UNKNOTTING TUNNELS AND SEIFERT SURFACES

MARTIN SCHARLEMANN AND ABIGAIL THOMPSON

ABSTRACT. Let K be a knot with an unknotting tunnel γ and suppose that K is not a 2-bridge knot. There is an invariant $\rho = p/q \in \mathbb{Q}/2\mathbb{Z}$, p odd, defined for the pair (K,γ).

The invariant ρ has interesting geometric properties: It is often straightforward to calculate; e. g. for K a torus knot and γ an annulus-spanning arc, $\rho(K,\gamma) = 1$. Although ρ is defined abstractly, it is naturally revealed when $K \cup \gamma$ is put in thin position. If $\rho \neq 1$ then there is a minimal genus Seifert surface F for K such that the tunnel γ can be slid and isotoped to lie on F. One consequence: if $\rho(K,\gamma) \neq 1$ then $genus(K) > 1$. This confirms a conjecture of Goda and Teragaito for pairs (K,γ) with $\rho(K,\gamma) \neq 1$.

1. INTRODUCTORY COMMENTS

In [GST] the following conjecture of Morimoto's was established: if a knot $K \subset S^3$ has a single unknotting tunnel γ, then γ can be moved to be level with respect to the natural height function on K given by a minimal bridge presentation of K. The repeated theme of the proof is that by "thinning" the 1-complex $K \cup \gamma$ one can simplify its presentation until the tunnel is either a level arc or a level circuit.

The present paper was originally motivated by two questions. One was a rather specialized conjecture of Goda and Teragaito: must a hyperbolic knot which has both genus and tunnel number one necessarily be a 2-bridge knot? A second question was this: Once the thinning process used in the proof of [GST] stops because the tunnel becomes level, can thin position arguments still tell us more?

With respect to the second question, it turns out that there is an obstruction to further useful motion of γ that can be expressed as an element $\rho \in \frac{Q}{2Z}$. Surprisingly, further investigation showed that, so long as K is *not* 2-bridge, the obstruction ρ can be defined in a way completely independent of thin position and thereby can be viewed as an invariant of the pair (K,γ). Moreover, this apparently new invariant has useful properties: It is not hard to calculate. If $\rho \neq 1$, then the tunnel can be isotoped onto a minimal genus

Date: August 12, 2002.

Research supported in part by grants from the National Science Foundation. The second author also gratefully acknowledges the support of the von Neumann Fund and the Weyl Fund through the Institute for Advanced Study.

Seifert surface. This, in combination with some work [EU] of Eudave-Munoz and Uchida, verifies a conjecture of Goda and Teragaito [GT] in the case in which $\rho(K,\gamma) \neq 1$. If $\rho = 1$ and the tunnel is a level edge, then the tunnel can be moved so that instead of connecting two maxima (say) of K, it connects two minima and vice versa. A future paper [Sc2] will expand on this observation, addressing the technically difficult and rather specialized case in which $\rho = 1$ with the goal of verifying the Goda-Teragaito conjecture in this final case.

Here are a few technical notes on conventions and notation used in the arguments that follow:

1. For $X \subset M$ a polyhedron, $\eta(X)$ will denote a closed regular neighborhood, whereas (abusing notation slightly) $M - \eta(X)$ will mean the closed complement of $\eta(X)$ in M.

2. Pairs of curves in surfaces will typically be regarded as having been isotoped to minimize the number of points in which they intersect. Only occasionally is care required with this convention. For example, if a surface S containing curves α and β is cut open along a circle c and the remnants $\alpha - c$ and $\beta - c$ are isotoped in $S - c$ to minimize their intersection (not necessarily fixing $\alpha \cap c$ or $\beta \cap c$) then when S is reassembled, new intersections are introduced because of twisting around c. In most contexts this will not matter, since it is the absence of intersections that typically complicates an argument.

3. When put in thin position as in [GST], a 1-complex Γ in S^3 will typically be regarded as having first been made "generic" with respect to the given height function on S^3; that is, all vertices will be of valence 3, with two edges incident from above (resp. below) and one from below (resp. above). At any height there will be at most one critical point or vertex. This convention leads to the following semantic problem: A process which puts Γ in thin position typically terminates when an edge of Γ is made level. Then Γ is no longer generic, but can be made generic by a small perturbation in which the height function on the edge becomes monotonic. To describe this situation we will sometimes say that the edge is a "perturbed" level edge.

