The tree of knot tunnels

SANGBUM CHO

DARRYL MCCULLOUGH

We present a new theory which describes the collection of all tunnels of tunnel number 1 knots in S^3 (up to orientation-preserving equivalence in the sense of Heegaard splittings) using the disk complex of the genus–2 handlebody and associated structures. It shows that each knot tunnel is obtained from the tunnel of the trivial knot by a uniquely determined sequence of simple cabling constructions. A cabling construction is determined by a single rational parameter, so there is a corresponding numerical parameterization of all tunnels by sequences of such parameters and some additional data. Up to superficial differences in definition, the final parameter of this sequence is the Scharlemann–Thompson invariant of the tunnel, and the other parameters are the Scharlemann–Thompson invariants of the intermediate tunnels produced by the constructions. We calculate the parameter sequences for tunnels of 2–bridge knots. The theory extends easily to links, and to allow equivalence of tunnels by homeomorphisms that may be orientation-reversing.

Introduction

In this work we present a new descriptive theory of the tunnels of tunnel number 1 knots in S^3, or equivalently, of genus–2 Heegaard splittings of tunnel number 1 knot exteriors. At its heart is a bijective correspondence between the set of equivalence classes of all tunnels of all tunnel number 1 knots and a subset of the vertices of a certain tree $\mathcal{T}$. In fact, $\mathcal{T}$ is bipartite, and the tunnel vertex subset is exactly one of its two classes of vertices. The construction of $\mathcal{T}$ uses the disk complex of the genus 2 handlebody, and $\mathcal{T}$ is a quotient of a spine of the subcomplex of nonseparating disks. The tree and its associated objects have a rich combinatorial structure. The work in this paper is a first step toward understanding how that structure is manifested in the topology of tunnel number 1 knots.

The theory has many consequences. It shows that every tunnel can be obtained by starting from the unique tunnel of the trivial knot and performing a uniquely determined sequence of simple constructions. Each construction is determined by a rational

parameter, which is essentially a Scharlemann–Thompson invariant [26]. This gives a natural numerical parameterization of all tunnels. We have computed this sequence of rational invariants for all tunnels of 2–bridge knots and torus knots.

The theory adapts easily to include tunnels of tunnel number 1 links, and to allow orientation-reversing equivalence.

The next two sections provide overviews before beginning the actual development of the theory. Section 1 emphasizes the context of the work, while Section 2 summarizes the main ideas and results of the paper and the contents of the individual sections.

1 Context of the work

There are several equivalent definitions of a tunnel for a knot K in S^3. One is that a tunnel is a 1–handle attached to a regular neighborhood of K in S^3 to produce a genus–2 handlebody that is unknotted, that is, one of the handlebodies of a genus–2 Heegaard splitting of S^3. Two such configurations are equivalent tunnels when the handlebodies are isotopic taking the copy of K in one handlebody to the corresponding copy in the other. Alternatively, one may think of a tunnel as an arc α meeting K only in its endpoints, such that a regular neighborhood of the "θ–curve" $K \cup \alpha$ is unknotted. The equivalence on such arcs must then allow not only isotopy but also "sliding," where the two endpoints can meet and pass through each other.

There is a stronger notion of equivalence, in which the isotopies and sliding must preserve the knot at all times. All of our work uses the weaker notion.

If one removes a small tubular neighborhood of K from the genus–2 handlebody produced by a tunnel, the remaining compression body and its complementary genus–2 handlebody form a genus–2 Heegaard splitting of the knot exterior, in the sense of Heegaard splittings of 3–manifolds with boundary. Indeed, the equivalence classes of knot tunnels correspond exactly to the equivalence classes up to homeomorphism of genus–2 Heegaard splittings of knot exteriors. For closed 3–manifolds, genus–0 Heegaard splittings are trivial, and genus–1 splittings are very restrictive, forming only lens spaces (including S^3 and $S^2 \times S^1$), while genus–2 splittings are already a very complicated class. In this context, the Heegaard splittings coming from tunnel number 1 knots might be considered to be a special class of "genus–$1\frac{1}{2}$" splittings, an intermediate case where one might hope to find structure restricted enough to be tractable, but rich enough to be of mathematical interest.

The historical development of the subject is consistent with this hope. An impressive amount of geometric theory of tunnel number 1 knots has been developed by a number of researchers. Recently, a general picture began to emerge, through work of M Scharlemann and A Thompson [26] which defines a rational invariant that detects a kind of cabled structure of K near the tunnel. One of the applications of our work is a complete clarification of how their invariant works and what information it is detecting. As mentioned above, it extends to a sequence of rational invariants which determine a unique sequence of simple constructions that produce the tunnel. These rational invariants, plus a bit more information, give a natural numerical parameterization of all knot tunnels.

Other recent work in the subject has begun to utilize connections between tunnel number 1 knots and the curve complex of the genus–2 surface. The curve complex provides an important measure of complexity of Heegaard splittings, called the *(Hempel) distance*. Applying it to the splittings that correspond to knot tunnels, J Johnson and A Thompson [15; 16] and Y Minsky, Y Moriah and S Schleimer [21] have obtained results on bridge number and other aspects of tunnels. In [4], we use our theory to define a new distance-type invariant for knot tunnels, the "depth", that is finer than distance (depth can be large even when distance is small). The depth invariant figures prominently in our exploration of bridge numbers of tunnel number 1 knots in [5]. In particular, it allows substantial improvements of known estimates of the growth rate of bridge number as distance increases.

There are additional reasons to believe that the class of tunnel number 1 knots is a nexus of interesting mathematical objects. The fundamental group of the genus–2 handlebody is the free group on two generators F_2, and because of this, the mapping class group of the genus–2 handlebody is related to the automorphism group $\mathrm{Out}(F_2)$ and thereby to the linear groups SL_2, which are very special and whose theory differs in many respects from SL_n with $n \geq 3$. It will be evident that the tree $\mathcal{T}$ that is the central object in our theory carries much of the structure of the well-known tree associated to $\mathrm{PSL}_2(\mathbb{Z})$.

2 Summary of the results

In this section, we will give an overview of the theory and applications developed in this paper.

Let H be an unknotted genus–2 handlebody in S^3. By $\mathcal{D}(H)$ we denote the 2–dimensional complex whose vertices are the isotopy classes of *nonseparating* properly-imbedded disks in H. A set of vertices spans a simplex of $\mathcal{D}(H)$ when they have

pairwise disjoint representatives. In the first barycentric subdivision $\mathcal{D}'(H)$ of $\mathcal{D}(H)$, the span of the vertices that are not vertices of $\mathcal{D}(H)$ is a tree $\widetilde{\mathcal{T}}$. Each vertex of $\widetilde{\mathcal{T}}$ is either a triple of (isotopy classes of) disks in H, or a pair. Figure 1 below shows a small portion of $\mathcal{D}(H)$ and $\widetilde{\mathcal{T}}$.

Each tunnel of a tunnel number 1 knot determines a collection of disks in H as follows. The tunnel is a 1–handle attached to a regular neighborhood of the knot to form an unknotted genus–2 handlebody. An isotopy carrying this handlebody to H carries a cocore 2–disk of that 1–handle to a nonseparating disk in H. The indeterminacy of this process is the group of isotopy classes of orientation-preserving homeomorphisms of S^3 that preserve H. This is the *Goeritz group* $\mathcal{G}$, which we discuss in Section 5 below. In particular, we will recall work of M Scharlemann [24] and E Akbas [2] that shows that $\mathcal{G}$ is finitely presented, and even provides a simple presentation of it.

Since $\mathcal{G}$ is exactly the indeterminacy of the cocore disk, moved to H, it is essential to understand the action of $\mathcal{G}$ on $\mathcal{D}(H)$. This action is closely related to a central concept of our viewpoint, called primitivity. A disk τ in H is *primitive* if there exists a properly imbedded disk τ' in the complementary handlebody $\overline{S^3 - H}$ such that the circles $\partial\tau$ and $\partial\tau'$ in ∂H intersect transversely in a single point. Viewed as tunnels, primitive disks are exactly the tunnels of the trivial knot, and are all equivalent. The portions of $\mathcal{D}(H)$ and $\widetilde{\mathcal{T}}$ corresponding to primitive disks form the "primitive region," in particular the pairs and triples of primitive disks span a $\mathcal{G}$–invariant subtree $\widetilde{\mathcal{T}}_0$ of $\widetilde{\mathcal{T}}$ called the *primitive subtree*. The primitive subtree $\widetilde{\mathcal{T}}_0$ is isomorphic to the tree used by Scharlemann and Akbas to understand the Goeritz group. As explained in Section 5 below, the 2–complex used by Scharlemann imbeds in a very natural way into $\mathcal{D}(H)$, in such a way that it deformation retracts to $\widetilde{\mathcal{T}}_0$ (see Figure 2). Indeed, it is fair to say that $\widetilde{\mathcal{T}}_0$ *is* the Scharlemann–Akbas tree. This viewpoint leads to a new proof of the main results of [24] and [2] that avoids many of the difficult geometric arguments of those important papers. This recasting of their work is carried out by the first author [3].

Using this viewpoint, it is not difficult to analyze the action of $\mathcal{G}$ on $\mathcal{D}(H)$ and $\widetilde{\mathcal{T}}$. The action of $\mathcal{G}$ on the primitive structures is as transitive as possible, while the action of $\mathcal{G}$ on the nonprimitive structures has stabilizers that are as small as possible—usually only the order 2 subgroup generated by the "hyperelliptic" element which acts trivially on all of $\mathcal{D}(H)$. This gives a picture of $\mathcal{D}(H)/\mathcal{G}$ as a tiny primitive region, together with additional portions which look exactly as they did in $\mathcal{D}(H)$. Figure 4 below shows the structure of $\mathcal{D}(H)/\mathcal{G}$ and $\widetilde{\mathcal{T}}/\mathcal{G}$ near the primitive region. The quotients $\mathcal{D}(H)/\mathcal{G}$ and $\widetilde{\mathcal{T}}/\mathcal{G}$ are described precisely in Section 10, after preliminary work which provides a useful framework for that discussion and is also used throughout the rest of the paper.

The tree $\mathcal{T}$ discussed in the introduction is defined to be $\widetilde{\mathcal{T}}/\mathcal{G}$. The vertices of $\mathcal{D}(H)/\mathcal{G}$ are not in $\mathcal{T}$, but their links in the barycentric subdivision $\mathcal{D}'(H)/\mathcal{G}$ are subcomplexes of $\mathcal{T}$. These links are infinite trees. In $\mathcal{T}$, there is a vertex θ_0 which is the unique $\mathcal{G}$–orbit of a triple of primitive disks in H (or dually it represents a planar θ–curve in S^3). For each vertex τ of $\mathcal{D}'(H)/\mathcal{G}$ that is the image of a vertex of $\mathcal{D}(H)$, ie for each tunnel, there is a unique shortest path in $\mathcal{T}$ from θ_0 to *the vertex in the link of τ that is closest to θ_0*. This closest vertex is a triple, called the *principal vertex* of τ, and the path is the *principal path* of τ. Figure 8 below illustrates these structures.

Those familiar with [26] will not be surprised to learn that the disks in the principal vertex, other than τ itself, are the disks μ^+ and μ^- used in the definition of the Scharlemann–Thompson invariant. That is, the principal vertex of τ is $\{\mu^+, \mu^-, \tau\}$ (Lemma 14.1).

The principal path and principal vertex of τ, along with the surrounding structure of $\mathcal{D}(H)/\mathcal{G}$, encode a great deal of geometric information about the tunnel τ and the knot K_τ of which it is a tunnel. In Section 13 we examine how the sequence of vertices in the principal path corresponds to a sequence of constructions of a very simple type, which will look familiar to experts. In a sentence, the construction is "Think of the union of K and the tunnel arc as a θ–curve, and cable the ends of the tunnel arc and one of the arcs of K in a neighborhood of the other arc of K." Each of these "cabling constructions" (often just called "cablings") is determined by a rational "slope" parameter. For the first cabling of the sequence, the indeterminacy coming from the Goeritz group necessitates some special treatment; in particular, its parameter is a "simple" slope taking values in $\mathbb{Q}/\mathbb{Z}$. Figures 9 and 10 below should give a fairly good idea of how cabling constructions change a pair consisting of a knot and tunnel to a more complicated pair.

More precisely, the vertices in the principal path of τ are a sequence of alternating triples and pairs, which we write as $\theta_0, \mu_0, \mu_0 \cup \{\tau_0\}, \mu_1, \ldots, \mu_n, \mu_n \cup \{\tau_n\}$, where $\tau_n = \tau$. Denoting by $(\mu; \sigma)$ the edge of $\mathcal{T}$ that goes from a pair μ to an adjacent triple $\mu \cup \sigma$, the exact information needed to describe a cabling construction is a succession from $(\mu_{i-1}; \tau_{i-1})$ to $(\mu_i; \tau_i)$ in the principal path. The first edge $(\mu_0; \tau_0)$ is special, and by itself determines a "simple" cabling. The uniqueness of the principal path gives one of our main results:

Unique Cabling Sequence Theorem 13.2 *Let τ be a tunnel of a nontrivial knot. Let $\theta_0, \mu_0, \mu_0 \cup \{\tau_0\}, \mu_1, \ldots, \mu_n, \mu_n \cup \{\tau_n\}$ with $\tau_n = \tau$ be the principal path of τ. Then the sequence of $n+1$ cablings consisting of the simple cabling determined by $(\mu_0; \tau_0)$ and the cablings determined by the successions from $(\mu_{i-1}; \tau_{i-1})$ to $(\mu_i; \tau_i)$ is the unique sequence of cablings beginning with the tunnel of the trivial knot and ending with τ.*

As detailed in Sections 7 and 8, an edge $(\mu; \sigma)$ defines a coordinate system in which each essential disk in H disjoint from the disks of μ and not parallel to either of them or to σ is assigned a rational slope. In particular, the $(\mu_i; \sigma_i)$–slopes of the τ_i, where σ_i is the unique disk in $\mu_{i-1} - \mu_i$, together with a special $\mathbb{Q}/\mathbb{Z}$–valued slope associated to the initial $(\mu_0; \tau_0)$, determine the exact cabling constructions in the sequence. This gives us a version of Unique Cabling Sequence Theorem 13.2 that describes the unique cabling sequences as a parameterization of all tunnels:

Parameterization Theorem 12.3 *Let τ be a knot tunnel with principal path θ_0, μ_0, $\mu_0 \cup \{\tau_0\}$, $\mu_1, \ldots, \mu_n$, $\mu_n \cup \{\tau_n\}$. Fix a lift of the principal path to $\mathcal{D}(H)$, so that each μ_i corresponds to an actual pair of disks in H.*

(1) *If τ is primitive, put $m_0 = [0] \in \mathbb{Q}/\mathbb{Z}$. Otherwise, let $m_0 = [p_0/q_0] \in \mathbb{Q}/\mathbb{Z}$ be the simple slope of τ_0.*

(2) *If $n \geq 1$, then for $1 \leq i \leq n$ let σ_i be the unique disk in $\mu_{i-1} - \mu_i$ and let $m_i = q_i/p_i \in \mathbb{Q}$ be the $(\mu_i; \sigma_i)$–slope of τ_i.*

(3) *If $n \geq 2$, then for $2 \leq i \leq n$ define $s_i = 0$ or $s_i = 1$ according to whether or not the unique disk of $\mu_i \cap \mu_{i-1}$ equals the unique disk of $\mu_{i-1} \cap \mu_{i-2}$.*

Then, sending τ to the pair $((m_0, \ldots, m_n), (s_2, \ldots, s_n))$ is a bijection from the set of all tunnels of all tunnel number 1 knots to the set of all elements

$$(([p_0/q_0], q_1/p_1, \ldots, q_n/p_n), (s_2, \ldots, s_n))$$

in $\qquad (\mathbb{Q}/\mathbb{Z}) \cup (\mathbb{Q}/\mathbb{Z} \times \mathbb{Q}) \cup \left(\bigcup_{n \geq 2} \mathbb{Q}/\mathbb{Z} \times \mathbb{Q}^n \times C_2^{n-1} \right)$

with all q_i odd.

This is actually proven earlier than Unique Cabling Sequence Theorem 13.2, since it does not require any interpretation of the principal path in terms of cabling constructions.

Section 14 explains how the Scharlemann–Thompson invariant is really the slope parameter for the final cabling construction—the m_n in the Parameterization Theorem 12.3. This rational number is called the *principal slope* of τ. Due to differing definitions, the invariants generally have different values, but they capture exactly the same geometric information. Intuitively, the Scharlemann–Thompson invariant is "the slope of the disk that the tunnel disk replaced as seen from the tunnel disk" while the principal slope is "the slope of the tunnel disk as seen from the disk it replaced," so it is not surprising that they are related by a continued fraction algorithm, which is described in Proposition 14.4 and has been computationally implemented [7]. In particular, when one of them is an integer, the other is also an integer, in fact, the negative of the first one.

