnel τ_0^n of $S^3 \smallsetminus K_n$ satisfies the same estimates on length. $\qquad\square$

Appendix A: Variations on the law of cosines

The goal of this appendix is to write down versions of the law of cosines that work for triangles with a mix of material and ideal vertices. The formulae of Lemmas A.2 and A.3 can likely be derived from the extensive tables compiled by Guo and Luo [24, Appendix]. We prefer to begin with the law of cosines for ordinary triangles in $\mathbb{H}^2$.

Lemma A.1 (Law of cosines) *Let Δ be a triangle in $\mathbb{H}^2$, with side-lengths a, b, c. Let α be the angle opposite side a. Then*

$$(A\text{-}1) \qquad\qquad \cosh a = \cosh b \, \cosh c - \sinh b \, \sinh c \, \cos\alpha.$$

Proof See [45, Equation 2.4.9] or [17, Section VI.3.5]. $\qquad\square$

When Δ has an ideal vertex, the natural analogue of an angle α is the length of a horocycle truncating this ideal vertex. As a result, the law of cosines takes the following form. (We prove Lemma A.2 for isosceles triangles, but the general case is not much harder.)
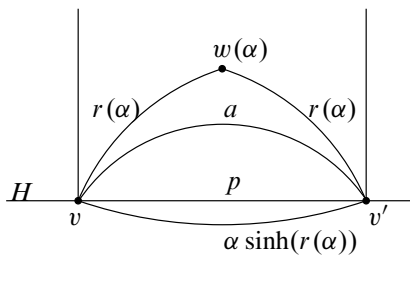
Figure 16: Lengths on 1/3–ideal triangle

Lemma A.2 *Let Δ be a 1/3 ideal triangle in $\mathbb{H}^2$. Let H be a horoball truncating the ideal vertex of Δ, and suppose that the two material vertices v, v' are equidistant from H. Let a denote the side-length opposite the ideal vertex, let b denote the distance from H to v or v', and let p denote the length of the horocyclical segment $\partial H \cap \Delta$. Then,*

$$(A\text{-}2) \qquad 2\sinh(a/2) = pe^b.$$

Proof First, observe that the quantity pe^b on the right-hand side of equation (A-2) is independent of the choice of horoball H. This is because increasing the size of H by hyperbolic distance d will increase the horocycle p by a factor of e^d while subtracting d from the side-length b. Thus the right-hand side of (A-2) remains unchanged. As a result, no generality is lost in assuming that $b = 0$, or equivalently that vertices $v, v' \in \partial H$.

For every angle $\alpha \in (0, \pi)$, let Δ_α be an isosceles triangle in $\mathbb{H}^2$ that has vertices at v and v', and a third vertex $w(\alpha)$ with angle α. Then the points v and v' lie on the same circle centered at $w(\alpha)$, whose radius we denote by $r(\alpha)$. The length of the circle arc from v to v' is $\alpha \sinh r(\alpha)$. See Figure 16.

By the law of cosines (A-1), the side-length a of Δ_α satisfies

$$(A\text{-}3) \qquad \cosh a = \cosh^2 r(\alpha) - \sinh^2 r(\alpha) \cos \alpha$$

$$= 1 + \sinh^2 r(\alpha)(1 - \cos \alpha)$$

$$= 1 + \left(\alpha \sinh r(\alpha)\right)^2 \left(\frac{1 - \cos \alpha}{\alpha^2}\right)$$

$$= 1 + \left(\lim_{\alpha \to 0} \alpha \sinh r(\alpha)\right)^2 \left(\lim_{\alpha \to 0} \frac{1 - \cos \alpha}{\alpha^2}\right)$$

$$= 1 + p^2 \cdot \tfrac{1}{2}.$$

The last equality holds because as $\alpha \to 0$, the triangle Δ_α converges to Δ, and the circle of radius $r(\alpha)$ converges to the horocycle ∂H through v and v'. Thus the circle arc of length $\alpha \sinh r(\alpha)$ converges to the horocyclical segment of length p. Now, solving equation (A-3) for p and recalling that $b = 0$ produces (A-2), as desired. $\qquad\square$

For $2/3$ ideal triangles, we have the following version of the law of cosines.