2. UNKNOTTED HANDLEBODIES IN S^3 AND THEIR SPLITTING SPHERES

Consider a standard genus two handlebody H in S^3 and suppose μ^+, μ^-, μ^t are three non-parallel, non-separating meridian disks for H, fixed throughout our discussion. Let Σ denote the 4-punctured sphere $\partial H - (\mu^+ \cup \mu^-)$,

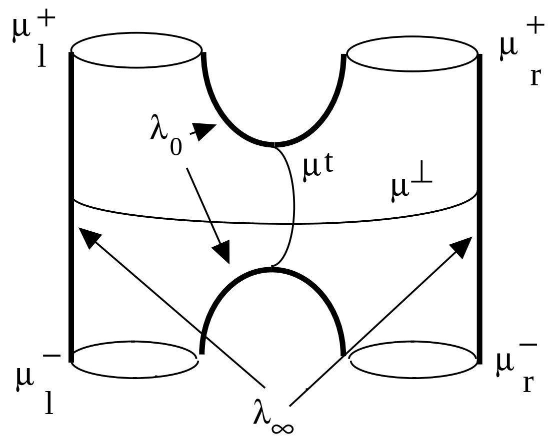

FIGURE 1

with boundary components $\mu_l^{\pm}, \mu_r^{\pm}$. Let $\mu^{\perp}$ denote a fixed separating meridian disk of H that is disjoint from $\mu^{\pm}$ and intersects μ^t in a single arc. There is a natural projection of Σ to the rectangle $I \times I$ so that $\mu^{\perp}$ projects to a horizontal bisector, μ^t to a vertical bisector, the two copies μ_l^+ and μ_r^+ of μ^+ in $\partial\Sigma$ project near the points $\partial I \times \{1\}$ and the two copies μ_l^- and μ_r^- of μ^- in $\partial\Sigma$ project near the points $\partial I \times \{-1\}$.

A *complete* pair of arcs in Σ will be a pair of arcs whose boundary has one point on each boundary component of Σ. A complete pair of arcs λ_∞ disjoint from μ^t is said to have *infinite slope* and a complete pair of arcs λ_0 that is disjoint from $\mu^{\perp}$ is said to have slope 0. The union $\lambda_\infty \cup \lambda_0$ divides Σ into two copies of $I \times I$, which we'll call the front face and the back face of Σ. See Figure 1. There is a natural correspondence between proper isotopy classes of complete pairs of arcs in Σ and the extended rationals $p/q \in \mathbb{Q} \cup \infty$. Here $|p|$ is the number of times one of the pair intersects $\mu^{\perp}$, $|q|$ is the number of times it intersects μ^t, and the fraction is positive (resp. negative) if the pair (when isotoped to have minimal intersection with $\lambda_\infty \cup \lambda_0$) is incident to the lower left corner μ_l^- on the front (resp. back) face. Note that a complete pair of arcs in Σ for which one end of each arc lies on μ^- and the other on μ^+ corresponds to a rational p/q with p odd. See Figure 2.

Definition 2.1. *Given two complete pairs of arcs λ and λ', with slopes p/q and p'/q' respectively, define $\Delta(\lambda, \lambda') = |pq' - p'q|$.*

Note that if $\Delta(\lambda, \lambda') \leq 1$ then the two pairs can be isotoped to be disjoint; otherwise, $|\lambda \cap \lambda'| = 2\Delta(\lambda, \lambda') - 2$.

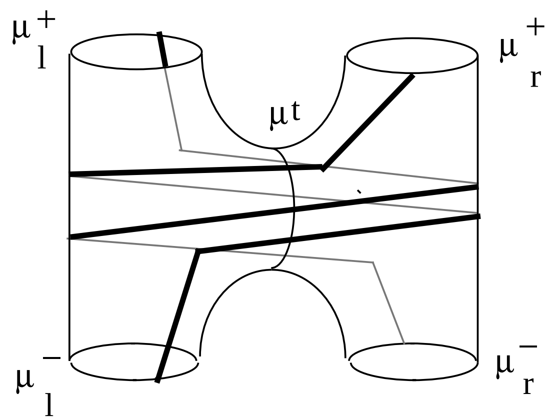

FIGURE 2. Slope 1/3

Definition 2.2. *Suppose $H \subset S^3$ is an unknotted genus two handlebody and S is a sphere that intersects ∂H in a single essential circle. Then S is a* splitting sphere *for H.*

Alternatively, we could define a splitting sphere to be a reducing sphere for the Heegaard splitting $S^3 = H \cup_{\partial H} (\overline{S^3 - H})$.