A tunnel produced from the tunnel of the trivial knot by a single cabling construction is called a *simple* tunnel. These are exactly the "upper and lower" tunnels of 2–bridge knots. According to the Parameterization Theorem 12.3, these are determined by a single $\mathbb{Q}/\mathbb{Z}$–valued parameter m_0, and this is of course a version of the standard rational parameter associated to the 2–bridge knot. Simple tunnels are examined in Section 11.

A nonsimple tunnel produced by a cabling sequence in which one of the original arcs of the trivial knot is retained is called a "semisimple" tunnel. In terms of the parameterization, the simple and semisimple tunnels are exactly the tunnels with all $s_i = 0$. Geometrically, the simple and semisimple tunnels are the "(1,1)" tunnels of $(1, 1)$–knots (ie knots which can be put in 1–bridge position with respect to the levels of a product neighborhood of an unknotted torus in S^3). As arcs, they are sometimes called "eyeglass" tunnels, meaning that the tunnel arc can be slid to an unknotted circle. The "semisimple region" in $\mathcal{D}(H)/\mathcal{G}$ appears to be where the more complicated phenomena involving tunnels occur. Indeed, we do not know an example of a knot with inequivalent tunnels which is not a $(1, 1)$–knot.

The tunnels of 2–bridge knots are examined in Section 15. It is known from work of several mathematicians (see Kobayashi [17; 18], Morimoto and Sakuma [22] and Uchida [27]) that a 2–bridge knot has at most four equivalence classes of tunnels (not six, for us, since we are considering tunnels only up to equivalence, rather than up to isotopy). Two of these are the upper and lower simple tunnels. The others are semisimple tunnels, and in Section 15 we determine their exact cabling sequences. Indeed, we have made software available [7] that computes them quite effectively.

In [8], we have computed the cabling sequences for all tunnels of torus knots. In other ongoing work [6], we have computed the cabling sequences for all $(1, 1)$–tunnels, in terms of a description of its $(1, 1)$–knot as a braid of two points in the torus. A complete translation between the braid word descriptions and the cabling slopes is obtained and has been implemented computationally.

The minimal length of a simplicial path in the 1–skeleton of $\mathcal{D}(H)/\mathcal{G}$ from a tunnel τ to the orbit of the primitive disks is called the *depth* of τ. The simple and semisimple tunnels are exactly the tunnels of depth 1. The depth invariant is used in our paper [5]. In particular, we show there that as a rather immediate consequence of work of Goda, Scharlemann and Thompson [11] and Scharlemann and Thompson [26], the bridge number grows exponentially with the depth. More precisely, we have the following from [5]:

Minimum Bridge Number Theorem *For $d \geq 1$, the minimum bridge number of a knot having a tunnel of depth d is given recursively by a_d, where $a_1 = 2$, $a_2 = 4$ and $a_d = 2a_{d-1} + a_{d-2}$ for $d \geq 3$. Explicitly,*

$$a_d = \frac{(1+\sqrt{2})^d}{\sqrt{2}} - \frac{(1-\sqrt{2})^d}{\sqrt{2}} ,$$

and consequently

$$\lim_{d \to \infty} a_d - \frac{(1+\sqrt{2})^d}{\sqrt{2}} = 0.$$

The lower bound in the Minimum Bridge Number Theorem improves Lemma 2 of [15], which is that bridge number grows linearly with Hempel distance. It also improves Proposition 1.11 of [11], which shows that bridge number grows asymptotically at least as fast as 2^d.

In addition, [5] contains a strengthening of the Tunnel Leveling Theorem of [11], which we call the Tunnel Leveling Addendum. It implies that for a knot having a tunnel with a given sequence of invariants $(s_2, s_3, \ldots, s_n)$ in the Parameterization Theorem 12.3, the bridge number lies in an easily computable set containing $2m - 2$ elements, where s_m is the first s_i that is equal to 1 (and $m = n + 1$ if all $s_i = 0$).

Our entire theory extends quite easily to links, as we discuss in Section 16. The cabling constructions are expanded to allow cablings that produce links, which are terminal in the sense that they do not allow further cabling constructions to be performed. The Parameterization Theorem 12.3 holds as stated, except allowing the final q_n to be even. One application of the link version of the theory is a very quick proof of the fact that the only tunnels of a 2–bridge link are its upper and lower tunnels [1; 19]. We also show that a tunnel number 1 link with an unknotted component must have torus bridge number 2 (Theorem 16.4).

Our entire theory also adapts easily to allow tunnel equivalences which may be orientation-reversing homeomorphisms of S^3. In the Parameterization Theorem 12.3, the cabling sequence of the mirror image of a tunnel τ has the same parameters as τ except that the slopes $(m_0, \ldots, m_n)$ become $(-m_0, \ldots, -m_n)$. This shows (Theorem 16.2) that apart from the tunnels of the trivial knot and trivial link, the only tunnel that is equivalent to itself under an orientation-reversing homeomorphism of S^3 is the tunnel of the Hopf link. In fact, the Hopf link and its tunnel have some unusual symmetries, which are analyzed in Section 17.

3 The disk complex of an irreducible 3–manifold

Let M be a compact, irreducible 3–manifold. The *disk complex* $\mathcal{K}(M)$ is the simplicial complex whose vertices are the isotopy classes of essential disks in M, such that a collection of $k+1$ vertices spans a k–simplex if and only if they admit a set of pairwise-disjoint representatives.

This is a good point at which to mention that to avoid endless repetition of "isotopy class", we often speak of "disks" and other objects when we really mean their isotopy classes, with the implicit understanding that we always choose representatives of the isotopy classes that are the simplest possible with respect to whatever we are doing (ie transversely intersecting in the minimum possible number of components, and so on). A "unique" disk means a unique isotopy class of disks. We often omit other formalism that should be obvious from context, so imbedded submanifolds are assumed to be essentially and properly imbedded, unless otherwise stated, and isotopies are assumed to preserve relevant structure. To initiate this massive abuse of language, we say that a collection of $k+1$ disks in M spans a k–simplex of $\mathcal{K}(M)$ if and only if they are pairwise disjoint.

The following was proven in [20, Theorem 5.3]:

Theorem 3.1 *If ∂M is compressible, then $\mathcal{K}(M)$ is contractible.*

Denote by $\mathcal{D}(M)$ the subcomplex of $\mathcal{K}(M)$ spanned by the *nonseparating* disks. From [20, Theorem 5.4], we have:

Theorem 3.2 *If M has a nonseparating compressing disk, then $\mathcal{D}(M)$ is contractible.*

The basic idea of these theorems is that, fixing a base disk D_0, one can start at any disk D and move steadily "closer" to D_0, in an appropriate sense, by repeatedly surgering D along an intersection with D_0 that lies outermost on D_0. Doing this with a certain amount of care allows one to produce a null-homotopy for any simplicial map of S^k into the complex. The arguments given in [20] are basically correct, but contain some minor misstatements. A much improved treatment is given in [3].

4 The disk complex and its spine

Fix a standard unknotted genus–2 handlebody $H \subset S^3$. From now on, disks in H are assumed to be *nonseparating* unless explicitly stated otherwise. Since at most three isotopy classes in H may be represented by disjoint disks, $\mathcal{D}(H)$ has dimension 2.
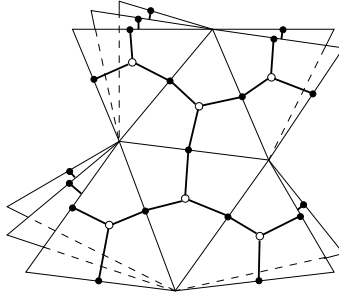
Figure 1: A portion of the nonseparating disk complex $\mathcal{D}(H)$ and the tree $\widetilde{\mathcal{T}}$. Countably many 2–simplices meet along each edge.

A portion of $\mathcal{D}(H)$ is shown in Figure 1. An edge of $\mathcal{D}(H)$ is a pair of disjoint nonisotopic disks in H and is called a *meridian pair*, or just a *pair*. Similarly, a 2–simplex of $\mathcal{D}(H)$ is a *triple*. A triple corresponds to a θ–*curve* in H; that is, a union of three arcs meeting only in their common endpoints that is a deformation retract of H. It is determined by the condition that each arc passes through exactly one of the disks of the triple in one point. That arc is called the arc *dual* to the disk it meets, and the disk is called the disk *dual* to the arc. We also say that the θ–curve is dual to the triple.

Also shown in Figure 1 is a tree $\widetilde{\mathcal{T}}$ which is a deformation retract of $\mathcal{D}(H)$. It is constructed as follows. Let $\mathcal{D}'(H)$ be the first barycentric subdivision of $\mathcal{D}(H)$. Denote by $\widetilde{\mathcal{T}}$ the subcomplex of $\mathcal{D}'(H)$ obtained by removing the open stars of the vertices of $\mathcal{D}(H)$. It is a bipartite graph, with "white" vertices of valence 3 represented by triples and "black" vertices of (countably) infinite valence represented by pairs. The valences reflect the fact that moving along an edge from a triple to a pair corresponds to removing one of its three disks, while moving from a pair to a triple corresponds to adding one of infinitely many possible third disks to a pair. The possible disjoint third disks that can be added are called the "slope disks" for the pair.

The link of a vertex of $\mathcal{D}(H)$ in $\mathcal{D}'(H)$ is an infinite graph contained in $\widetilde{\mathcal{T}}$, and structurally very similar to $\widetilde{\mathcal{T}}$, except that its white vertices have valence 2 rather than valence 3.

It was proven in [20, Theorem 5.5] and in [3] that the link in $\mathcal{D}'(H)$ of each vertex of $\mathcal{D}(H)$ is contractible. Combined with Theorem 3.2, this implies that:

Theorem 4.1 $\widetilde{\mathcal{T}}$ *is a tree.*

It is well known that complexes such as the disk complex admit actions of mapping class groups, indeed this is one of the important motivations for studying them. The

mapping class group $\pi_0(\text{Diff}_+(H))$ acts on $\mathcal{D}(H)$, simply by $[f] \cdot \langle [D_0], \ldots, [D_k] \rangle = \langle [f(D_0)], \ldots, [f(D_k)] \rangle$. It is rather obvious that the quotient of $\widetilde{\mathcal{T}}$ is a single edge, so as seen in [20], the Bass–Serre theory of group actions on trees shows that $\pi_0(\text{Diff}_+(H))$ is a free product with amalgamation. In particular, it is finitely generated, and an explicit presentation can be worked out by examination of the vertex and edge stabilizers.

5   The Goeritz groups and the Scharlemann–Akbas tree

The *Goeritz group* $\mathcal{G}$ is the group of isotopy classes of orientation-preserving diffeomorphisms of S^3 that take H onto itself. Since every orientation-preserving diffeomorphism of S^3 is isotopic to the identity, $\mathcal{G}$ is exactly the indeterminacy when one takes an arbitrary unknotted handlebody in S^3 and moves it to H by an isotopy.

At times, mostly in Section 16, we will use the *extended Goeritz group* $\mathcal{G}_\pm$, in which the diffeomorphisms of S^3 are allowed to reverse orientation.

Some very important recent work of M Scharlemann and E Akbas provides a precise description of $\mathcal{G}$. Their methodology is also of interest, and as we will see it can be considerably simplified by using $\mathcal{D}(H)$.

L Goeritz [12] gave generators for $\mathcal{G}$, and M Scharlemann [24] provided a modern proof that they do generate. That proof uses a 2–dimensional complex whose vertices are isotopy classes of *splitting spheres* for H, that is, 2–spheres in S^3 that intersect H in one essential disk (necessarily a separating disk), up to isotopy through such spheres. This complex, which we call the *Scharlemann complex*, is rather difficult to work with, since the adjacency condition for two spheres in the complex is not disjointness—this condition would not work, because vertices with disjoint representatives are equal, and the complex would be a discrete set. Instead, one must use minimal intersection, that is, the intersections of the spheres with ∂H meet in only 4 points. The Goeritz group acts on the Scharlemann complex in the usual way, with quotient a finite complex and with finitely generated stabilizers. Using highly nontrivial geometric arguments, Scharlemann proved:

Theorem 5.1 (Scharlemann)   *The Scharlemann complex is connected.*

It follows by well-known algebraic considerations that $\mathcal{G}$ is finitely generated.

By additional complicated geometric arguments, E Akbas [2] obtained the following:

Theorem 5.2 (Akbas)   *The natural 1–dimensional deformation retract of the Scharlemann complex is a tree.*

Akbas worked out the vertex and edge stabilizers of the tree, enabling him to give a pleasantly transparent presentation of $\mathcal{G}$.

It turns out that the work of Scharlemann and Akbas is intimately related to the disk complex and begins to give insight into the role of $\widetilde{\mathcal{T}}$. Recall that a disk τ in H is called *primitive* when there exists a properly imbedded disk τ' in the complementary handlebody $\overline{S^3 - H}$ such that the circles $\partial\tau$ and $\partial\tau'$ in ∂H intersect transversely in a single point. We call τ' a *dual disk* of τ.

A *primitive pair* (respectively, *primitive triple*) is a pair (respectively, triple) of disjoint (nonisotopic) primitive disks. *The splitting spheres for H correspond exactly to the primitive pairs.* For a splitting sphere cuts H into two unknotted solid tori, each containing a unique nonseparating disk which is primitive. On the other hand, any primitive pair can be shown [3, Lemma 2.2] to have a unique pair of dual disks that are disjoint from each other and each disjoint from the other disk of the primitive pair, and the splitting sphere is the boundary of a small regular neighborhood of the union of either of the disks with its dual.

The *primitive subtree* $\widetilde{\mathcal{T}}_0$ is the subcomplex of $\widetilde{\mathcal{T}}$ spanned by the vertices that are primitive pairs and primitive triples. It is routine to check that two splitting spheres represent adjacent vertices in the Scharlemann complex if and only if the corresponding primitive pairs are contained in a primitive triple. So sending a splitting sphere to its cor-
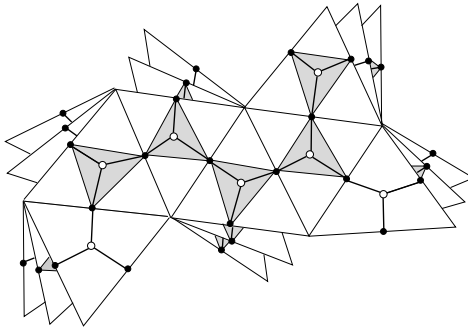

Figure 2: A portion of the nonsimplicial imbedding of the Scharlemann complex into $\mathcal{D}(H)$. The vertices of the complex map to the vertices of $\mathcal{D}'(H)$ that are primitive pairs, and each 2–simplex is spanned by the three primitive pairs contained in a primitive triple.

responding primitive pair determines a (nonsimplicial) imbedding of the Scharlemann complex into $\mathcal{D}(H)$, as shown in Figure 2, and there is an obvious deformation from the image to the primitive tree. In this way, *the Scharlemann–Akbas tree is naturally identified with the primitive subtree $\widetilde{\mathcal{T}}_0$ of $\widetilde{\mathcal{T}}$.*

Notice that our observations to this point show that Scharlemann's connectedness theorem immediately implies Akbas's tree theorem, since any connected subcomplex of a tree is a tree; also, Scharlemann's theorem verifies that the primitive subtree really is a tree. But there is now an independent proof that $\widetilde{\mathcal{T}}_0$ is a tree, based on the following key fact about primitive disks [3, Theorem 2.3]:

Theorem 5.3 *Let τ and σ be primitive disks in H which intersect transversely. Let τ_1 and τ_2 be the disks that result from surgering τ along an intersection arc which is outermost on σ. Then τ_1 and τ_2 are primitive.*

The proof of Theorem 5.3 is not trivial, but neither is it lengthy. Theorem 5.3 easily implies that $\widetilde{\mathcal{T}}_0$ is contractible, so implies the theorem of Akbas, and it is straightforward to work out the stabilizers and recover his presentation of $\mathcal{G}$.

We remark that Theorem 5.3 is not true for higher genera, so does not provide a proof that the higher-genus Goeritz groups are finitely generated.