Lemma A.3 *Let Δ be a $2/3$ ideal triangle in $\mathbb{H}^2$. Suppose that the two sides of Δ meeting at the material vertex v are labeled ρ and ρ', and the angle at v is α. Suppose the third side of Δ is labeled g. Choose horoball neighborhoods H and H' about the two ideal vertices. Then, relative to the horoballs H and H',*

$$\text{(A-4)} \qquad \ell(g) = \ell(\rho) + \ell(\rho') + \ln\left(\frac{1-\cos\alpha}{2}\right).$$

Proof As in the last proof, we begin by observing that the quantity

$$\text{(A-5)} \qquad \ell(g) - (\ell(\rho) + \ell(\rho'))$$

is independent of the choice of horoball neighborhoods. This is because expanding horoball H will decrease both $\ell(g)$ and $\ell(\rho)$ by the same amount. Similarly, adjusting the size of H' will affect $\ell(g)$ and $\ell(\rho')$ by the same amount. Thus, no generality is lost in assuming that $\ell(\rho) = \ell(\rho')$. Given this assumption, we will show that the quantity (A-5) is equal to $\ln(1 - \cos\alpha)/2$.

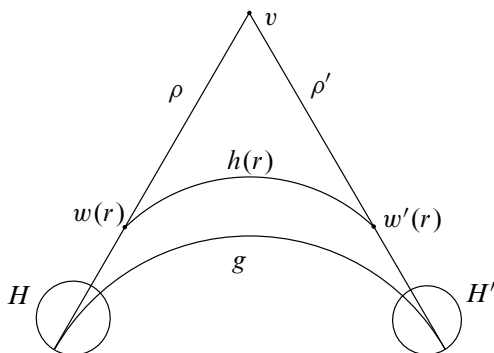

Figure 17: The setup of Lemma A.3

For each $r > 0$, let $w(r)$ be the point on ρ that is distance r from v, and let $w'(r)$ be the corresponding point on ρ'. See Figure 17. By Lemma A.1, the distance between $w(r)$ and $w'(r)$ is

$$h(r) = \cosh^{-1} f(r), \quad \text{where } f(r) = \cosh^2 r - \sinh^2 r \cos\alpha.$$

Now, as $r \to \infty$, the geodesic between $w(r)$ and $w'(r)$ approaches g. In particular, this geodesic fellow-travels ρ and ρ' for a greater and greater portion of its length. Thus $h(r) - 2r$ becomes a better and better approximation to the quantity $\ell(g) - 2\ell(\rho)$. Therefore,

$$
\begin{aligned}
\ell(g) - 2\ell(\rho) &= \lim_{r \to \infty} h(r) - 2r \\
&= \lim_{r \to \infty} \cosh^{-1}(f(r)) - 2r \\
&= \lim_{r \to \infty} \ln\left(f(r) + \sqrt{f(r)^2 - 1} \right) - 2r \\
&= \lim_{r \to \infty} \ln(2f(r)) - 2r \\
&= \lim_{r \to \infty} \ln(2\cosh^2 r - 2\sinh^2 r \cos\alpha) - 2r \\
&= \lim_{r \to \infty} \ln\left(\frac{e^{2r}}{2} - \frac{e^{2r}}{2}\cos\alpha \right) - 2r \\
&= \lim_{r \to \infty} 2r + \ln\left(\frac{1 - \cos\alpha}{2} \right) - 2r \\
&= \ln\left(\frac{1 - \cos\alpha}{2} \right). \qquad \square
\end{aligned}
$$

References

[1] **C Adams**, *Unknotting tunnels in hyperbolic 3–manifolds*, Math. Ann. 302 (1995) 177–195   MR1329452

[2] **C C Adams**, *Waist size for cusps in hyperbolic 3–manifolds*, Topology 41 (2002) 257–270   MR1876890

[3] **C C Adams**, *Waist size for cusps in hyperbolic 3–manifolds, II*, preprint available from the author (2002)

[4] **C C Adams**, **A W Reid**, *Unknotting tunnels in two-bridge knot and link complements*, Comment. Math. Helv. 71 (1996) 617–627   MR1420513