Suppose S is a splitting sphere for $H \subset S^3$, so $D = S \cap H$ is an essential separating disk in H. One possibility is that D lies entirely inside the ball $H - (\mu^+ \cup \mu^-)$. But, if not, then an outermost arc in D of $D \cap (\mu^+ \cup \mu^-)$ cuts off a disk D_0 and $D_0 \cap \partial H$ is an arc α with both ends at one of the boundary components of Σ, say μ_l^-. An arc in Σ, such as α, with both ends at μ_l^-, say, is called a *wave* based at μ_l^-. A simple counting argument (as many ends of arcs of $\partial D \cap \Sigma$ lie on μ_l^- as on μ_r^-) shows that there is also a wave α' based at μ_r^- and that one of the components of $\Sigma - \alpha$ is an annulus whose other end is one of μ_r^+ or μ_l^+. A spanning arc for that annulus unambiguously gives us an arc with one end on μ_l^- and the other end on one of μ_r^+ or μ_l^+. Similarly α' unambiguously gives us an arc from μ_r^- to the other choice of μ_r^+ or μ_l^+. Thus, given a splitting sphere, either its intersection circle with ∂H lies entirely in Σ or there is unambiguously defined a complete pair of arcs, each of which has one end on μ^- and one end on μ^+. On the other hand, knowing that a specific essential pair λ is the result of this construction, we do not know whether the waves are based on μ^- or on μ^+. In other words, to any choice of essential pairs of arcs, each of which has one end on μ^- and one end on μ^+, there correspond exactly two possible (pairs of) waves.

Definition 2.3. *Let H be the standard genus two handlebody in S^3, let μ^+, μ^-, μ^t be three non-parallel, non-separating meridian disks for H and*

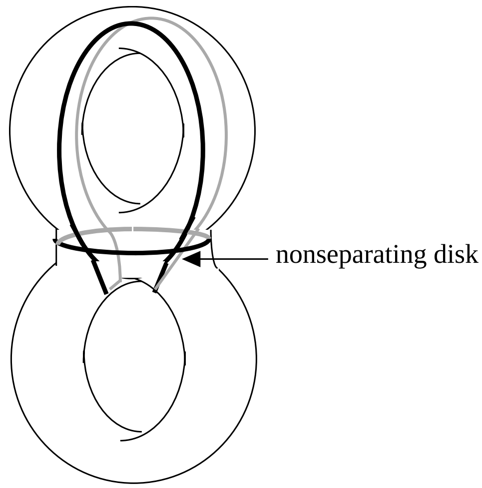

FIGURE 3

let $\mu^\perp$ a fourth, separating, meridian disk that is disjoint from $\mu^\pm$ and which intersects μ^t in a single arc. Finally let S be a splitting sphere for H.

Define $\rho_\perp(\mu^+,\mu^-,\mu^t,\mu^\perp,S) \in \mathbb{Q}\cup\infty$ to be the slope associated to the waves of $S\cap\Sigma$ as defined above.

Two splitting spheres S,S' are said to have the same augmented slope *(with respect to $\mu^+,\mu^-,\mu^t,\mu^\perp$), if*

$$\rho_\perp(\mu^+,\mu^-,\mu^t,\mu^\perp,S) = \rho_\perp(\mu^+,\mu^-,\mu^t,\mu^\perp,S')$$

and the associated waves are based at the same meridian μ^+ or μ^-.

A natural question is to what extent $\rho_\perp(\mu^+,\mu^-,\mu^t,\mu^\perp,S)$, or indeed the augmented slope, depends on our choices. Let us begin by considering different choices of splitting spheres.

Definition 2.4. *Let S and S' be two splitting spheres for $H\subset S^3$ and let C and C' be the corresponding separating curves $\partial H\cap S$ and $\partial H\cap S'$. Define the* intersection number *$S\cdot S'$ to be the minimum number of points in $C\cap C'$.*

The relation between spheres with low intersection number is easy to understand and describe, as we now outline. (For more detail see [Sc1].)