6 Tunnels as disks

Consider a knot tunnel, regarded as a 1–handle attached to a regular neighborhood of the knot. Since S^3 has a unique Heegaard splitting of each genus, we can move the neighborhood of the knot and the 1–handle by isotopy to be the standard H. The cocore 2–disk of the 1–handle is then a nonseparating disk in H, ie a vertex of $\mathcal{D}(H)$. Allowing for the indeterminacy in this process measured by the Goeritz group, we have our definition of a knot tunnel:

Definition 6.1 A *tunnel* is a $\mathcal{G}$–orbit of disks in H. Thus the tunnels correspond exactly to the vertices of the quotient $\mathcal{D}(H)/\mathcal{G}$ of $\mathcal{D}(H)$ by the action of $\mathcal{G}$.

Notation 6.2 If τ is a tunnel, then it is a *tunnel of the knot* K_τ which is a core of the solid torus obtained by cutting H along τ. We regard K_τ as defined only up to isotopy in S^3.

Figure 3 shows tunnels for the right-handed trefoil and figure-8 knots.

One may also consider tunnels and knots up to equivalence that may be orientation-reversing, simply by replacing $\mathcal{G}$ by $\mathcal{G}_\pm$ in the definition of tunnel. This requires only very minor modifications to the theory. In our exposition, we usually consider only orientation-preserving equivalence, but will occasionally point out what happens in the other case.
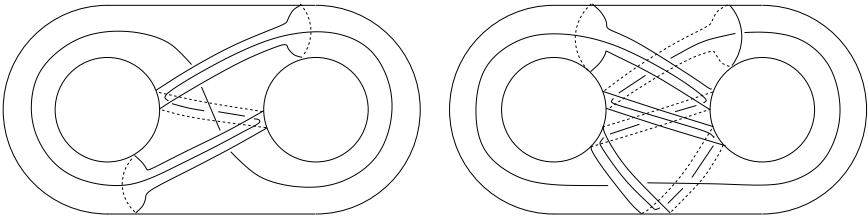
Figure 3: Knot tunnels for the right-handed trefoil and figure-8 knots.

Our definition of tunnel agrees with the standard definition of "equivalent" knot tunnels using a 1–handle attached to a regular neighborhood of a knot, and hence with any definition, but it is still worth thinking through the tunnel arc viewpoint. Given a tunnel arc attached to the knot, take a regular neighborhood of the knot and arc and move it to H by isotopy. The knot and tunnel arc form a θ–curve, and the disks dual to the arcs of the θ–curve form a triple that contains τ. Conversely, any triple containing τ determines such a θ–curve whose arcs not dual to τ form (a knot isotopic to) K_τ. Thus the isotopy classes in S^3 of arcs that determine the tunnel correspond exactly to the white vertices of the link of τ in $\mathcal{D}'(H)/\mathcal{G}$. The moves usually called "sliding" change the θ–curve and correspond to moving through the link of τ in $\mathcal{D}'(H)/\mathcal{G}$.

It is not difficult to see that there is a unique $\mathcal{G}$–orbit π_0 of primitive disks. Using [3, Lemma 2.2], one can show that there is a unique $\mathcal{G}$–orbit μ_0 of primitive pairs. Also from [3], there is a unique orbit θ_0 of primitive triples.

The quotient $\mathcal{D}'(H)/\mathcal{G}$ inherits a triangulation from $\mathcal{D}'(H)$. The quotient of $\widetilde{\mathcal{T}}$ by the action of $\mathcal{G}$ is a subcomplex of $\mathcal{D}'(H)/\mathcal{G}$. We call this quotient $\mathcal{T}$. A portion of $\mathcal{D}(H)/\mathcal{G}$ and $\mathcal{T}$ is shown in Figure 4.

Theorem 6.3 $\mathcal{T}$ *is a tree.*

Proof Since all primitive triples are equivalent, and the three pairs in a triple are also equivalent under the stabilizer of the triple, the quotient of $\widetilde{\mathcal{T}_0}$ by $\mathcal{G}$ is the single edge $\langle \mu_0, \theta_0 \rangle$. Since $\widetilde{\mathcal{T}_0}$ is connected, each point p of $\widetilde{\mathcal{T}} - \widetilde{\mathcal{T}_0}$ has a well-defined minimal distance from $\widetilde{\mathcal{T}_0}$, which is just the length of the unique shortest arc in $\widetilde{\mathcal{T}}$ from p to $\widetilde{\mathcal{T}_0}$ (as is usual, we use a path metric with the length of each edge equal to 1). This arc must end at a primitive pair, since all edges incident to a primitive triple lie in $\widetilde{\mathcal{T}_0}$. The action of $\mathcal{G}$ leaves $\widetilde{\mathcal{T}_0}$ invariant, hence preserves the shortest distance, so the image of the arc from p to $\widetilde{\mathcal{T}_0}$ is the unique shortest arc from the image of p to $\widetilde{\mathcal{T}_0}$. Thus $\mathcal{T}$ has a unique arc from each point to μ_0, so $\mathcal{T}$ is a tree.   □

In Section 9 we will analyze $\mathcal{D}(H)/\mathcal{G}$ and $\mathcal{T}$ in quite a bit more detail, after setting up some useful terminology in the next two sections.
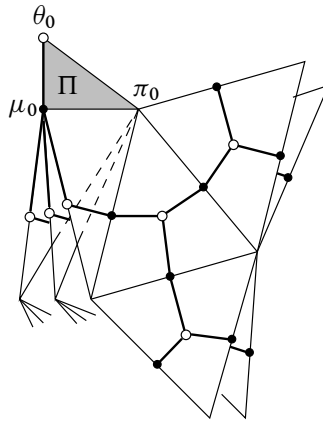
Figure 4: A portion of $\mathcal{D}(H)/\mathcal{G}$ and $\mathcal{T}$ near the primitive orbits

7 Slope disks, cables and waves

Throughout this section, we consider a pair of disks λ and ρ (for "left" and "right") in H, as shown abstractly in Figure 5. Since λ and ρ are arbitrary, the true picture in S^3 might look very different from the standard-looking pair shown here. Let B be H cut along $\lambda \cup \rho$. The frontier of B in H consists of four disks which appear vertical in Figure 5. Denote this frontier by F, and let Σ be $B \cap \partial H$, a sphere with four holes.
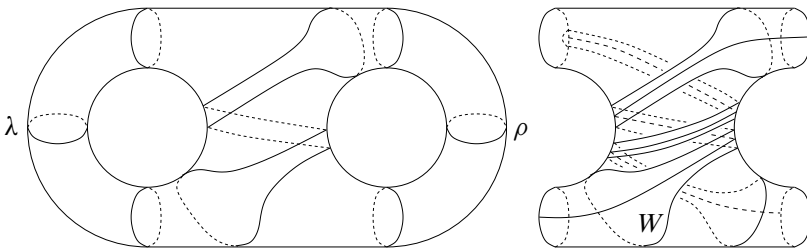

Figure 5: A slope disk and associated structures. The disk W is one of the four waves that correspond to the slope disk.

Definition 7.1 A *slope disk for* $\{\lambda, \rho\}$ is an essential disk, possibly separating, which is contained in B and not isotopic to any component of F.

The boundary of a slope disk always separates Σ into two pairs of pants, conversely any loop in Σ that is not homotopic into $\partial\Sigma$ is the boundary of a unique slope disk. If two slope disks are isotopic in H, then they are isotopic in B.

An arc in Σ whose endpoints lie in two different boundary circles of Σ is called a *cabling arc*, and a pair of disjoint cabling arcs whose four endpoints lie in the four different boundary circles of Σ is called a *cable*. Figure 5 shows a cable disjoint from a slope disk. A slope disk is disjoint from a unique cable. On the other hand, each cabling arc α determines a unique slope disk: if the endpoints of α lie in the frontier disks F_1 and F_2 of B, then the frontier of a regular neighborhood of $F_1 \cup \alpha \cup F_2$ in B is the slope disk. Finally, each cabling arc determines a unique cable, the cable disjoint from the slope disk that it determines.

A *wave* is a disk in B that meets F in a single arc and is essential in (B, F), that is, not parallel through such disks to a disk in F. A wave W determines a unique slope disk: if F_0 is the component of F that meets W, then the frontier of a regular neighborhood of $F_0 \cup W$ in B consists of two disks, one a slope disk and the other parallel to a component of F. A cabling arc α determines two waves: if F_1 and F_2 are the disks of F that contain a boundary point of α, the frontier of a regular neighborhood of each $\alpha \cup F_i$ in B is a wave which determines the same slope disk as α does. A slope disk is produced by any of four waves, the two pairs produced from a cabling that determines the slope disk. One such wave is shown in Figure 5.

In summary:

(1) A slope disk determines a cable, either of whose cabling arcs determines the slope disk and hence the other cabling arc of the cable.

(2) A slope disk determines four waves, each of which determines the slope disk.

Since disjoint slope disks in B are parallel, disjoint waves in B determine the same slope disk. If E is any disk in H, separating or not, which intersects F minimally, then each arc of $E \cap F$ that is outermost on E cuts off a subdisk of E that is a wave. We call it a *wave of E with respect to $\{\lambda, \rho\}$*.

Any disk D in H, separating or not, determines a unique slope disk of $\{\lambda, \rho\}$, provided that D itself is not one of λ or ρ. If it is not already a slope disk, then it has a wave with respect to $\{\lambda, \rho\}$, which determines a unique slope disk of $\{\lambda, \rho\}$. Since all waves of D with respect to $\{\lambda, \rho\}$ are disjoint, this slope disk is well-defined, and we call it the *slope disk of $\{\lambda, \rho\}$ determined by D*. Of course if D is already slope disk for $\{\lambda, \rho\}$, then we define the slope disk determined by D to be D itself.

8 General slope coordinates

In this section, we will explain how each choice of nonseparating slope disk for a pair $\mu = \{\lambda, \rho\}$ determines a correspondence between $\mathbb{Q} \cup \{\infty\}$ and the set of all slope

disks of μ. As we saw in Section 7, such a correspondence associates a value to each cabling arc, cable and wave as well. In fact, it will be most convenient to use cabling arcs to define the value, as the method is simply a version of the standard procedure for associating a parameter to a rational tangle.

Let F and Σ be as in the previous section. Fixing a nonseparating slope disk τ for μ, we will write $(\mu; \tau)$ for the ordered pair consisting of μ and τ.

Definition 8.1 A *perpendicular disk* for $(\mu; \tau)$ is a disk $\tau^\perp$ with the following properties:

(1) $\tau^\perp$ is a slope disk for μ.

(2) τ and $\tau^\perp$ intersect transversely in one arc.

(3) $\tau^\perp$ separates H.

There are infinitely many choices for $\tau^\perp$, but because $H \subset S^3$ there is a natural way to choose a particular one, which we call τ^0. It is illustrated in Figure 6. To construct it, start with any perpendicular disk and change it by Dehn twists of H about τ until the core circles of the complementary solid tori have linking number 0 in S^3.
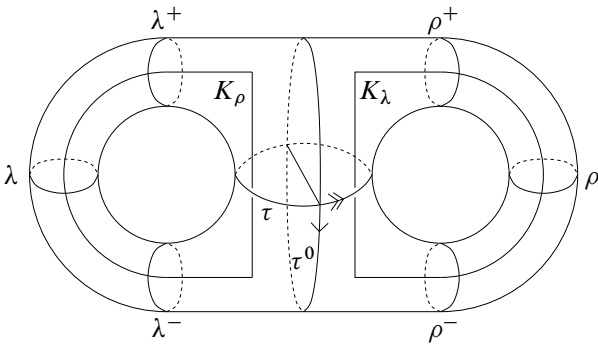

Figure 6: The slope-zero perpendicular disk τ^0. It is chosen so that K_λ and K_ρ have linking number 0.

For calculations, it is convenient to draw the picture as in Figure 6, and orient the boundaries of τ and τ^0 so that the orientation of τ^0 (the "x–axis"), followed by the orientation of τ (the "y–axis"), followed by the outward normal of H, is a right-hand orientation of S^3. At the other intersection point, these give the left-hand orientation, but we will see that the coordinates are unaffected by changing the choices of which of $\{\lambda, \rho\}$ is λ and which is ρ, or changing which sides are $+$ and which are $-$, provided that the $+$ sides both lie on the same side of $\lambda \cup \rho \cup \tau$ in Figure 6.

Let $\widetilde{\Sigma}$ be the covering space of Σ such that:

(1) $\widetilde{\Sigma}$ is the plane with an open disk of radius $1/8$ removed from each point with half-integer coordinates.

(2) The components of the preimage of τ are the vertical lines with integer $x-$ coordinate.

(3) The components of the preimage of τ^0 are the horizontal lines with integer $y-$coordinate.

Figure 7 shows a picture of Σ, and a picture of $\widetilde{\Sigma}$ showing a fundamental domain for the action of its group of covering transformations, which is the orientation-preserving subgroup of the group generated by reflections in the half-integer lattice lines (that pass through the centers of the missing disks). Each circle of $\partial\widetilde{\Sigma}$ double covers a circle of $\partial\Sigma$. The curves in Σ labeled $\lambda^\pm$, $\rho^\pm$, τ and τ^0 are the boundaries of the corresponding disks in H.
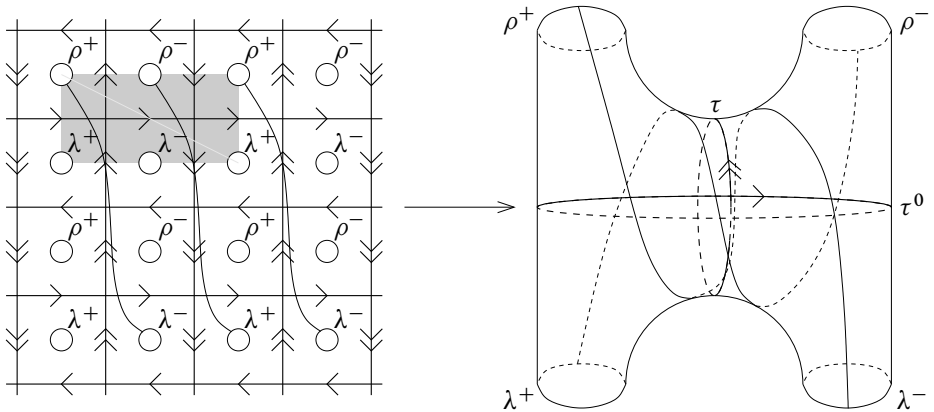

Figure 7: The covering space $\widetilde{\Sigma} \to \Sigma$ and some lifts of a $[1,-3]$–cabling arc. The shaded region is a fundamental domain.

If we lift any cabling arc in Σ to $\widetilde{\Sigma}$, the lift runs from a boundary circle of $\widetilde{\Sigma}$ to one of its translates by a vector (p,q) of signed integers, defined up to multiplication by the scalar -1. Thus each cabling arc receives a *slope pair* $[p,q] = \{(p,q),(-p,-q)\}$ and is called a $[p,q]$–*cabling arc*. Of course, each slope disk, cable, or wave receives a slope pair $[p,q]$ as well.

An important observation is that a $[p,q]$–slope disk is nonseparating in H if and only if q is odd. Both happen exactly when the corresponding cabling arc has one endpoint in λ^+ or λ^- and the other in ρ^+ or ρ^-.

Definition 8.2 The $(\mu; \tau)$–*slope* of a $[p, q]$–slope disk, cabling arc, cable, or wave is the rational number q/p.

Finally, we clarify why the choices of λ and ρ and of $+$ and $-$ sides do not affect the final slope. Interchanging λ and ρ changes the covering space in Figure 7 by a vertical translation by 1, while interchanging the $+$ and $-$ sides changes it by a horizontal translation by 1. The orientation on one of τ^0 or τ is reversed, but the set of lifts in Figure 7 is preserved, so the slope pairs are unchanged.

In fact, any disk D in H, separating or not, has a well-defined $(\mu; \tau)$–slope provided that D is not one of the two disks in μ. We just define it to be the $(\mu; \tau)$–slope of the slope disk of μ determined by D. This has the following property, which will be useful in Section 14.

Proposition 8.3 Let D and D' be disjoint disks in H, separating or not, neither of which is a disk in μ. Then D and D' have the same $(\mu; \tau)$–slope.

Proof If both are slope disks of μ, they must be parallel in B. If only one is a slope disk, then it is the unique slope disk disjoint from a wave with respect to μ of the other, again producing the same slope. If neither is a slope disk, then they have disjoint waves with respect to μ and hence the same $(\mu; \tau)$–slope. □

9 Slope disks of primitive pairs

In this section, we will examine the set of slope disks for primitive pairs. In particular, we will show that the equivalence classes modulo $\mathcal{G}$ correspond, by taking the "simple slope," to $\mathbb{Q}/\mathbb{Z} \cup \{\infty\}$.