[5] **H Akiyoshi**, *Finiteness of polyhedral decompositions of cusped hyperbolic manifolds obtained by the Epstein–Penner's method*, Proc. Amer. Math. Soc. 129 (2001) 2431–2439   MR1823928

[6] **H Akiyoshi**, **M Sakuma**, **M Wada**, **Y Yamashita**, *Punctured torus groups and 2–bridge knot groups, I*, Lecture Notes in Mathematics 1909, Springer, Berlin (2007) MR2330319

[7] **D Bachman**, *Normalizing Heegaard–Scharlemann–Thompson splittings* `arXiv: math/0309049`

[8]   **R Benedetti**, **C Petronio**, *Lectures on hyperbolic geometry*, Universitext, Springer, Berlin (1992)   MR1219310

[9]   **S A Bleiler**, **C D Hodgson**, *Spherical space forms and Dehn filling*, Topology 35 (1996) 809–833   MR1396779

[10] **S A Bleiler**, **Y Moriah**, *Heegaard splittings and branched coverings of* B^3, Math. Ann. 281 (1988) 531–543   MR958258

[11] **J F Brock**, **K W Bromberg**, *On the density of geometrically finite Kleinian groups*, Acta Math. 192 (2004) 33–93   MR2079598

[12] **B A Burton**, *Regina: normal surface and* 3*–manifold topology software* Available at `http://regina.sourceforge.net/`

[13] **D Cooper**, **M Lackenby**, **J S Purcell**, *The length of unknotting tunnels*, Algebr. Geom. Topol. 10 (2010) 637–661   MR2606795

[14] **M Culler**, **N M Dunfield**, **J R Weeks**, *SnapPy, a computer program for studying the geometry and topology of* 3*–manifolds* Available at `http://snappy.computop.org`

[15] **M Culler**, **P B Shalen**, *Margulis numbers for Haken manifolds*, Israel J. Math. 190 (2012) 445–475   MR2956250

[16] **N M Dunfield**, **D P Thurston**, *A random tunnel number one 3-manifold does not fiber over the circle*, Geom. Topol. 10 (2006) 2431–2499   MR2284062

[17] **W Fenchel**, *Elementary geometry in hyperbolic space*, de Gruyter Studies in Mathematics 11, Walter de Gruyter & Co., Berlin (1989)   MR1004006

[18] **D Futer**, *Involutions of knots that fix unknotting tunnels*, J. Knot Theory Ramifications 16 (2007) 741–748   MR2341313

[19] **D Futer**, **E Kalfagianni**, **J S Purcell**, *Dehn filling, volume, and the Jones polynomial*, J. Differential Geom. 78 (2008) 429–464   MR2396249

[20] **D Futer**, **J S Purcell**, *Explicit Dehn filling and Heegaard splittings*, Comm. Anal. Geom. 21 (2013) 625–650

[21] **F Guéritaud**, *Géométrie hyperbolique effective et triangulations idéales canoniques en dimension trois*, PhD thesis, Université de Paris–XI (Orsay) (2006)

[22] **F Guéritaud**, *On canonical triangulations of once-punctured torus bundles and two-bridge link complements*, Geom. Topol. 10 (2006) 1239–1284   MR2255497 With an appendix by David Futer

[23] **F Guéritaud**, **S Schleimer**, *Canonical triangulations of Dehn fillings*, Geom. Topol. 14 (2010) 193–242   MR2578304

[24] **R Guo**, **F Luo**, *Rigidity of polyhedral surfaces, II*, Geom. Topol. 13 (2009) 1265–1312   MR2496046

[25] **J Hass**, *Genus two Heegaard splittings*, Proc. Amer. Math. Soc. 114 (1992) 565–570   MR1070519

[26] **D J Heath**, **H-J Song**, *Unknotting tunnels for* $P(-2, 3, 7)$, J. Knot Theory Ramifications 14 (2005) 1077–1085   MR2196648

[27] **C D Hodgson**, **S P Kerckhoff**, *Universal bounds for hyperbolic Dehn surgery*, Ann. of Math. 162 (2005) 367–421   MR2178964

[28] **M Lackenby**, *Word hyperbolic Dehn surgery*, Invent. Math. 140 (2000) 243–282   MR1756996