Lemma 2.5. *If S and S' are not isotopic then $S\cdot S'\geq 4$. If $S\cdot S'=4$ then $S\cap S'$ is a single circle and each of the 4 bigons formed in ∂H by adjoining an arc of $C-C'$ to an arc of $C'-C$ bounds a non-separating disk in exactly one of H or S^3-H. (See Figure 3).*

Proof. If $S\cdot S'=0$ then $S\cap S'=\emptyset$. In a genus two handlebody, any two disjoint separating disks are parallel, so S would be isotopic to S'. If $S\cdot S'=2$

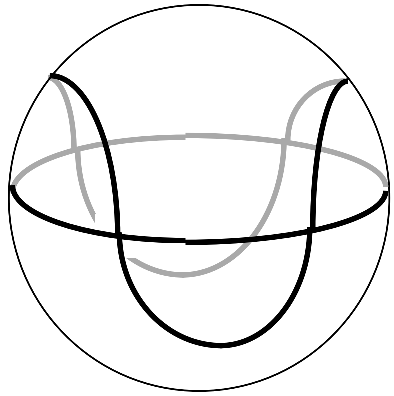

FIGURE 4

then C and C' would be two separating circles in ∂H that intersect in two points, hence they would be isotopic, a contradiction.

So suppose $S \cdot S' = 4$, i. e. $|C \cap C'| = 4$. Let γ represent the curve(s) $S \cap S'$. It is easy to remove, by isotopy of S or S' rel ∂H, any component of γ that does not intersect ∂H (hence intersect it at least twice). Each point in $C \cap C'$ is really a point in $S \cap S' \cap \partial H$, so $C \cap C' = C \cap \gamma = C' \cap \gamma$. If γ has more than one component then it has exactly two (since each must contain at least two points in $C \cap C'$). If γ does have two components then of course they are parallel in S and in S'. So $C \cup \gamma$ cuts four bigonal disk components from S and $C' \cup \gamma$ cuts four bigonal disk components from S'. These may be assembled (gluing along the four arc components of $\gamma - \partial H$) to give disks in both H and $S^3 - H$ with the same boundary in ∂H, namely two bigons formed from arcs of $C - C'$ and $C' - C$ in ∂H. This means that each of these bigons in ∂H is the intersection with a sphere, although each is non-separating. This is impossible.

So we conclude that $\gamma = S \cap S'$ is a single circle. We can then exploit the fact that there is essentially only one way for a pair of circles in the sphere to intersect in 4 points. See Figure 4. In particular, note that each of the four arcs $\gamma - \partial H$ is adjacent to exactly one bigonal disk in $S - (C \cup \gamma)$ and one bigonal disk in $S' - (C' \cup \gamma)$. Assembling these bigonal disks, just as above, gives the required non-separating disks in H or $S^3 - H$. □

Proposition 2.6. *If S and S' are two splitting spheres for $H \subset S^3$ then there is a sequence of splitting spheres*

$$S = S_0 \to S_1 \to \ldots \to S_m = S'$$

so that for $i = 1, \ldots, m$, $S_{i-1} \cdot S_i = 0$ or 4.

Proof. There is an obvious (but obviously not unique) orientation-preserving homeomorphism $h: S^3 \to S^3$ with the property that $h(H) = H$ and $h(S) = S'$. By the Alexander trick, h is isotopic to the identity.

In [Go] Goeritz shows that any isotopy of S^3 that ends in a homeomorphism carrying H to H is a product of particularly simple such isotopies, whose effect on a fixed separating sphere S_0 is easy to describe. In each case, either S_0 is preserved or the intersection number of S_0 with its image is 4.

The upshot is this: the homeomorphism h is the composition of homeomorphisms $h = h_1 \circ h_2 \circ \ldots \circ h_m$ where each h_i is the H-preserving homeomorphism of S^3 obtained by one of the simple isotopies. To obtain a sequence of splitting spheres we take $S_i = h_1 \circ h_2 \circ \ldots \circ h_i(S_0)$. Then notice that $S_i \cap S_{i-1}$ can be understood by viewing it as the image under the homeomorphism $h_1 \circ h_2 \circ \ldots \circ h_{i-1}$ of $h_i(S_0) \cap S_0$, so $S_i \cdot S_{i-1} = h_i(S_0) \cdot S_0 = 0$ or 4. □

Remark: This argument can be extended to Heegaard splittings of arbitrary genus, using work of Powell [Po]. See [Sc1].

Lemma 2.7. *If two splitting spheres for H have different augmented slopes, then there is an essential disk in $S^3 - H$ which intersects each of μ^+ and μ^- at most once.*

Proof. The conclusion is obvious if any splitting sphere S intersects H in a disk disjoint from $\mu^+ \cup \mu^-$, for just use the disk $S - H$. So we may as well assume that every splitting sphere defines an augmented slope.