Fix a primitive pair $\mu_0 = \{\lambda_0, \rho_0\}$. We use the notation of the previous section, but add "0" subscripts as in Σ_0 and $\widetilde{\Sigma}_0$ to remind ourselves that we are in the primitive case.

Akbas [2] gave explicit generators for the stabilizer $\mathcal{G}_{\mu_0}$ of μ_0 under the action of $\mathcal{G}$. To describe them, we refer to Figure 6 (take λ_0 to be the disk λ in Figure 6 and ρ_0 to be ρ). The "hyperelliptic" involution α rotates H by π about a horizontal axis that meets H in three arcs that are diameters of λ_0, ρ_0 and τ. The "half-twist" β is the identity on the half of H to the left of τ^0 and agrees with α on the 1–handle connecting ρ_0^+ and ρ_0^-. It is a "left-hand" twist, taking the τ in Figure 6, which has slope pair $[0, 1]$, to a slope disk with slope pair $[1, 1]$, and β^2 is a Dehn twist about τ^0. Finally, the "rotation" γ is an involution that rotates H by π about the

axis $\tau \cap \tau^0$, interchanging λ_0 and ρ_0 and preserving each of τ and τ^0 but acting as reflection across the fixed axis on each of the latter two. As usual, we make no notational distinction between these maps and their isotopy classes.

As is well known, the restriction of α to ∂H is central in the mapping class group of ∂H and acts trivially on every simple closed loop in ∂H. Consequently α is a central involution in $\mathcal{G}$ which acts trivially on every disk in H. It is easily seen that $\beta\gamma\beta^{-1} = \alpha\gamma$. From [2] and also, using the present viewpoint, from [3], we have:

Proposition 9.1   *The stabilizer $\mathcal{G}_{\mu_0}$ is the subgroup generated by α, β and γ. In fact, $\mathcal{G}_{\mu_0}$ is the semidirect product $(C_2 \times \mathbb{Z}) \circ C_2$, where $\langle \alpha, \beta \rangle$ is the normal subgroup $C_2 \times \mathbb{Z}$ and γ acts by $\gamma\alpha\gamma^{-1} = \alpha$ and $\gamma\beta\gamma^{-1} = \alpha\beta$.*

Now we analyze the action of $\mathcal{G}$ on the slope disks associated to μ_0. Each cable for μ_0, and hence each slope disk for μ_0, is invariant under α and γ. In terms of the covering space $\widetilde{\Sigma}_0$ in Figure 7, α lifts to a horizontal translation by 1, and γ lifts to multiplication by $-I$.

To understand the action of β, fix some primitive slope disk τ_0 for μ_0, and consider $(\mu_0; \tau_0)$–slope pairs. The associated perpendicular disk τ_0^0 is exactly the intersection of H with the splitting sphere disjoint from $\lambda_0 \cup \rho_0$. A lift of β to $\widetilde{\Sigma}_0$ sends (x, y) to $(x + y, y)$, so β sends a cabling arc with slope pair $[p, q]$ to one with slope pair $[p + q, q]$.

The next observation is that the action of $\mathcal{G}_{\mu_0}$ is transitive on primitive slope disks for μ_0. Perhaps the easiest way to see this is that the β–translates of τ_0 are the only slope disks for μ_0 whose complementary tori in H are unknotted in S^3 (for the others, the core circle of the complementary torus is a nontrivial 2–bridge knot; see Section 11).

We claim that slope disks of μ_0 are equivalent under $\mathcal{G}$ only when they are equivalent under $\mathcal{G}_{\mu_0}$. We have just seen that the action of $\mathcal{G}_{\mu_0}$ on the primitive slope disks of μ_0 is transitive. Consider two nonprimitive slope disks τ_1 and τ_2 of μ_0 that are equivalent by an element g of $\mathcal{G}$. The shortest path ω_i in $\widetilde{\mathcal{T}}$ from any vertex in the link of τ_i in $\mathcal{D}'(H)$ to a vertex of the primitive tree $\widetilde{\mathcal{T}}_0$ is the edge of $\widetilde{\mathcal{T}}$ from $\mu_0 \cup \{\tau_i\}$ to μ_0. Since the primitive tree is invariant, g must take ω_1 to ω_2 and hence $g \in \mathcal{G}_{\mu_0}$.

Since α and γ act trivially on slope disks associated to μ_0, while β sends a $[p, q]$–disk to a $[p + q, q]$–disk, it follows that *sending a $[p, q]$–slope disk associated to μ_0 to p/q induces a bijection from the set of $\mathcal{G}$–orbits of slope disks associated to μ_0 to the set $\mathbb{Q}/\mathbb{Z} \cup \{\infty\}$*. The slope disk that corresponds to ∞ is the one with slope pair $[1, 0]$, that is, τ_0^0.

Since any other choice of τ_0 differs from our previous one by the action of a power of β, the bijection to $\mathbb{Q}/\mathbb{Z} \cup \{\infty\}$ is independent of the choice of τ_0. Moreover, this shows that the primitive slope disks are those having slope pair of the form $[p, 1]$, so they correspond to $[0] \in \mathbb{Q}/\mathbb{Z} \cup \{\infty\}$.

The slope disks for a primitive pair play a key role in our theory, so we introduce some special terminology.

Definition 9.2 A possibly separating disk in H is called *simple* if it is nonprimitive and is a slope disk for a primitive pair.

Definition 9.3 The *simple slope* of a simple disk is the corresponding element $[p/q] \in (\mathbb{Q}/\mathbb{Z} - \{0\}) \cup \{\infty\}$, which has q odd if and only if the disk is nonseparating.

10 The tree of knot tunnels

In this section we will analyze the quotient $\mathcal{D}(H)/\mathcal{G}$ and the tree $\mathcal{T}$. We begin with a summary description, which we will establish in the remainder of this section. It may be useful at this point to refer to Figure 4. To fix notation, let θ_0 be a primitive triple $\{\lambda_0, \rho_0, \tau_0\}$, containing the primitive pair $\mu_0 = \{\lambda_0, \rho_0\}$. We recall that π_0 denotes the $\mathcal{G}$–orbit of primitive disks, which is representable by any of λ_0, ρ_0, or τ_0. As before, we will use μ_0 and θ_0 to denote their images in $\mathcal{D}(H)/\mathcal{G}$, the unique orbits or primitive pairs and primitive triples.

(1) The quotient of the subcomplex of $\mathcal{D}(H)$ spanned by primitive vertices is a single simplex Π spanned by π_0, μ_0 and θ_0. It is a copy of a 2–simplex of $\mathcal{D}'(H)$ and meets $\mathcal{T}$ in the edge $\langle \mu_0, \theta_0 \rangle = \widetilde{\mathcal{T}}_0/\mathcal{G}$.

(2) Each 2–simplex of $\mathcal{D}(H)$ having two primitive vertices and one simple vertex gets identified with some other such simplices, then folded in half by the action of some conjugate of the element γ of Section 9, and the quotient half-simplex is attached to Π along the edge $\langle \pi_0, \mu_0 \rangle$. These half-simplices correspond to the elements of $\mathbb{Q}/\mathbb{Z} - \{0\}$ with odd denominator. The intersection of $\mathcal{T}$ with such a half-simplex is a pair of edges.

(3) Each component of the remainder of $\mathcal{D}(H)$ descends injectively to $\mathcal{D}(H)/\mathcal{G}$. In fact, it is identified with some other such components, and the result is attached to a half-simplex of $\mathcal{D}(H)/\mathcal{G}$ along the edge from π_0 to the simple vertex. The intersection of $\mathcal{T}$ with each component of the remainder of $\mathcal{D}(H)/\mathcal{G}$ is exactly a copy of the intersection of $\widetilde{\mathcal{T}}$ with a component of the remainder of $\mathcal{D}(H)$.

Now we present more detail of this description. Part (1) should be clear since the action of $\mathcal{G}$ is transitive on primitive triples, and the stabilizer of θ_0 acts as the permutation group on the three disks in θ_0.

Consider a 2–simplex of $\mathcal{D}(H)$ which has λ_0 and ρ_0 as two of its vertices, and third vertex a simple disk. We saw in Section 9 that the generator α of $\mathcal{G}_{\mu_0}$ acts trivially on every 2–simplex, and the generator β preserves each of λ_0 and ρ_0, but acts on slope disks for μ_0 by sending the disk with slope pair $[p, q]$ to the one with slope pair $[p + q, q]$. Finally, γ interchanges λ_0 and ρ_0, while preserving each slope disk associated to μ_0. So the effect of $\mathcal{G}_{\mu_0}$ on the collection of 2–simplices of $\mathcal{D}(H)$ having λ_0, ρ_0, and a simple disk as vertices is to identify those for which the ratio p/q differs by an integer, and to fold the resulting 2–simplices in half, producing half-sized simplices in $\mathcal{D}(H)/\mathcal{G}$ with short edges attached to Π along the 1–simplex $\langle \pi_0, \mu_0 \rangle$. There is one such half-simplex for each orbit of nonseparating simple disks under $\mathcal{G}_{\mu_0}$, and as seen in Section 9, these orbits correspond to the elements of $\mathbb{Q}/\mathbb{Z} - \{0\}$ with odd denominator.

To understand the remaining portions of $\mathcal{D}(H)/\mathcal{G}$, fix a simple disk τ, and consider a portion of $\mathcal{D}(H)$ attached to the 2–simplex $\langle \lambda_0, \rho_0, \tau \rangle$ along the edge $\langle \rho_0, \tau \rangle$. We know that the action of β moves it to similar portions attached along edges of other 2–simplices having λ_0 and ρ_0 as vertices. The action of γ interchanges λ_0 and ρ_0, so interchanges the portion with another one attached along $\langle \lambda_0, \tau \rangle$. Finally, α acts trivially. So the original portion descends injectively into $\mathcal{D}(H)/\mathcal{G}$ onto a copy attached along the edge $\langle \pi_0, \tau \rangle$.

The quotient of the primitive subtree $\widetilde{\mathcal{T}}_0$ is the edge $\langle \theta_0, \mu_0 \rangle$ of Π. The rest of $\mathcal{T}$ should be clear from the previous discussion and Figure 4. In $\mathcal{T}$, the vertex θ_0 has valence 1, and μ_0 has countable valence. Each additional edge emanating from μ_0 goes to a white vertex of valence 2, representing a triple with two primitive and one simple disk as vertices. From then on, $\mathcal{T}$ looks exactly like a portion of $\widetilde{\mathcal{T}}$.

We now have a complete picture of $\mathcal{D}(H)/\mathcal{G}$ and $\mathcal{T}$. It will also be useful to know the stabilizers of the action of $\mathcal{G}$ on nonprimitive disks in H.

Proposition 10.1 *The stabilizer in $\mathcal{G}$ of a nonseparating, nonprimitive disk τ in H is as follows:*

(i) *If τ simple, then its stabilizer is conjugate to the subgroup $C_2 \times C_2$ generated by α and γ.*

(ii) *Otherwise, the stabilizer is C_2, generated by α.*

Proof Suppose first that τ simple. There is a unique primitive pair μ which together with τ spans a 2–simplex of $\mathcal{D}(H)$, for if there were two such pairs μ and μ', then the 1–simplices $\langle \mu \cup \{\tau\}, \mu \rangle$ and $\langle \mu' \cup \{\tau\}, \mu' \rangle$, connected by an arc in the primitive tree, would form an imbedded loop in $\widetilde{\mathcal{T}}$. By conjugation, we may assume that μ is the standard primitive pair μ_0.

Let h be in $\mathcal{G}_\tau$. Since h also preserves the primitive subcomplex of $\mathcal{D}(H)$, it must preserve μ_0. From Proposition 9.1, any element of $\mathcal{G}_{\mu_0}$ can be written in the form $\alpha^{\epsilon_1} \beta^n \gamma^{\epsilon_2}$, where the $\epsilon_i \in \{0, 1\}$ and $n \in \mathbb{Z}$. Since h preserves the slope pair of τ, we have $n = 0$ and part (i) follows.

For (ii), suppose that τ is not simple. Let θ be the vertex of the link of τ in $\mathcal{D}'(H)$ that is closest in $\widetilde{\mathcal{T}}$ to the primitive subtree $\widetilde{\mathcal{T}}_0$. Since the action of $\mathcal{G}$ fixes τ and preserves $\widetilde{\mathcal{T}}_0$, it must fix the path from θ to the nearest point μ of $\widetilde{\mathcal{T}}_0$ (we will see in Section 12 that the image of this path in $\mathcal{D}(H)/\mathcal{G}$, together with θ_0, is the *principal path* of τ, illustrated in Figure 8). By conjugation, we may assume that $\mu = \mu_0$. Write this path as $\mu_0, \mu_0 \cup \{\tau_0\}, \mu_1, \mu_1 \cup \{\tau_1\}, \ldots, \theta$. As in part (i), any element of $\mathcal{G}_\tau$ fixes this path, so must be in the stabilizer $\langle \alpha, \gamma \rangle$ of τ_0. But γ interchanges the disks of μ_0, so acts as reflection on the 2–simplex spanned by $\mu_0 \cup \{\tau_0\}$ and cannot fix μ_1. Part (ii) follows. $\qquad\square$

Corollary 10.2 *Let $\langle \tau_1, \tau_2, \tau_3 \rangle$ be a 2–simplex of $\mathcal{D}(H)$, with τ_3 nonprimitive. Suppose that $\langle \tau_1, \tau_2, \tau_3 \rangle$ is stabilized by an element h of $\mathcal{G}$ other than the identity or α. Then τ_3 is simple, τ_1 and τ_2 are primitive, and h is conjugate to γ or $\alpha\gamma$.*

Proof Of the three vertices $\langle \tau_1, \tau_2 \rangle$, $\langle \tau_1, \tau_3 \rangle$ and $\langle \tau_2, \tau_3 \rangle$, let $\langle \tau_i, \tau_j \rangle$ be the one closest to $\widetilde{\mathcal{T}}_0$ (possibly in $\widetilde{\mathcal{T}}_0$), and let τ_k be the vertex of $\langle \tau_1, \tau_2, \tau_3 \rangle$ different from τ_i and τ_j. The action of h must preserve $\langle \tau_i, \tau_j \rangle$, so must fix τ_k.

Since $\langle \tau_i, \tau_j \rangle$ is the closest vertex to $\widetilde{\mathcal{T}}_0$, τ_k is nonprimitive. Proposition 10.1 then shows that τ_k is simple and h is conjugate to either γ or $\alpha\gamma$. Again since $\langle \tau_i, \tau_j \rangle$ is the closest vertex to $\widetilde{\mathcal{T}}_0$, τ_i and τ_j must both be primitive, so $\tau_k = \tau_3$ and the corollary is established. $\qquad\square$

11 Simple, semisimple and regular tunnels

A $\mathcal{G}$–orbit of nonseparating simple disks is called a *simple tunnel*.

Proposition 11.1 *The simple tunnels are exactly the "upper" and "lower" tunnels of 2–bridge knots.*

Proof   Let Σ_0 be as in Section 9. If σ is a simple tunnel, then K_σ is isotopic to the union of the cable in Σ_0 determined by σ, plus two arcs in $\overline{\partial H - \Sigma_0}$, each crossing λ_0 or ρ_0 in one point. A dual arc to σ is an arc cutting once across σ and connecting the cable arcs. This is a standard description of the upper and lower tunnels of 2–bridge knots.   □

From Sections 9 and 10, we have immediately:

Proposition 11.2   *The simple tunnels are classified up to equivalence by their simple slopes in $\mathbb{Q}/\mathbb{Z} - \{0\}$, and up to possibly orientation-reversing equivalence by the pairs $\{[p/q], [-p/q]\}$ with q odd.*

Of course, $[p/q]$ is a version of the standard rational invariant that classifies 2–bridge knots. This will be examined further in Section 15.

Recall that a θ–curve in S^3 is called *unknotted* if its complement is an open handlebody and *planar* if it is isotopic into a standard plane. The results of Section 10 allow us to describe the automorphisms of nonplanar unknotted θ–curves.

Corollary 11.3   *Let Θ be a nonplanar unknotted θ–curve in S^3, and let h be an orientation-preserving homeomorphism of S^3 that preserves Θ. Then, either*

  (i)   *h is isotopic preserving Θ to a homeomorphism which is the identity or the hyperelliptic involution on a neighborhood of Θ, or*

  (ii)   *Θ is the union of a 2–bridge knot and one of its simple tunnel arcs, and h preserves the tunnel arc and interchanges the two arcs of the knot.*

Proof   By isotopy we may assume that $\Theta \subset H$ and $h \in \mathcal{G}$. In $\mathcal{D}(H)$, the dual disks of the arcs of Θ span a 2–simplex, which is preserved by h. Since Θ is nonplanar, the dual disk of the tunnel is not primitive. Applying Corollary 10.2, either case (i) holds, or the tunnel disk is simple and h is conjugate to γ or to $\alpha\gamma$, giving case (ii).   □

Remark 11.4   Corollary 11.3 holds as stated without the assumption that h is orientation-preserving, since Proposition 17.2 below shows that nonplanar unknotted θ–curves have no orientation-reversing automorphisms.