[29] **T Li**, *Heegaard surfaces and measured laminations, I: The Waldhausen conjecture*, Invent. Math. 167 (2007) 135–177   MR2264807

[30] **M Lustig**, **Y Moriah**, *Are large distance Heegaard splittings generic?* `arXiv: 1002.4292`

[31] **A D Magid**, *Deformation spaces of Kleinian surface groups are not locally connected*, Geom. Topol. 16 (2012) 1247–1320   MR2967052

[32] **J Maher**, *Random Heegaard splittings*, J. Topol. 3 (2010) 997–1025   MR2746344

[33] **R Meyerhoff**, *A lower bound for the volume of hyperbolic* 3–*manifolds*, Canad. J. Math. 39 (1987) 1038–1056   MR918586

[34] **Y Moriah**, **H Rubinstein**, *Heegaard structures of negatively curved* 3-*manifolds*, Comm. Anal. Geom. 5 (1997) 375–412   MR1487722

[35] **H Moser**, *Proving a manifold to be hyperbolic once it has been approximated to be so*, Algebr. Geom. Topol. 9 (2009) 103–133   MR2471132

[36] **W D Neumann**, **A W Reid**, *Arithmetic of hyperbolic manifolds*, from: "Topology '90", (B Apanasov, W D Neumann, A W Reid, L Siebenmann, editors), Ohio State Univ. Math. Res. Inst. Publ. 1, de Gruyter, Berlin (1992) 273–310   MR1184416

[37] **Y Rieck**, **E Sedgwick**, *Persistence of Heegaard structures under Dehn filling*, Topology Appl. 109 (2001) 41–53   MR1804562

[38] **J H Rubinstein**, *Polyhedral minimal surfaces, Heegaard splittings and decision problems for* 3–*dimensional manifolds*, from: "Geometric topology", (W H Kazez, editor), AMS/IP Stud. Adv. Math. 2, Amer. Math. Soc. (1997) 1–20   MR1470718

[39] **M Sakuma**, **J Weeks**, *Examples of canonical decompositions of hyperbolic link complements*, Japan. J. Math. 21 (1995) 393–439   MR1364387

[40] **M Scharlemann**, *Tunnel number one knots satisfy the Poenaru conjecture*, Topology Appl. 18 (1984) 235–258   MR769294

[41] **M Scharlemann**, **M Tomova**, *Alternate Heegaard genus bounds distance*, Geom. Topol. 10 (2006) 593–617   MR2224466

[42] **S Schleimer**, *Sphere recognition lies in NP*, from: "Low-dimensional and symplectic topology", (M Usher, editor), Proc. Sympos. Pure Math. 82, Amer. Math. Soc. (2011) 183–213

[43] **P B Shalen**, *A generic Margulis number for hyperbolic* 3*–manifolds*, from: "Topology and geometry in dimension three", (W Li, L Bartolini, J Johnson, F Luo, R Myers, J H Rubinstein, editors), Contemp. Math. 560, Amer. Math. Soc. (2011) 103–109 MR2866926

[44] **M Stocking**, *Almost normal surfaces in* 3*–manifolds*, Trans. Amer. Math. Soc. 352 (2000) 171–207   MR1491877

[45] **W P Thurston**, *Three-dimensional geometry and topology, Vol. 1*, Princeton Mathematical Series 35, Princeton Univ. Press (1997)   MR1435975

[46] **J L Tollefson**, *Involutions of sufficiently large* 3*–manifolds*, Topology 20 (1981) 323–352   MR617370

[47] **F Waldhausen**, *On irreducible* 3*–manifolds which are sufficiently large*, Ann. of Math. 87 (1968) 56–88   MR0224099

Department of Mathematics, University of California, Santa Barbara
Santa Barbara, CA 93106, USA

Department of Mathematics, Temple University
Philadelphia, PA 19147, USA

Department of Mathematics, Brigham Young University
Provo, UT 84602-6539, USA

cooper@math.ucsb.edu, dfuter@temple.edu, jpurcell@math.byu.edu

http://www.math.ucsb.edu/~cooper/, http://math.temple.edu/~dfuter,
http://www.math.byu.edu/~jpurcell/