Suppose S and S' are splitting spheres that give rise to two different augmented slopes. Then there is a sequence of splitting spheres, beginning with S and ending with S', such that each has intersection number 4 with the previous splitting sphere. Since the first and last terms have different augmented slopes, somewhere there is a pair in sequence with different augmented slopes. So we may as well assume that $S \cdot S' = 0$ or 4.

Let λ and λ' be the complete pair of arcs associated to the waves of $S \cap \partial H$ and $S' \cap \partial H$ respectively. First notice that $\Delta(\lambda, \lambda') \leq 1$. For if not, then $|\lambda \cap \lambda'| \geq 2$. If we double each of the four arcs, the total number of intersection points is 8, and converting a doubled arc into a wave can never remove intersection points, only add them.

Suppose next that $\Delta(\lambda, \lambda') = 1$. Then λ and λ' can be made disjoint, but not the waves that define them. Indeed, if one pair of waves has its ends on μ^- and the other on μ^+ then each wave from S intersects each wave from S' in at least two points, a total of at least $2 \cdot 2 \cdot 2 = 8$ points. See Figure 5.

On the other hand, if $\Delta(\lambda, \lambda') = 1$ and both pairs of waves have their ends on μ^- (or both on μ^+) then each wave from S intersects each pair of waves from S' in at least 2 points, a total intersection of *just the waves* of 4

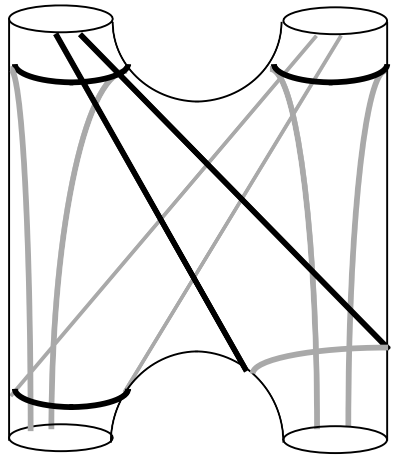

FIGURE 5

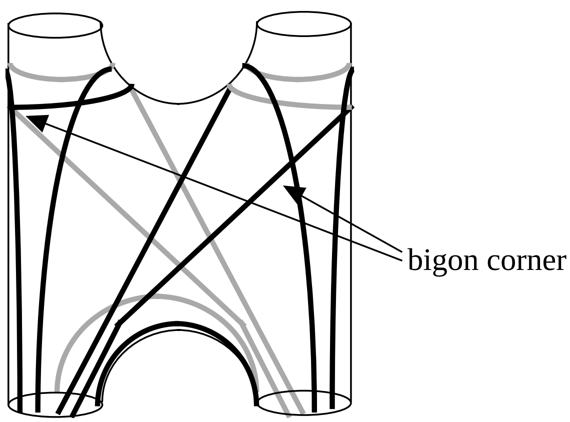

FIGURE 6

points. Any other arc of S with the same slope will intersect a wave of S', and vice versa, so S and S' must both be disjoint from μ^+. Following 2.5 we can say more: the bigons determined by the intersection points bound non-separating disks, two in H and two in $S^3 - H$. In our case each relevant bigon in $S^3 - H$ is made up of a union of arcs, each with one end on μ_r^- and one end on μ_l^-. In particular, the bigon intersects μ^- twice or more, always with the same orientation. (The extra intersections arise from arcs of $C \cap \Sigma$ and/or $C' \cap \Sigma$ that run from μ_r^- to μ_l^-. See Figure 6.) But no such curve can bound a disk in $S^3 - H$, for the union of the solid torus $H - \mu^+$ with the disk would define a punctured lens space in S^3.

Finally, suppose $\Delta(\lambda, \lambda') = 0$ but S has waves on μ^- and S' on μ^+. The first observation is that there is a meridian ν of H, with $\partial\nu \subset \Sigma$, so that ν is disjoint from both C and C'. (Just take ν to have the same slope as the waves.) Let W be the solid torus $H-\eta(\nu)$, with $\nu_\pm \subset \partial W$ the two remnant copies of ν. That is, let $\nu_\pm = \eta(\nu)\cap\partial W$. Pick a bigon $\beta \subset \partial W$, cut out by $C\cup C'$, so that β bounds a non-separating disk in $(S^3-H)\subset(S^3-W)$. The only way an essential curve on ∂W can bound a disk in S^3-W is if it is isotopic to a longitude of the knot, so β is isotopic in ∂W to a longitude of W. Moreover, because the waves are based at different meridians $\mu^\pm$, the arc components of $\beta-\mu^\pm$ adjacent to $\nu_\pm$ each have one end on μ^- and the other end on μ^+. This implies that every arc of $\beta-\mu^\pm$ has one end on μ^- and one end on μ^+. The only closed curve on ∂W that both has this property and is also isotopic (in ∂W) to a longitude, is a curve that is isotopic also in $(\partial W-\nu_\pm)\subset\partial H$ to a longitude, that is, to a curve that intersects each meridian $\mu^\pm$ of W exactly once. □