Corollary 11.3 gives immediately a result of D Futer [9]:

Theorem 11.5   (Futer)   *Let A be a tunnel arc for a nontrivial knot $K \subset S^3$. Then A is fixed pointwise by a strong inversion of K if and only if K is a two-bridge knot and A is its upper or lower tunnel.*

Definition 11.6 A tunnel τ is called *semisimple* if τ is not primitive or simple but τ lies in the link in $\mathcal{D}(H)$ of a primitive disk. It is called *regular* if it is neither primitive, simple, or semisimple.

Recall that a $(1, 1)$–knot is a knot of the form $t_1 \cup t_2$, where t_1 and t_2 are trivial (boundary parallel) arcs in the tori W_1 and W_2 of a genus–1 Heegaard splitting of S^3. A $(1, 1)$–*tunnel* is a tunnel of a $(1, 1)$–knot which is representable by an arc t in W_1 which meets t_1 in its endpoints and together with the arc that ∂t bounds in t_1 forms a core circle of W_1. There is also a $(1, 1)$–tunnel obtained by the corresponding construction in W_2, which may be equivalent to the one in W_1. The knot may have more $(1, 1)$–tunnels coming from other $(1, 1)$–descriptions, or may have a regular tunnel (as occurs for most torus knots). The following fact is well-known; see for example [22, Proposition 1.3]. A fairly brief proof can be obtained using Gordon's Theorem [13]. We give an independent proof from our viewpoint.

Proposition 11.7 *A tunnel is a $(1, 1)$–tunnel for a nontrivial $(1, 1)$–knot if and only if it is simple or semisimple.*

Proof Suppose first that τ is simple or semisimple. By definition, τ lies in a 2–simplex $\{\pi, \lambda, \tau\}$ with π primitive. Let α be a tunnel arc of K_τ for which the θ–curve $K_\tau \cup \alpha$ is dual to this triple. Let N be a small regular neighborhood of π in H, and consider the solid torus $W = \overline{H - N}$. Since π is primitive, W is unknotted. The existence of a dual disk to π in $S^3 - H$ shows that $K_\tau \cap N$ is a trivial arc in $\overline{S^3 - W}$. Moreover, since $K_\tau \cup \alpha$ is dual to $\{\pi, \lambda, \tau\}$, $K_\tau \cap W$ is a trivial arc in W and $(K_\tau \cup \alpha) \cap W$ contains a core circle of W. Thus τ is a $(1, 1)$–tunnel.

Conversely, suppose that τ is a $(1, 1)$–tunnel. Then the union of one of the arcs of K_τ with α is a core circle of an unknotted solid torus, so the disk λ or ρ dual to the other arc of K_τ is primitive. Therefore τ is simple or semisimple. $\square$

Remark 11.8 Most torus knots have a regular tunnel and two semisimple tunnels [8]. H Goda and C Hayashi [10] give an example of a nontorus knot with a regular tunnel and a semisimple tunnel. We do not know an example of a knot that has more than one regular tunnel.

12 Principal paths, principal vertices and parameterization

Using the coordinates from Section 8, we will obtain a natural numerical parameterization of all knot tunnels. First, we give an important definition.

Definition 12.1 Let τ be a tunnel. The *principal vertex* of τ is the vertex of the link of τ in $\mathcal{D}'(H)/\mathcal{G}$ that is closest to θ_0 in $\mathcal{T}$. The *principal path* of τ is the unique path in $\mathcal{T}$ from θ_0 to the principal vertex of τ.

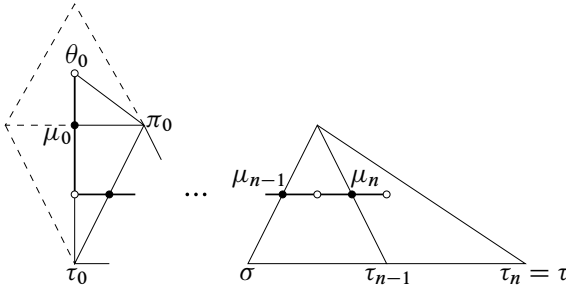Figure 8 illustrates the principal path and principal vertex of a typical tunnel.

Figure 8: The principal path of τ is the path in $\mathcal{T}$ from θ_0 to the principal vertex $\mu_n \cup \{\tau\}$ of τ. The "trailing" disk σ, which is the disk of $\mu_{n-1} - \mu_n$, plays an important role in the calculation of slope invariants.

Remark 12.2 We will see in Lemma 14.1 that the principal vertex of τ is $\langle \mu^+, \mu^-, \tau \rangle$, where μ^+ and μ^- is the pair of disks used in the definition of the Scharlemann–Thompson invariant [26]. Thus in Figure 8, $\mu_n = \{\mu^+, \mu^-\}$.

We can now give the numerical parameterization.

Parameterization Theorem 12.3 *Let τ be a knot tunnel with principal path θ_0, μ_0, $\mu_0 \cup \{\tau_0\}$, $\mu_1, \ldots, \mu_n$, $\mu_n \cup \{\tau_n\}$. Fix a lift of the principal path to $\mathcal{D}(H)$, so that each μ_i corresponds to an actual pair of disks in H.*

(1) *If τ is primitive, put $m_0 = [0] \in \mathbb{Q}/\mathbb{Z}$. Otherwise, let $m_0 = [p_0/q_0] \in \mathbb{Q}/\mathbb{Z}$ be the simple slope of τ_0.*

(2) *If $n \geq 1$, then for $1 \leq i \leq n$ let σ_i be the unique disk in $\mu_{i-1} - \mu_i$ and let $m_i = q_i/p_i \in \mathbb{Q}$ be the $(\mu_i; \sigma_i)$–slope of τ_i.*

(3) *If $n \geq 2$, then for $2 \leq i \leq n$ define $s_i = 0$ or $s_i = 1$ according to whether or not the unique disk of $\mu_i \cap \mu_{i-1}$ equals the unique disk of $\mu_{i-1} \cap \mu_{i-2}$.*

Then, sending τ to the pair $((m_0, \ldots, m_n), (s_2, \ldots, s_n))$ is a bijection from the set of all tunnels of all tunnel number 1 knots to the set of all elements

$$((\,[p_0/q_0],\, q_1/p_1, \ldots, q_n/p_n), (s_2, \ldots, s_n))$$

in $$\left(\mathbb{Q}/\mathbb{Z}\right) \cup \left(\mathbb{Q}/\mathbb{Z} \times \mathbb{Q}\right) \cup \left(\cup_{n \geq 2}\, \mathbb{Q}/\mathbb{Z} \times \mathbb{Q}^n \times C_2^{n-1}\right)$$

with all q_i odd.

Proof To see that the slopes are well-defined, consider another lift of the principal path of τ, given by a sequence of vertices θ'_0, μ'_0, $\mu'_0 \cup \{\tau'_0\}$, $\mu'_1, \ldots, \mu'_n$, $\mu'_n \cup \{\tau'_n\}$ of $\widetilde{\mathcal{T}}$ in $\mathcal{D}(H)$. There is an element h of the Goeritz group taking τ_0 to τ'_0. Each of τ_0 and τ'_0 is adjacent to only two primitive vertices, so h must take μ_0 to μ'_0. According to Proposition 10.1, the stabilizer of μ'_0 is generated by α and (an appropriate conjugate of) γ; since α acts trivially on all of $\mathcal{D}(H)$, $h(\mu_1)$ and μ'_1 must either be equal or differ by the action of γ. Since γ preserves μ'_0, we may compose h with γ if necessary to assume that h also takes μ_1 to μ'_1. Again by Proposition 10.1, the stabilizer of each remaining vertex of the two lifts is generated by α, so h must carry the entire first lift to the second. Since all slopes are defined to be invariant under the application of an element of $\mathcal{G}$, the vector $([p_0/q_0], q_1/p_1, \ldots, q_n/p_n)$ is well-defined.

From the definitions in Sections 8 and 9, all q_i must be odd, and any such vector as in the theorem must occur for some path in $\mathcal{D}(H)$ that projects to the principal path of some vertex τ of $\mathcal{D}(H)/\mathcal{G}$.

Finally, the sequence $(s_2, \ldots, s_n)$ needs explanation. Referring to Figure 4, we note that the value of m_0 determines a unique simple tunnel τ_0 (unless $m_0 = [0]$, which determines the primitive tunnel), and the value of m_1 determines a unique choice of τ_1. From then on, one must make a choice of which disk of μ_{i-1} will be retained in μ_i, and the number s_i simply records this choice. As with the slopes, this vector is well-defined, and any sequence can be achieved. $\square$

Remark 12.4 Theorem 16.1 below is a version of the Parameterization Theorem 12.3 that includes tunnels of tunnel number 1 links. The main difference is that the final slope m_n may have q_n even.

Remark 12.5 Reversing the orientation of S^3 has the effect of negating each slope invariant, so the classification of tunnels up to arbitrary homeomorphism of S^3 is obtained from that of the Parameterization Theorem 12.3 by adding the equivalence $(m_0, m_1, \ldots, m_n) \sim (-m_0, -m_1, \ldots, -m_n)$.

13 The cabling construction

In this section, we will see how the tree $\mathcal{T}$ specifies a unique sequence of "cabling constructions" that produce a given tunnel. Roughly speaking, a path of length 2 from a white vertex to a white vertex corresponds to one cabling construction. The principal path of τ encodes the unique sequence of cabling constructions that produces τ.

The cabling construction is simple and will look familiar to experts. In a sentence, it is "Think of the union of K and the tunnel arc as a θ–curve, and rationally tangle the

ends of the tunnel arc and one of the arcs of K in a neighborhood of the other arc of K." We sometimes call this "swap and tangle," since one of the arcs in the knot is exchanged for the tunnel arc, then the ends of the other arc of the knot and the tunnel arc are connected by a rational tangle.

The cabling operation is a very restricted special case of the "tunnel moves" described in Goda, Scharlemann and Thompson [11]. A tunnel move means a replacement of K_τ by any knot in ∂H that crosses $\partial \tau$ exactly once. A tunnel move is a composition of an arbitrarily large number of cabling constructions that retain the same arc of the knot. As mentioned in Section 2, tunnel moves are studied from our viewpoint in [4].

We begin with some terminology.

Definition 13.1 Let μ be a pair, and let τ be a disk in H. We say that μ is a *meridian pair of* τ when τ is a slope disk of μ. In this case, the pair $(\mu; \tau)$ corresponds to the directed 1–simplex in $\mathcal{T}$ (or $\widetilde{\mathcal{T}}$) from μ to $\mu \cup \{\tau\}$.

The meridian pairs of τ correspond exactly to the white vertices of the link of τ in $\mathcal{D}'(H)$.

Geometrically, $(\mu; \tau)$ corresponds to an isotopy class of tunnel arc of K_τ. In a θ–curve corresponding to $\mu \cup \{\tau\}$, the union of the arcs dual to μ is K_τ, and the arc dual to τ is a tunnel arc for K_τ. Here, isotopy class refers to the isotopy class in H when working in $\widetilde{\mathcal{T}}$, and the isotopy class in S^3 (possibly moving K_τ along with the arc) when working in $\mathcal{T}$.
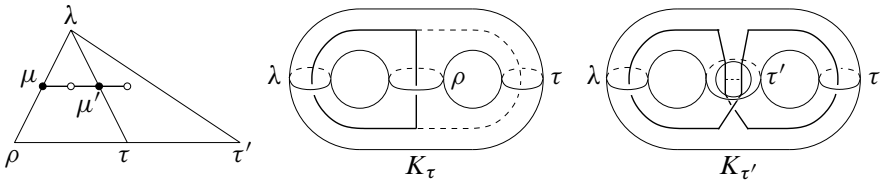

Figure 9: Schematic for the general cabling construction. In the middle ball in the right-hand picture of H, the two vertical arcs form some rational tangle, disjoint from the disk τ'.

Moving through the tree determines a sequence of steps in which one of the two disks of a pair $\{\lambda, \rho\}$ is replaced by a tunnel disk τ, and a slope disk τ' of the new pair μ' (with τ' nonseparating in H) is chosen as the new tunnel disk. As illustrated in Figure 9, the way the path determines the particular cabling operation is:

(1) The selection of λ or ρ corresponds to which edge one chooses to move out of the white vertex $\{\lambda, \rho, \tau\}$.

(2) The selection of the new slope disk τ' corresponds to which edge one chooses to continue out of the black vertex μ'.
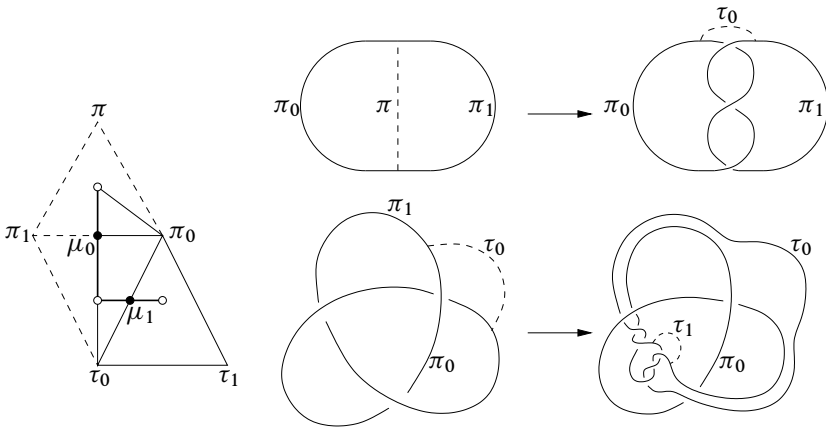
Figure 10: Examples of the cabling construction

[Figure 10](#) shows the effects of a specific sequence of two cabling constructions, starting with the trivial knot and obtaining the trefoil, then starting with the tunnel of the trefoil.

As usual, let μ_0 be a primitive pair, and let τ_0 be a simple disk for μ_0, with simple slope m_0. The segment $(\mu_0; \tau_0)$ determines a cabling construction starting with the tunnel of the trivial knot and producing a 2–bridge knot. We call this a *simple cabling of slope m_0*.

Now, consider a path of length 3 in $\mathcal{T}$ determined by the four vertices μ, $\mu \cup \{\tau\}$, μ', $\mu' \cup \{\tau'\}$, that is, by a succession from $(\mu; \tau)$ to $(\mu'; \tau')$, as in [Figure 9](#). Let σ be the unique disk of $\mu - \mu'$ (the "trailing" disk). In the case shown in [Figure 9](#), σ happens to be ρ, so that $\mu' = \{\lambda, \tau\}$. Denoting by m the $(\mu'; \sigma)$–slope of τ', we call the corresponding cabling construction a *cabling of slope m*. We require that $\tau' \neq \sigma$, that is, cablings do not allow one to "backtrack" in $\mathcal{T}$. In terms of slope, such a cabling would have $m = \infty$.

On the other hand, any cabling construction in H corresponds to an arc of length 3 starting at a black vertex in $\widetilde{\mathcal{T}}$, and any cabling construction in S^3 that does not produce the trivial tunnel or a simple tunnel corresponds to such an arc in $\mathcal{T}$. One of our main theorems is now immediate; it is just a geometric restatement of the [Parameterization Theorem 12.3](#).

Unique Cabling Sequence Theorem 13.2 *Let τ be a tunnel of a nontrivial knot. Let θ_0, μ_0, $\mu_0 \cup \{\tau_0\}$, $\mu_1, \ldots, \mu_n$, $\mu_n \cup \{\tau_n\}$ with $\tau_n = \tau$ be the principal path of τ. Then the sequence of $n + 1$ cablings consisting of the simple cabling determined by $(\mu_0; \tau_0)$ and the cablings determined by the successions from $(\mu_{i-1}; \tau_{i-1})$ to $(\mu_i; \tau_i)$ is the unique sequence of cablings beginning with the tunnel of the trivial knot and ending with τ.*

Of course, the values $[p_0/q_0]$ and q_i/p_i in the Parameterization Theorem 12.3 are exactly the slopes of the cablings.

One of the slope parameters in the Unique Cabling Sequence Theorem 13.2 will be the subject of Section 14:

Definition 13.3 Let τ be a tunnel for a nontrivial knot. If τ is not simple, then the slope m_n is called the *principal slope* of τ. When τ is simple, its principal slope is undefined.