We have this corollary:

Corollary 2.8. *If different splitting spheres define different augmented slopes for $(\mu^+,\mu^-,\mu^t,\mu^\perp)$ then the knot core of the solid torus $H-\eta(\mu^t)$ is a 2-bridge knot.*

Proof. We can regard H as the regular neighborhood of a 1-vertex figure-8 graph Γ in S^3, in which $\mu^\pm$ are meridians of neighborhoods of the two edges of the graph. Let $k^\pm\subset\Gamma$ denote the subknots of Γ corresponding to the meridian disks $\mu^\pm$. It suffices to show that Γ is a standard unknotted figure-8 graph in S^3 since then the boundary of a regular neighborhood of the vertex of Γ serves as a bridge sphere for a 2-bridge presentation of the knot core of $H-\eta(\mu^t)$. Following the unpublished [HR] (see [ST]) it suffices then to show that each of the knots $k^\pm\subset\Gamma$ is the unknot.

Suppose that $E\subset S^3-H$ is an essential disk, given by Lemma 2.7, that intersects each of $\partial\mu^+$ and $\partial\mu^-$ at most once. If ∂E is disjoint from exactly one of the meridians, say μ^-, then E is an unknotting disk for k^+, and $H\cup\eta(E)$ is an unknotted solid torus whose core is k^-. Similarly, if ∂E is disjoint from both meridians, then E divides S^3-H into two solid tori, each of whose meridians is an unknotting disk for one of $k^\pm$. If ∂E intersects both meridians, then $H\cup\eta(E)$ is an unknotted solid torus in which both k^+ and k^- can be viewed (individually) as core curves. □

The next lemma shows that, given μ^t, there is a natural choice of meridians $\mu^\pm$. First note that if $\rho_\perp(\mu^+,\mu^-,\mu^t,\mu^\perp,S)$ is finite, then different choices of $\mu^\perp$ will change its value by a finite amount. Indeed, any other possible $\mu^\perp$ will differ from the given one by some number of full Dehn twists around μ^t, and such a Dehn twist changes $\rho_\perp$ by ±2.

Lemma 2.9. *Suppose that μ^t is a non-separating meridian disk for an unknotted genus two handlebody H and that S is a splitting sphere for H. Then*

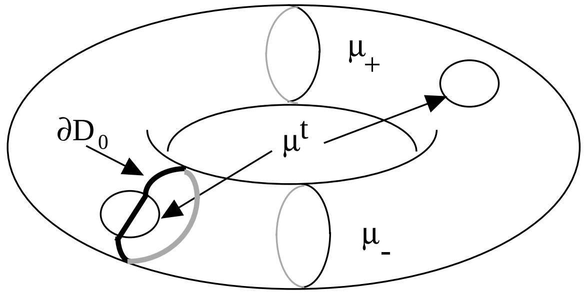

FIGURE 7

there is exactly one pair $\mu^{\pm}$ *of meridians disks for H such that* $\{\mu^t, \mu^+, \mu^-\}$ *is a complete set of meridian disks for H and* $\rho_\perp(\mu^+, \mu^-, \mu^t, \mu^\perp, S)$ *is finite for any (hence every) choice of* $\mu^\perp$ *that is disjoint from* $\mu^{\pm}$ *and intersects* μ^t *in a single arc.*

Proof. Suppose, to begin, that there is an extension of μ^t to a set of meridians $\{\mu^t, \mu^+, \mu^-, \mu^\perp\}$ with respect to which $\rho_\perp$ is finite. Because $\rho_\perp$ is finite, an outermost disk of $D = S \cap H$ cut off by the pair of meridians $\mu^{\pm} \subset H$ intersects μ^t. Then an outermost subdisk D_0 of this subdisk, cut off by μ^t, is disjoint from both meridians $\mu^{\pm}$. Furthermore, ∂D_0 intersects μ^t in a single arc dividing μ^t into two subdisks. The union of each of those subdisks with D_0 gives meridian disks for H parallel to $\mu^{\pm}$. (See Figure 7).