In some sense, the Unique Cabling Sequence Theorem 13.2 enables one to distinguish any two tunnels. If one finds any sequence of cablings that produce the tunnel, it must be the *unique* such sequence, and when the sequences are different for two tunnels, the tunnels are inequivalent. Theoretically this sequence of cablings, ie the principal path, can be determined algorithmically starting from a specific representative disk $D \subset H$ of the tunnel: Working in $\widetilde{\mathcal{T}}$, fix a primitive pair $\mu_0 = \{\lambda_0, \rho_0\}$, and let W be a wave for D with respect to μ_0 with W meeting, say, ρ_0. Let τ_0 be the slope disk of μ_0 determined by W. Then D has fewer components of intersection with $\lambda_0 \cup \tau_0$ than with $\lambda_0 \cup \rho_0$. Put $\mu_1 = \{\lambda_0, \tau_0\}$ and repeat this process inductively using a wave of D with respect to μ_1. When D has no wave, one is at the principal vertex $\mu_n \cup \{D\}$. Some of the initial disks $\tau_0, \tau_1, \dots$ may be primitive, but the portion starting with the last $\mu_i \cup \{\tau_i\}$ that is a primitive triple will descend to the principal path in $\mathcal{T}$. This does not seem to be a practical algorithm.

14   The Scharlemann–Thompson invariant

The Scharlemann–Thompson invariant, developed in [26] and further used in [25; 23], is essentially the principal slope of τ. The construction in [26] proceeds as follows: (1) Intersect a splitting sphere S with H, obtaining a separating disk E. (2) Take an arc of $E \cap \tau$ outermost on E and cutting off a subdisk E' of E. (3) Take as μ^+ and μ^- the two components of the frontier of a regular neighborhood of $\tau \cup E'$ that are not parallel to τ. Then the invariant is defined to be the $(\{\mu^+, \mu^-\}; \tau)$–slope of a wave of E with respect to $\{\mu^+, \mu^-\}$. No selection of a canonical perpendicular disk τ^0 is made, so the invariant is regarded as an element of $\mathbb{Q}/2\mathbb{Z}$ since changing the choice of $\tau^\perp$ changes the slope by a multiple of 2. We will regard the Scharlemann–Thompson invariant as $\mathbb{Q}$–valued, by using τ^0 as the perpendicular disk.

To understand the relation between the Scharlemann–Thompson invariant and the principal slope, we use the following lemma.

Lemma 14.1 *Let τ be a nonsimple tunnel of a nontrivial knot. Then:*

(i) *The disks called μ^+ and μ^- in the definition of the Scharlemann–Thompson invariant form the principal meridian pair of τ. That is, the principal vertex of τ is $\{\tau, \mu^+, \mu^-\}$.*

(ii) *The Scharlemann–Thompson invariant is the $(\mu_n; \tau)$–slope of the unique disk σ_n of $\mu_{n-1} - \mu_n$.*

Proof Before beginning the proof, we remind the reader that any finite collection of essential disks in H can be moved by isotopies so that any two of them intersect minimally. A nice way to do this is to choose a hyperbolic structure on ∂H, move the boundaries of the disks by isotopies to be geodesics, and eliminate simple closed curve intersections of their interiors by further isotopies. If we choose our hyperbolic structure so that the hyperbolic involution α of H is an isometry, then the boundaries will be invariant under α, indeed every simple closed curve is invariant up to isotopy so every geodesic is invariant. In particular, we may assume that α preserves the boundaries of τ, μ^+, μ^- and E.

As usual, denote the vertices in the principal path of τ by θ_0, μ_0, $\mu_0 \cup \{\tau_0\}$, $\mu_1, \ldots,$ μ_n, $\mu_n \cup \{\tau\}$, so that $\mu_n \cup \{\tau\}$ is the principal vertex of τ, and denote the unique disk of $\mu_{n-1} - \mu_n$ by σ_n. We will first show that if μ is any black vertex of the link of τ in $\mathcal{D}'(H)/\mathcal{G}$ other than μ_n, then the $(\mu; \tau)$–slope of E is ∞.
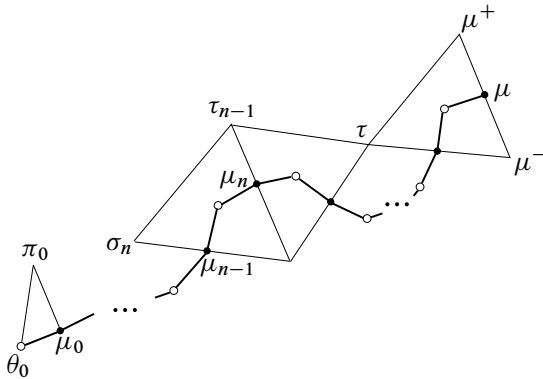

Figure 11: Verification that the principal vertex is $\{\mu^+, \mu^-, \tau\}$

Figure 11 illustrates the path in $\mathcal{T}$ from θ_0 to μ if the principal vertex is not $\mu \cup \{\tau\}$. It is possible that one of the disks of μ equals a disk of μ_n, in which case the two rightmost triangles in Figure 11 share a side.

The splitting sphere S that contains E is disjoint from a primitive disk D_0. Assuming that $\mu \neq \mu_n$, there is a sequence D_0, $D_1, \ldots, D_k$ of disks in H such that:

(1) For each $1 \le i \le k$, D_{i-1} is disjoint from D_i.

(2) $D_{k-2} = \sigma_n$ and $D_k = \tau$.

(3) No D_i is a disk of μ (if it happens that one of the disks of μ is a disk of μ_n, then D_{k-1} will be the other disk of μ_n).
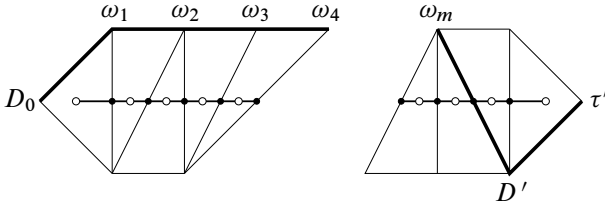

Figure 12: Constructing a path that avoids μ

Figure 12 illustrates one way to obtain such a sequence: Choose a representative disk τ' of τ, and choose a primitive pair ν_1 such that $\nu_1 \cup \{D_0\}$ is a primitive triple, ie a representative of θ_0. Let $\nu_1 \cup \{D_0\}$, ν_1, $\nu_1 \cup \delta_1, \ldots, \nu_{\ell-1} \cup \delta_{\ell-1}$, ν_ℓ, $\nu_\ell \cup \{\tau'\}$ be the shortest path in $\widetilde{\mathcal{T}}$ from $\nu_1 \cup \{D_0\}$ to the link of τ'. Some final segment of this path projects to the principal path of τ, and in particular ν_ℓ projects to μ_n. The 2–simplices of $\mathcal{D}(H)$ that contain one of the white vertices of this path form a "corridor" which is topologically a disk and is illustrated in Figure 12. The boundary of this corridor contains two arcs in the 1–skeleton of $\mathcal{D}(H)$ that run from D_0 to the disk ω that projects to σ_n; let D_0, $\omega_1, \ldots, \omega_m = \omega$ be the one that does not contain τ'. Since ω_m projects to σ_n, it is a vertex of a 2–simplex whose other vertices are the disks of ν_ℓ. Therefore ω_m is disjoint from both of the disks of ν_ℓ. At least one of these disks, say D', is not in μ, so one sequence of disks satisfying the three conditions given above is D_0, $\omega_1, \ldots, \omega_m$, D', τ'.

Proposition 8.3 now shows that the $(\mu; \tau)$–slope of E equals the $(\mu; \tau)$–slope of τ, which is ∞.

Now put $\mu = \{\mu^+, \mu^-\}$. To prove part (i), it suffices to show that E does not have infinite $(\mu; \tau)$–slope. If E is a slope disk for μ, then it has finite slope since it separates H. Suppose it is not a slope disk, but has infinite $(\mu; \tau)$–slope. Let W be a wave of E with respect to μ, and consider the arc $W \cap \partial H$. Either this arc or its image under α lies on the side of $\mu^+ \cup \tau \cup \mu^-$ that contains E'. Since E is invariant under α, both of these arcs lie in E and are disjoint from E'. But this contradicts the fact that W is essential.

For part (ii), we have seen that $\mu = \mu_n$ and that for any splitting sphere S, the disks $S \cap H$ and D_{k-2} from a sequence as above have the same $(\mu_n; \tau)$–slope. That is, the $(\mu_n; \tau)$–slope of a wave of $S \cap H$ is the same as the slope of σ_n.   □

The previous argument clarifies the fact that the $(\mu; \tau)$–slope of E is infinite for any meridian pair μ other than the principal meridian pair [26, Lemma 2.9]. It also explains why the Scharlemann–Thompson definition gives nonunique values for the upper and lower tunnels of 2–bridge knots [26, Corollary 2.8]. For these cases, the principal path is just θ_0, μ_0, $\mu_0 \cup \{\tau\}$. In the sequence of disks $D_0, D_1, \ldots, D_k$ in the proof of Lemma 14.1, different choices of E and hence of D_0 can give different choices for D_{k-2}, with different slopes. From our viewpoint, these are the simple tunnels, for which the principal slope is undefined.

Theorem 14.5 below gives the expression for the Scharlemann–Thompson invariant and the principal slope m_n in terms of each other. To obtain this, we must understand how to change coordinates on the slope disks at the principal meridian pair of the tunnel, and we set this up as a general principle. Recall that for integers $a_1, \ldots, a_k$, the continued fraction $[a_1, \ldots, a_k]$ is defined inductively by $[a_1] = a_1$ and $[a_1, \ldots, a_k] = a_1 + 1/[a_2, \ldots, a_k]$. Sometimes one may choose to allow some of the a_i to be ∞. We have $-[a_1, \ldots, a_k] = [-a_1, \ldots, -a_k]$, and

$$[\ldots a_{i-1}, 0, a_{i+1} \ldots] = [\ldots a_{i-1} + a_{i+1} \ldots]$$

so in particular

$$[\ldots, a_{k-1}, a_k, 0] = [\ldots, a_{k-1}, a_k, 0, \infty] = [\ldots, a_{k-1}, \infty] = [\ldots, a_{k-1}].$$

Another basic fact is:

Lemma 14.2 *Let $q/p \in \mathbb{Q}$. Then q/p may be written as a continued fraction as $[2a_1, 2b_1, 2a_2, \ldots, 2b_{k-1}, 2a_k]$ or $[2a_1, 2b_1, 2a_2, \ldots, 2b_{k-1}, 2a_k, b_k]$, with all entries nonzero except possibly a_1, according as q is even or odd. When q is odd, the parity of b_k equals the parity of p. The expression is unique, provided that a_k and b_k do not have different signs when $b_k = \pm 1$.*

There is a very well-known connection between continued fraction decompositions and $SL_2(\mathbb{Z})$. Define

$$U = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad L = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Lemma 14.3 *If*

$$U^{a_1} L^{b_1} \cdots U^{a_k} L^{b_k} = \begin{pmatrix} q & s \\ p & r \end{pmatrix}$$

then

$$q/p = [a_1, b_1, \ldots, a_k, b_k], \; s/r = [a_1, b_1, \ldots, a_k],$$
$$q/s = [b_k, a_k, \ldots, b_1, a_1] \quad \text{and} \quad p/r = [b_k, a_k, \ldots, b_1].$$

The first two equalities can be proven by a straightforward induction, and the last two follow by taking transposes.

Proposition 14.4 Let $\mu = \{\lambda, \rho\}$ be a pair of disks in H and let σ and τ be two nonseparating slope disks for μ. Write the $(\mu; \tau)$–slope of σ as

$$q/p = [2a_1, 2b_1, 2a_2, \ldots, 2b_{n-1}, 2a_n, b_n].$$

If we regard slope pairs $[r, s]$ as column vectors $\binom{s}{r}$, then the change-of-basis matrix from $(\mu; \tau)$–slopes to $(\mu; \sigma)$–slopes is

$$U^{2a_1} L^{2b_1} U^{2a_2} \cdots U^{2a_n} L^{b_n} U^{-(-1)^p 2a},$$

where $a = \sum a_i$.

Proof Referring to the picture of H in Figure 6, let u be the Dehn twist of H about τ that sends an object with slope pair $[p, q]$ to one with slope pair $[p, q + 2p]$ (u is a "left-handed" Dehn twist about τ). Similarly, let ℓ be the homeomorphism of H that preserves λ and ρ and sends an object with slope pair $[p, q]$ to one with slope pair $[p + q, q]$ (a half-twist of H about τ^0, whose effect looks like the restriction of β from Section 9, when H is viewed as in Figure 6).

Regarding slope pairs $[r, s]$ as column vectors $\binom{s}{r}$, the effects of u and ℓ are multiplication by U^2 and L respectively. That is, for the $(\mu; \tau)$–slopes determined by the basis $\{\tau, \tau^0\}$, these are the matrices of u and ℓ.

Since σ is nonseparating, we can use Lemma 14.2 to write its $(\mu; \tau)$–slope in the form $[2a_1, 2b_1, \ldots, 2b_{n-1}, 2a_n, b_n]$, where b_n has the parity of p and all terms except possibly $2a_1$ are nonzero. The composition $u^{a_1} \ell^{2b_1} \cdots u^{a_n} \ell^{b_n}$ of ℓ has matrix $U^{2a_1} L^{2b_1} \cdots U^{2a_n} L^{b_n}$. Since τ has slope pair $[0, 1]$, ie $\binom{1}{0}$, the "q/p" case of Lemma 14.3 shows that $u^{a_1} \ell^{2b_1} \cdots u^{a_n} \ell^{b_n}$ sends τ to σ. It takes τ^0 to a perpendicular disk for σ, but not necessarily σ^0. Since u preserves τ, the homeomorphism $u^{a_1} \ell^{2b_1} \cdots u^{a_n} \ell^{b_n} u^{-(-1)^p a}$, where $a = \sum a_i$, also takes τ to σ. In the remainder of the proof, we will show that it takes τ^0 to σ^0 as well, and the lemma follows.

Consider any separating slope disk for μ, and let c_1 and c_2 be the core circles of its complementary solid tori in H. Each c_i has intersection number 0 with τ^0, but applying ℓ reverses the sides of ρ, so reverses the orientation of exactly one of the c_i. It follows that $\mathrm{lk}(\ell(c_1), \ell(c_2)) = -\mathrm{lk}(c_1, c_2)$. On the other hand, each c_i has intersection number ± 1 with τ, since c_i has intersection number ± 1 with one of λ or ρ and 0 with the other. Therefore $\mathrm{lk}(u(c_1), u(c_2)) = \mathrm{lk}(c_1, c_2) \pm 1$, the sign depending on conventions.

In particular, if C_1 and C_2 are the core circles for the complementary components of τ^0, we have

$$\mathrm{lk}(\ell^{b_n} u^{-(-1)^p a}(C_1), \ell^{b_n} u^{-(-1)^p a}(C_2)) = \mathrm{lk}(u^{-a}(C_1), u^{-a}(C_2))$$

since b_n and p have the same parity. The remaining ℓ^{2b_i} do not change linking numbers, and $a = \sum a_i$, so

$$\mathrm{lk}(u^{a_1} \ell^{2b_1} \cdots u^{a_n} \ell^{b_n} u^{-(-1)^p a}(C_1), u^{a_1} \ell^{2b_1} \cdots u^{a_n} \ell^{b_n} u^{-(-1)^p a}(C_2)) = 0 .$$

These are core circles of the complementary tori of $u^{a_1} \ell^{2b_1} \cdots u^{a_n} \ell^{b_n} u^{-(-1)^p a}(\tau^0)$. Since this disk intersects $u^{a_1} \ell^{2b_1} \cdots u^{a_n} \ell^{b_n} u^{-(-1)^p a}(\tau) = \sigma$ in a single arc, it must be σ^0. □

Theorem 14.5 If $q/p = [2a_1, 2b_1, 2a_2, \ldots, 2b_{n-1}, 2a_n, b_n]$ is one of the Scharlemann–Thompson invariant or the principal slope of a tunnel, then the other one is $[(-1)^p 2a, -b_n, -2a_n, \ldots, -2a_2, -2b_1]$, where $a = \sum a_i$. In particular, if one is a (necessarily odd) integer, then the other is the negative of that integer.