Now $\partial D_0 - \mu^t$ is an essential arc α in the twice punctured torus $T_0 = \partial H - \partial\mu^t$, and the arc has both its ends on a single puncture. Let $\alpha^{\pm}$ be closed curves in T_0 parallel to $\alpha \cup \{puncture\}$, lying on either side of $\alpha \cup \{puncture\}$. If β is any other arc of $T_0 \cap S$ which has both its ends on a single puncture, then β is disjoint from $\alpha^{\pm}$; this is obvious if the ends of β lie on the other puncture, and follows from a counting argument on ends of arcs in $S \cap T_0$ if the ends of β lie on the same puncture as those of α. Any non-parallel separating pair of closed curves, e. g. $\partial\mu^{\pm}$, in $T_0 - \alpha$ must be parallel to $\alpha^{\pm}$. So we see that $\mu^{\pm}$ are determined precisely by taking closed essential curves in T_0 that are parallel to $\alpha \cup \{puncture\}$. (See Figure 8)

It is now easy to see that there is always some such pair. Consider an outermost disk D_0 of D cut off by μ^t in H. Then the union of D_0 with each of the two subdisks of μ^t into which D_0 splits μ^t produces two natural meridian disks $\mu^{\pm}$ for the solid torus $H - \mu^t$. These, together with μ^t comprise a complete collection of meridian disks for H. Moreover, $\partial D_0 \cap T_0$ is an essential arc that is disjoint from both meridians $\mu^{\pm}$ of the solid torus $H - \mu^t$ bounded by T_0. Since some arc of $S \cap T_0$ lying in the pairs of pants $T_0 - \mu^{\pm}$ has both ends on $\partial\mu^t$, the wave described just before Definition 2.3 (and so the spanning arc disjoint from the wave whose slope determines $\rho_\perp$) must intersect $\partial\mu^t$. Hence $\rho_\perp$ is finite with respect to the meridian set $\{\mu^+, \mu^-, \mu^t\}$. □

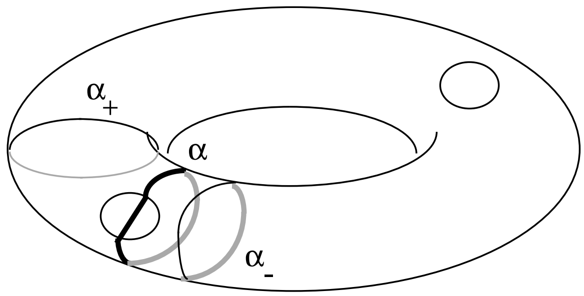

FIGURE 8

3. KNOTS WITH A SINGLE UNKNOTTING TUNNEL

Definition 3.1. *A knot K has* tunnel number one *if it is possible to attach a single arc γ to K in $S^3 - K$ so that $S^3 - \eta(K \cup \gamma)$ is a solid handlebody.*

Put another way, K has tunnel number one if there is a meridian disk μ^t for a standard genus two handlebody $H \subset S^3$ so that the solid torus $H - \mu^t$ has core knot isotopic to K.

Definition 3.2. *Let K be a knot and let γ be an unknotting tunnel for K. Let S be a splitting sphere for $H = \eta(K \cup \gamma)$ and $\mu^t \subset H$ be a meridian disk for γ. Let $\mu^{\pm}$ be the meridians of $H - \mu^t$ given by Lemma 2.9 and $\mu^{\perp}$ be a meridian of $H - \mu^{\pm}$ that intersects μ^t in a single arc. Define $\rho(K, \gamma, S) \in \mathbb{Q}/2\mathbb{Z}$ to be the value, mod 2 of $\rho_{\perp}(\mu^+, \mu^-, \mu^t, \mu^{\perp}, S)$.*

Since different choices of $\mu^{\perp}$ change $\rho_{\perp}$ by multiples of 2, $\rho(K, \gamma, S)$ is well-defined. Moreover, by Corollary 2.8, if K is not 2-bridge, ρ is independent of S and so can be written $\rho(K, \gamma)$. Here we examine some features that ρ reveals about the knot and its tunnel.

Much is already known about their geometry. The central theorem of [GST] says that if the graph $K \cup \gamma$, viewed as a trivalent graph in S^3, is put in *thin position*, then K is, on its own, in thin position and in bridge position, and γ is a (perturbed) level edge. Moreover, the tunnel γ either has one end on each of two different maxima (or minima) or is a (perturbed) level loop, or "eyeglass", whose endpoint lies on a single maximum (or minimum) and which encircles all the other bridges of K.