Proof From Lemma 14.1, the invariants are related by the fact that for some pair of slope disks σ and τ for a meridian pair μ, one invariant is the $(\mu; \sigma)$–slope of τ, and the other is the $(\mu; \tau)$–slope of σ. Using Proposition 14.4, the change-of-basis matrix from $(\mu; \sigma)$–slopes to $(\mu; \tau)$–slopes is $U^{(-1)^p 2a} L^{-b_n} U^{-2a_n} \cdots U^{-2a_2} L^{-2b_1} U^{-2a_1}$. The first column of this matrix gives the $(\mu; \tau)$–slope of σ. By the "q/p" case of Lemma 14.3, with $b_k = 0$, it is $[(-1)^p 2a, -b_n, -2a_n, \ldots, -2a_2, -2b_1]$. □

Corollary 14.6 If q/p is either the Scharlemann–Thompson invariant or the principal slope of a tunnel, then the other one is of the form q'/p where $qq' \equiv -1 \pmod{p}$.

Proof If the change-of-basis matrix in the proof of Theorem 14.5 has the form

$$\begin{pmatrix} q & s \\ p & r \end{pmatrix},$$

then its inverse is

$$\begin{pmatrix} r & -s \\ -p & q \end{pmatrix},$$

so the invariants are q/p and $-r/p$ where $qr - ps = 1$. □

We have implemented the formula of Theorem 14.5 to convert between the invariants computationally [7]. Some sample calculations are:

```
STinvariant> convert 55
-55

STinvariant> convert (59/35)
-299/35

STinvariant> convert (-299/35)
59/35

STinvariant> convertRange 100102 17255 17265
17255/100102, -2843767/100102
17257/100102, -6541753/100102
17259/100102, 345051565/100102
17261/100102, 5593835/100102
17263/100102, 1775313/100102
17265/100102, 158447/100102
```

The last command produces the corresponding pairs of invariants containing each $q/100102$ for odd q with $17255 \le q \le 17265$.

15 Tunnels of two-bridge knots

It is known from work of Kobayashi [17; 18], Morimoto and Sakuma [22] and Uchida [27] that a 2–bridge knot has at most four equivalence classes of tunnels (not six, for us, since we are considering tunnels only up to equivalence, rather than up to isotopy). Two of these are the upper and lower simple tunnels. In this section, we will locate the other tunnels in $\mathcal{D}(H)/\mathcal{G}$ and compute their slope parameters.
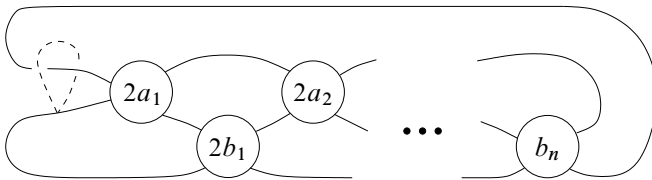

Figure 13

From the above references and standard repositioning of 2–bridge knots by isotopy, each tunnel of a 2–bridge knot either is simple or is equivalent to one like that shown in Figure 13, where each circle indicates a block of some nonzero number of half-twists.

Each of the blocks in the middle row has an even number $2a_i$ of half-twists, and those on the bottom row have an even number $2b_i$ of half-twists, except that the last one has a number b_n that may be odd. Our convention is that a_i is positive for left-hand twists, and b_i is positive for right-hand twists.

There is a well-known classification of 2–bridge knots based on continued fraction expansions of a rational parameter b/a with b odd (the case of b even gives 2–bridge links). One description of the invariant is that the 2–fold branched cover of S^3 over the knot is $L(b, a)$, but we will describe it here in a way that is more suited to our purposes.

Given b/a, change a by multiples of b until $|b/a| > 1$. Either of two possible values of a may be used. Expand b/a as a continued fraction $[2a_1, 2b_1, 2a_2, 2b_2, \ldots, 2a_n, b_n]$ as in Lemma 14.2. Additionally, if $b_n = \pm 1$, adjust a_n and b_n so that they have the same sign. Under these conditions, the expansion of b/a is uniquely determined, and the corresponding 2–bridge knot is the one shown in Figure 13.

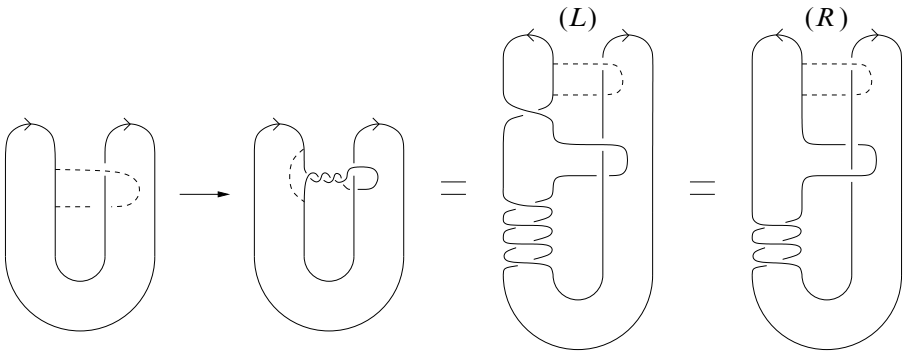Figure 14 shows the type of cabling construction used to produce the tunnel in Figure 13.

Figure 14

It is described by a nonzero parameter k that tells the number of right-hand twists of the two horizontal arcs from the original tunnel (the case $k = 0$ would produce a cabling with infinite slope, ie not a cabling construction). The cabling shown in Figure 14 has $k = -4$. As indicated in that figure, a knot and tunnel arc resulting from such a cabling can be moved by isotopy so that the full twist of the middle two strands is either left-handed (configuration (L)) or right-handed (configuration (R)), then repositioned so that the tunnel arc has the same appearance as the original one. As shown in Figure 14, this will produce either k half-twists below the full twist and -1 half-twists above, as in configuration (L), or $k + 1$ half-twists below the full twist, as in configuration (R).

Starting from the right-hand end of Figure 13, we perform a sequence of these cablings, one for each of the full twists of the middle two strands. Thus, the total number of cablings is $\sum |a_i|$. At each step, the value of k in the cablings must be selected to produce the correct number $2b_i$ (or b_n) of half-twists, as we will detail below. The condition that b_n has the same sign as a_n, when $b_n = \pm 1$, ensures that the first cabling produces a nontrivial knot. By the Unique Cabling Sequence Theorem 13.2, this sequence of cablings is the unique sequence producing this tunnel.

We will now calculate the slopes of these cablings. The calculation depends on the parity, at the time a given cabling is to be performed, of the number of crossings of the left-hand two strands that lie below the tunnel when it is positioned in Figure 14. In the example of Figure 14, the parity is odd both in configuration (L) and in configuration (R). When the parity is even, either of the two orientations of the knot orients the middle strands so that near the tunnel one is upward and the other is downward, as occurs in the trivial knot, but when the parity is odd, both are upward or both are downward.
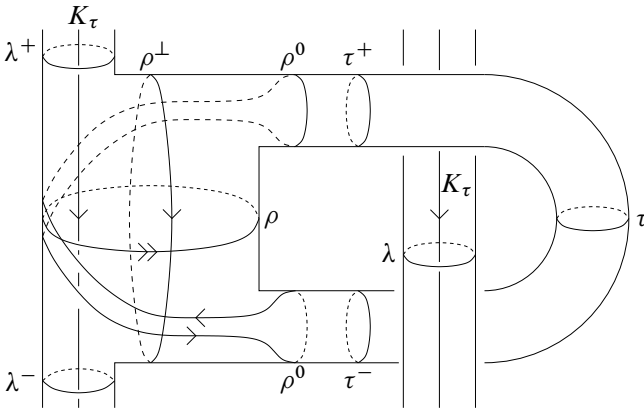

Figure 15: The zero-slope disk for the case of odd parity (compare with Figure 6)

The parity affects which disk will be the zero-slope disk in the calculation of the slope coefficients. Figure 15 shows the case when the parity is odd. The disk ρ is the one that is replaced by the cabling construction. The zero-slope disk ρ^0 is obtained from $\rho^\perp$ by a left-hand Dehn twist about ρ. For this ρ^0, a K_λ that is disjoint from ρ^0 has linking number 0 with the K_τ that is disjoint from ρ^0. In the case of even parity, ρ^0 is obtained from $\rho^\perp$ by a right-hand Dehn twist.

To obtain the actual slope coefficient for a given cabling of the type shown in Figure 14, one may draw a cabling arc for the cabling and laboriously calculate its slope using ρ^0 as the zero-slope disk, but there is a quick way to find it. If the disk $\rho^\perp$ shown

in Figure 15 were used in place of ρ^0 in calculating the slope, then the slope pairs would be $[k, 1]$. The Dehn twist about ρ that moves ρ^0 to $\rho^{\perp}$ moves the cabling arc to one whose slope pair using $\rho^{\perp}$ as the zero-slope disk is $[k, 1 - 2k]$ (see the proof of Proposition 14.4, where this Dehn twist would be denoted by u^{-1}), so this is the slope pair of the original cabling arc using ρ^0. Consequently the $(\{\lambda, \tau\}; \rho)$–slope of the cabling arc is $-2 + 1/k$. In the even parity case, the only difference is that the twist moving ρ^0 to $\rho^{\perp}$ is in the opposite sense, changing $[k, 1]$ to $[k, 1 + 2k]$ and producing the slope $2 + 1/k$.

Consider the initial cabling construction. Let k be the number of right-hand half twists in the cabling, so that the slope is $2 + 1/k$.

Suppose first that $a_n > 0$. According to configuration (L) of Figure 14, $b_n = k$. The parity of the trivial knot is even, to the slope pair of the cabling is $[k, 2k + 1] = [b_n, 2b_n + 1]$, giving $m_0 = [b_n/(2b_n + 1)]$.

Suppose now that $a_n < 0$. From configuration (R) of Figure 14, we see that $b_n = k + 1$. The slope pair of the cabling is $[k, 1 + 2k] = [b_n - 1, 2b_n - 1]$, giving $m_0 = [(b_n - 1)/(2b_n - 1)]$.

For the cablings beyond the first, it will be convenient to rewrite the continued fraction by expanding each $2a_i$ to $[2, 0, 2, 0, \ldots, 2, 0, 2]$ or $[-2, 0, -2, 0, \ldots, -2, 0, -2]$, thereby assuming that each $a_i = \pm 2$ and allowing some $b_i = 0$.

Consider the cabling that produces the full twist of the middle two strands corresponding to a_i. The cabling that produces the twist corresponding to a_{i+1} has just been completed. Suppose first that $a_{i+1} > 0$. As in configuration (L) of Figure 14, there is already a left-hand half twist in the left two strands. Since all $2b_j$ are even, the parity is the opposite of the parity of b_n. Again referring to Figure 14, we see that to end up with exactly b_i half-twists of the left two strands, we need to use $k = b_i + 1$ if $a_i > 0$, and $k = b_i$ if $a_i < 0$.

Suppose now that $a_{i+1} < 0$. From configuration (R) of Figure 14, the parity is just equal to that of b_n. To achieve $2b_i$ half-twists after the cabling, we need $k = 2b_i$ if $a_i > 0$, and $k = 2b_i - 1$ if $a_i < 0$.

We now have a complete algorithm to determine the values k_i in the cablings and the cabling slopes: Write b/a as $[2a_1, 2b_1, \ldots, 2a_n, b_n]$ where:

(1) Each $a_i = \pm 1$, and some b_i other than b_n may be 0.

(2) If $b_n = \pm 1$, then a_n and b_n have the same sign.

For the first cabling:

(1) If $a_n = 1$, then $k_n = b_n$ and $m_0 = [b_n/(2b_n + 1)]$.

(2) If $a_n = -1$, then $k_n = b_n - 1$ and $m_0 = [(b_n - 1)/(2b_n - 1)]$.

For the remaining cablings, the slope m_i is $2 + 1/k_i$ or $-2 + 1/k_i$ according as the parity is even or odd. The parity and the value of k_i are computed as follows:

(1) If $a_{i+1} = 1$, then:
   (a) The parity equals the parity of $b_n + 1$.
   (b) If $a_i = 1$, then $k_i = 2b_i + 1$, and if $a_i = -1$, then $k_i = 2b_i$.

(2) If $a_{i+1} = -1$, then:
   (a) The parity equals the parity of b_n.
   (b) If $a_i = 1$, then $k_i = 2b_i$, and if $a_i = -1$, then $k_i = 2b_i - 1$.

We have implemented the algorithm computationally [7]. Some sample calculations are:

```
TwoBridge> slopes (33/19)
[ 1/3 ], 3, 5/3

TwoBridge> slopes (64793/31710)
[ 2/3 ], -3/2, 3, 3, 3, 3, 3, 7/3, 3, 3, 3, 3, 49/24

TwoBridge> slopes (3860981/2689048)
[ 13/27 ], 3, 3, 3, 5/3, 3, 7/3, 15/8, -5/3, -1, -3

TwoBridge> slopes (5272967/2616517)
[ 5/9 ], 11/5, 21/10, -23/11, -131/66
```

Of course, the slope parameters that we have calculated are the parameters m_i that appear in the Parameterization Theorem 12.3. The parameters s_j are all 0, since these are semisimple tunnels (the primitive disk called λ in Figure 15 is retained in every cabling construction).

16 Tunnels of links

A quick summary of how the theory adapts to include tunnels of tunnel number 1 links is that one just adds the separating disks as possible slope disks. The cabling sequence ends with the first separating slope disk and cannot be continued. The Parameterization Theorem 12.3 holds as stated, except allowing q_n to be even.

In a bit more detail, one way to allow links is to include separating disks in the theory from the start, that is, to use the full disk complex $\mathcal{K}(H)$ rather than the nonseparating disk complex $\mathcal{D}(H)$. Very little additional complication actually occurs. A separating disk E in H is disjoint from only two other disks, both nonseparating, so is a vertex of only one 2–simplex $\langle E, \tau_1, \tau_2 \rangle$ attached to $\mathcal{D}(H)$ along $\langle \tau_1, \tau_2 \rangle$. Each 1–simplex

$\langle \tau_1, \tau_2 \rangle$ of $\mathcal{D}(H)$ is a face of countably many such 2–simplices, one for each separating slope disk of $\{\tau_1, \tau_2\}$. In slope coordinates, these disks correspond to the q/p with q even.

The link of E in the first barycentric subdivision $\mathcal{K}'(H)$ consists of two 1–simplices meeting in the principal vertex $\{E, \tau_1, \tau_2\}$ of E. The spine of $\mathcal{K}(H)$ is obtained from $\widetilde{\mathcal{T}}$ simply by adding a "Y" in each 2–simplex $\langle E, \tau_1, \tau_2 \rangle$, which meets $\widetilde{\mathcal{T}}$ only in the vertex $\{\tau_1, \tau_2\}$.

A *primitive* separating disk is a disk in H that is contained in a splitting sphere of H. Note that both of the nonseparating disks disjoint from a primitive separating disk are primitive.

To obtain $\mathcal{K}(H)/\mathcal{G}$ from $\mathcal{D}(H)/\mathcal{G}$, we first add one half-simplex to the primitive region. It meets the primitive simplex Π along the edge called $\langle \mu_0, \pi_0 \rangle$ in Section 10, and its third vertex is the unique orbit of primitive separating disks. Next, a half-simplex for each simple separating disk is added along $\langle \mu_0, \pi_0 \rangle$. These correspond to the $[p/q] \in \mathbb{Q}/\mathbb{Z}$ with q even and are the upper and lower tunnels of 2–bridge links. The remaining added 2–simplices are attached along the other 1–simplices of $\mathcal{D}(H)/\mathcal{G}$ as they were in $\mathcal{D}(H)$. A tunnel of a link has a principal path from the primitive nonseparating triple θ_0 to its principal vertex, which is the only white vertex in the link of the tunnel.

The trivial link is the link associated to the orbit of primitive separating disks. It arises from the tunnel of the trivial knot by a cabling construction of simple slope $[1/0] = \infty$. This is the only case in which ∞ is an allowable slope parameter. With this convention, we can state the general Parameterization Theorem:

Theorem 16.1 *Let τ be a knot or link tunnel with principal path θ_0, μ_0, $\mu_0 \cup \{\tau_0\}$, $\mu_1, \ldots, \mu_n$, $\mu_n \cup \{\tau_n\}$. Fix a lift of the principal path to $\mathcal{K}(H)$, so that each μ_i corresponds to an actual pair of disks in H.*

(1) *If τ is primitive, put $m_0 = [0] \in \mathbb{Q}/\mathbb{Z}$ or $m_0 = [1/0] = \infty$, according as τ is the tunnel of the trivial knot or the trivial link. Otherwise, let $m_0 = [p_0/q_0] \in \mathbb{Q}/\mathbb{Z}$ be the simple slope of τ_0.*

(2) *If $n \geq 1$, then for $1 \leq i \leq n$ let σ_i be the unique disk in $\mu_{i-1} - \mu_i$ and let $m_i = q_i/p_i \in \mathbb{Q}$ be the $(\mu_i; \sigma_i)$–slope of τ_i.*

(3) *If $n \geq 2$, then for $2 \leq i \leq n$ define $s_i = 0$ or $s_i = 1$ according to whether or not the unique disk of $\mu_{i-1} \cap \mu_i$ equals the unique disk of $\mu_{i-1} \cap \mu_{i-2}$.*

Then, sending τ to the pair $((m_0, \ldots, m_n), (s_2, \ldots, s_n))$ is a bijection from the set of all tunnels of all tunnel number 1 knots and links to the set of all elements $(([p_0/q_0], q_1/p_1, \ldots, q_n/p_n), (s_2, \ldots, s_n))$ in

$$\left(\mathbb{Q}/\mathbb{Z} \cup \{\infty\}\right) \cup \left(\mathbb{Q}/\mathbb{Z} \times \mathbb{Q}\right) \cup \left(\bigcup_{n \geq 2} \mathbb{Q}/\mathbb{Z} \times \mathbb{Q}^n \times C_2^{n-1}\right)$$

with all q_i odd except possibly q_n. The tunnel is a tunnel of a knot or a link according as q_n is odd or even.

The linking number of the two components of a tunnel number 1 link, up to sign, is half the numerator q_n of the principal slope q_n/p_n of σ (or half the denominator of the simple slope, if the tunnel is simple). This is immediate from the construction of general slope coordinates in Section 8.

Theorem 16.1 implies that a tunnel is almost never equivalent to itself by an orientation-reversing equivalence:

Theorem 16.2 *Let τ be a tunnel of a tunnel number 1 knot or link. Suppose that τ is equivalent to itself by an orientation-reversing equivalence. Then τ is the tunnel of the trivial knot, the trivial link, or the Hopf link.*

Proof As noted in Remark 12.5, the classification of tunnels up to arbitrary homeomorphism of S^3 is obtained from that of the Parameterization Theorem 12.3 by adding the equivalence $([m_0], m_1, \ldots, m_n) \sim ([-m_0], -m_1, \ldots, -m_n)$. The only tuples equal to themselves under this move are $([0])$, (∞) and $([1/2])$, which correspond to the trivial knot, the trivial link and the Hopf link. $\square$

The tunnel of the Hopf link will be examined more closely in Section 17 below.

The Unique Cabling Sequence Theorem 13.2 holds as stated for links as well as knots.

As in Proposition 11.1, the simple tunnels of links are exactly the upper and lower tunnels of 2–bridge links, and the statement of Proposition 11.2 holds allowing q even. Concerning these tunnels, we can give a quick proof of a theorem of C Adams and A Reid [1] and M Kuhn [19]:

Theorem 16.3 (Adams–Reid, Kuhn) *The only tunnels of a 2–bridge link are its upper and lower tunnels.*

Proof Since each component of a 2–bridge link is unknotted, the tunnel disk is disjoint from a primitive pair, hence is simple. $\square$

We can also understand semisimple tunnels of links, that is, tunnels whose cocore disk is disjoint from a primitive disk, but not from a primitive pair. In this case, one of the components of the link is unknotted. Such links are the topic of the following theorem, which slightly strengthens a result of T Harikae [14]:

Theorem 16.4 *Let L be a nontrivial tunnel number 1 link with an unknotted component. Then the other component of L is a $(1, 1)$–knot. Moreover, every tunnel of L is simple or semisimple, and L has torus bridge number 2.*

Proof Let σ be a tunnel of L. Then σ is disjoint from exactly two nonseparating disks, τ_1 and τ_2, and one of them, say τ_1, must be primitive. Therefore τ_2 is simple or semisimple, and K_{τ_2} is a $(1, 1)$–knot.

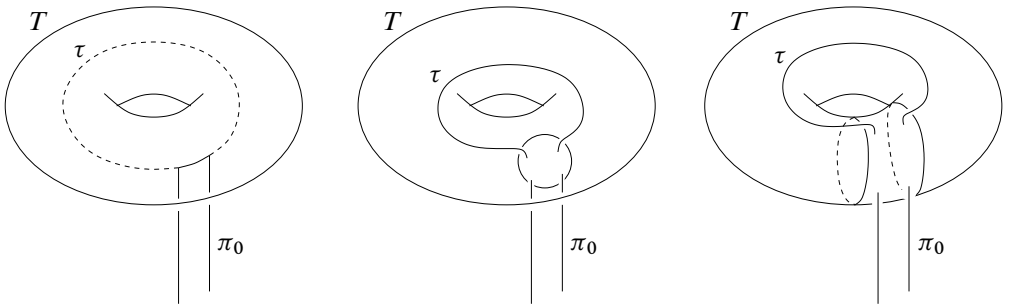To prove that L has torus bridge number 2, we refer to Figure 16. The left drawing

Figure 16

shows a torus level T and the tunnel τ before the cabling that produces σ, and the middle picture shows schematically the result of the cabling. The drawing on the right shows an isotopic repositioning of L_σ. The τ–arc is pushed slightly outside of T, and the sphere for the cabling is expanded to the union of an annulus in T and two meridian disks. The cabling arcs may be moved off of the meridian disks by isotopy, to lie in A. From there, they can be pushed slightly inside T. Then, L meets each of the complementary solid tori of T in a trivial pair of arcs. □

17 The Hopf link

In Theorem 16.2, we saw that the tunnel of the Hopf link is the only tunnel of a nontrivial knot or link that is preserved by an orientation-reversing equivalence. In this section, we examine this equivalence more closely, obtaining a version of Futer's Theorem 11.5 for links.

First, we work out the vertex stabilizers of the action of $\mathcal{G}_{\pm}$ on $\mathcal{K}(H)$. As usual, fix a primitive pair $\mu_0 = \{\lambda_0, \rho_0\}$. To obtain a generating set for the stabilizer $(\mathcal{G}_{\pm})_{\mu_0}$, we add to the generators α, β and γ of $\mathcal{G}_{\mu_0}$ an orientation-reversing involution R. Recalling the description of α, β and γ in Section 9, R is a reflection through the plane of the page in Figure 6; it preserves λ_0 and ρ_0, reflecting each across a diameter. It commutes with α and γ, and conjugates β to β^{-1}.

In the description of slope coordinates in Section 8, R lifts to the involution of $\widetilde{\Sigma}$ reflecting across the vertical line bisecting the fundamental domain shown in Figure 7. This shows R sends a simple disk of slope pair $[p, q]$ to one with slope pair $[-p, q]$;

We have the following version of Proposition 9.1:

Proposition 17.1 *The stabilizer $(\mathcal{G}_{\pm})_{\mu_0}$ is the subgroup generated by α, β, γ and R. In fact, $(\mathcal{G}_{\pm})_{\mu_0}$ is the semidirect product $(C_2 \times \mathbb{Z}) \circ (C_2 \times C_2)$, where $\langle \alpha, \beta \rangle$ is the normal subgroup $C_2 \times \mathbb{Z}$, $\langle \gamma, R \rangle$ is the subgroup $C_2 \times C_2$, α is central, $\gamma \beta \gamma^{-1} = \alpha\beta$ and $R\beta R^{-1} = \beta^{-1}$.*

Proof We have $R \in \mathcal{G}_{\pm} - \mathcal{G}$, and R stabilizes the primitive pair μ_0, so the subgroup S generated by α, β, γ and R properly contains $\mathcal{G}_{\mu_0}$ and is contained in $(\mathcal{G}_{\pm})_{\mu_0}$. But $\mathcal{G}$ has index 2 in $\mathcal{G}_{\pm}$, so $\mathcal{G}_{\mu_0}$ has index at most 2 in $(\mathcal{G}_{\pm})_{\mu_0}$. Therefore S must equal $(\mathcal{G}_{\pm})_{\mu_0}$. Since R is an involution, $(\mathcal{G}_{\pm})_{\mu_0}$ is the semidirect product $\mathcal{G}_{\mu_0} \circ \langle R \rangle$, and using Proposition 9.1 and the relations $R\alpha R^{-1} = \alpha$, $R\beta R^{-1} = \beta^{-1}$ and $R\gamma R^{-1} = \gamma$ gives our description of $(\mathcal{G}_{\pm})_{\mu_0}$. □

Proposition 17.2 *The stabilizer in $\mathcal{G}_{\pm}$ of a possibly separating nonprimitive disk E in H is as follows:*

(i) *If E is simple with simple slope $[1/2]$, then its stabilizer is conjugate to the dihedral subgroup of order 8 generated by the involutions βR and γ.*

(ii) *If E is simple with simple slope not $[1/2]$, then its stabilizer is conjugate to the subgroup $C_2 \times C_2$ generated by α and γ.*

(iii) *Otherwise, the stabilizer is C_2, generated by α.*

Proof Assume first that E is simple with simple slope $[1/2]$. Any element of $\mathcal{G}_{\pm}$ that preserves E must also preserve the unique pair of primitive disks that are disjoint from E. Conjugating in $\mathcal{G}_{\pm}$, we may assume that this pair is μ_0.

Conjugating further by a power of β, we may assume that E has slope pair $[1, 2]$ with respect to μ_0. This slope pair is preserved by βR, γ and α. On the other hand, using

the action of the four generators of $(\mathcal{G}_\pm)_{\mu_0}$ on slope pairs, any word that stabilizes E can be written as $\alpha^{\epsilon_1}\gamma^{\epsilon_2}(\beta R)^{\epsilon_3}$ where the $\epsilon_i \in \{0, 1\}$. Since $(\gamma \beta R)^2 = \alpha$, the stabilizer is dihedral as in (i). (Figure 17 illustrates the effect of βR.)

If E is simple but does not have simple slope $[1/2]$, then R does not preserve its simple slope. So any element $\alpha^{\epsilon_1}\beta^n\gamma^{\epsilon_2}R^{\epsilon_3}$ stabilizing E must have $\epsilon_3 = 0$. As in Proposition 10.1(i), part (ii) follows. Part (iii) is proven as in Proposition 10.1(ii); note that the disk called τ_0 there must be nonseparating, since otherwise the path could not continue on to θ. □
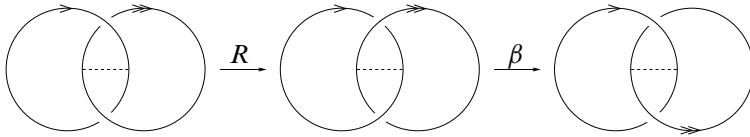

Figure 17: An orientation-reversing automorphism of the Hopf link and its tunnel arc

Arguing as in Corollary 10.2, we have:

Corollary 17.3 Let $\langle \tau_1, \tau_2, \tau_3 \rangle$ be a 2–simplex of $\mathcal{D}(H)$, with τ_3 nonprimitive. Suppose that $\langle \tau_1, \tau_2, \tau_3 \rangle$ is stabilized by an element h of $\mathcal{G}_\pm$ other than the identity or α. Then τ_3 is simple, and τ_1 and τ_2 are primitive.

Using Corollary 17.3 and Proposition 17.2, we adapt Corollary 11.3 to links:

Corollary 17.4 Let W be the union of a nontrivial tunnel number 1 link L and a tunnel arc for L, and suppose that h is a homeomorphism of S^3 that preserves W. Then either

(i) h is isotopic preserving W to a homeomorphism which is the identity or the hyperelliptic involution on a neighborhood of W, or

(ii) L is a 2–bridge link.

Moreover, if h is orientation-reversing, then L is the Hopf link.

Using these results, we obtain a version of Futer's Theorem 11.5 for links:

Theorem 17.5 Let A be a tunnel arc for a nontrivial link $L \subset S^3$. Then:

(a) There exists an orientation-preserving homeomorphism of S^3 that preserves $L \cup A$ and interchanges the components of L if and only if L is a two-bridge link.

(b) There exists an orientation-reversing homeomorphism of S^3 that preserves $L \cup A$ if and only if L is the Hopf link.

Proof   For a 2–bridge link, γ interchanges the components and preserves the tunnel arc up to isotopy, so Corollary 17.4 gives part (a). Part (b) follows easily from Proposition 17.2. □

References

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

[2] **E Akbas**, *A presentation for the automorphisms of the 3-sphere that preserve a genus two Heegaard splitting*, Pacific J. Math. 236 (2008) 201–222   MR2407105

[3] **S Cho**, *Homeomorphisms of the 3–sphere that preserve a Heegaard splitting of genus two*, Proc. Amer. Math. Soc. 136 (2008) 1113–1123   MR2361888

[4] **S Cho**, **D McCullough**, *Constructing knot tunnels using giant steps*   arXiv: 0812.1382

[5] **S Cho**, **D McCullough**, *Tunnel leveling, depth, and bridge numbers*   arXiv: 0812.1396

[6] **S Cho**, **D McCullough**, *Semisimple tunnels*, in preparation

[7] **S Cho**, **D McCullough**, *Software*   Available at   http://www.math.ou.edu/~dmccullough/research/software.html

[8] **S Cho**, **D McCullough**, *Cabling sequences of tunnels of torus knots*, Algebr. Geom. Topol. 9 (2009) 1–20

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

[10] **H Goda**, **C Hayashi**, *Genus two Heegaard splittings of exteriors of* 1*–genus* 1*–bridge knots*, preprint

[11] **H Goda**, **M Scharlemann**, **A Thompson**, *Levelling an unknotting tunnel*, Geom. Topol. 4 (2000) 243–275   MR1778174

[12] **L Goeritz**, *Die Abbildungen der Brezelfläche und der Volbrezel vom Gesschlect* 2, Abh. Math. Sem. Univ. Hamburg 9 (1933) 244–259

[13] **C M Gordon**, *On primitive sets of loops in the boundary of a handlebody*, Topology Appl. 27 (1987) 285–299   MR918538

[14] **T Harikae**, *On the triviality of bouquets and tunnel number one links*, from: "Proceedings of the Workshop on Graph Theory and Related Topics (Sendai, 1999)", volume 7 (2001) 1–3   MR1874572

[15] **J Johnson**, *Bridge number and the curve complex*   arXiv:math.GT/0603102

[16] **J Johnson**, **A Thompson**, *On tunnel number one knots which are not* $(1, n)$ arXiv: J. Knot Theory Ramifications

[17] **T Kobayashi**, *A criterion for detecting inequivalent tunnels for a knot*, Math. Proc. Cambridge Philos. Soc. 107 (1990) 483–491 MR1041480

[18] **T Kobayashi**, *Classification of unknotting tunnels for two bridge knots*, from: "Proceedings of the Kirbyfest (Berkeley, CA, 1998)", (J Hass, M Scharlemann, editors), Geom. Topol. Monogr. 2 (1999) 259–290 MR1734412

[19] **M Kuhn**, *Tunnels of* $2-bridge$ *links*, J. Knot Theory Ramifications 5 (1996) 167–171 MR1395777

[20] **D McCullough**, *Virtually geometrically finite mapping class groups of* $3-manifolds$, J. Differential Geom. 33 (1991) 1–65 MR1085134

[21] **Y N Minsky**, **Y Moriah**, **S Schleimer**, *High distance knots*, Algebr. Geom. Topol. 7 (2007) 1471–1483 MR2366166

[22] **K Morimoto**, **M Sakuma**, *On unknotting tunnels for knots*, Math. Ann. 289 (1991) 143–167 MR1087243

[23] **T Saito**, *Scharlemann–Thompson invariant for knots with unknotting tunnels and the distance of* $(1, 1)-splittings$, J. London Math. Soc. (2) 71 (2005) 801–816 MR2132384

[24] **M Scharlemann**, *Automorphisms of the* $3-sphere$ *that preserve a genus two Heegaard splitting*, Bol. Soc. Mat. Mexicana (3) 10 (2004) 503–514 MR2199366

[25] **M Scharlemann**, *There are no unexpected tunnel number one knots of genus one*, Trans. Amer. Math. Soc. 356 (2004) 1385–1442 MR2034312

[26] **M Scharlemann**, **A Thompson**, *Unknotting tunnels and Seifert surfaces*, Proc. London Math. Soc. (3) 87 (2003) 523–544 MR1990938

[27] **Y Uchida**, *Detecting inequivalence of some unknotting tunnels for two-bridge knots*, from: "Algebra and topology 1990 (Taejon, 1990)", (S H Bae, G T Jin, editors), Korea Adv. Inst. Sci. Tech., Taejŏn (1990) 227–232 MR1098731

Department of Mathematics, University of California at Riverside
Riverside, CA 92521, USA

Department of Mathematics, University of Oklahoma
Norman, Oklahoma 73019, USA

scho@math.ucr.edu, dmccullough@math.ou.edu

www.math.ou.edu/~dmccullough/