The first claim is that the slope $\rho(K, \gamma, S)$ is naturally revealed by some thin positioning of $K \cup \gamma$. That is, there is a thin positioning of $K \cup \gamma$ so that the two isotopy classes of meridians of $K - \gamma$ with which level spheres intersect $K - \gamma$ are the classes $\mu^{\pm}$ identified in Lemma 2.9.

First we consider the case in which, upon thinning, γ becomes a (perturbed) level eyeglass. We can equivalently take, in this case, γ to be a level

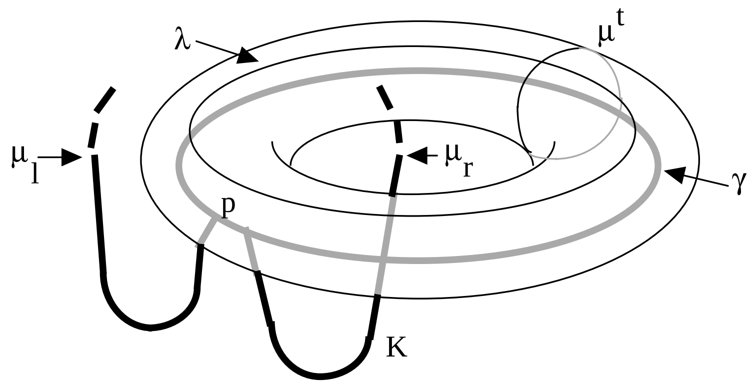

FIGURE 9

edge with both ends incident to K in the same point $p \in K$, i. e. γ is a cycle. Let P be the level sphere in which γ lies. Following [GST], we can furthermore take p to be the lowest maximum (or highest minimum) of the knot K; every bridge of K other than the one containing p has one end on each disk component of $P - \gamma$. By a meridian of $K - p$ we will mean any meridian of K that is disjoint from the (vertical) meridian of K corresponding to the maximum p.

Lemma 3.3. *Suppose S is a splitting sphere for $H = \eta(K \cup \gamma)$ and let $D = S \cap H$. Then any outermost disk D_0 of D cut off by a meridian μ of $K - p$ intersects the meridian μ^t of the tunnel γ. Moreover, a subdisk of D_0 cut off by μ^t is disjoint from a horizontal longitude of $\eta(\gamma)$.*

Proof. Cut H open along the copies of μ, denoted μ_l and μ_r, corresponding to the points of $K \cap P$ that are nearest to p in K. Then μ_r and μ_l lie on different (disk) components of $P - \gamma$. Cutting H open along these meridians leaves one component that is a solid torus W whose core is the cycle γ and whose boundary contains disks corresponding to μ_l and μ_r. The meridian μ^t of γ and the curves $P \cap \partial W$ (parallel in H) naturally define respectively a meridian curve (which we continue to call μ^t) and horizontal longitudes in the twice-punctured torus $T_0 = \partial H \cap W$. We want to understand the pattern of arcs $\Gamma = S \cap T_0$. See Figure 9.

We begin by examining how Γ intersects the twice punctured annulus A obtained by cutting open T_0 along the horizontal longitude λ at the top of ∂H. Note that μ^t intersects A in a single spanning arc. The boundary of A can be thought of as the two copies $\partial_0 A, \partial_1 A$ of λ that lie in $P \cap T_0$.

Claim: Among the arcs in $\Gamma \cap A$ that intersect μ^t, either there is one that has both ends on the same component of ∂A and separates the punctures

μ_r and μ_l, or there is one that has one end on ∂A and the other end on a puncture μ_r or μ_l.

Proof of Claim: Let $E = S - H$ be the exterior disk, and consider an outermost disk E_0 cut off from E by an outermost arc β of $E \cap (P - H)$. (It is easy to remove all closed components of $E \cap P$, since K is thin.) Let $\alpha = \partial E_0 - \beta$. Since β obviously lies in a single component of $P - H$, it follows easily that E_0 must lie below P (all arcs of $\Gamma - A$ can be assumed to be essential and so each spans the annulus $T_0 - A$) and so $E_0 \cap H$ lies in A. The arc β cannot have one end on each of μ_l and μ_r since these lie in distinct components of $P - \gamma$. If the ends of β (hence the ends of α) both lie on the meridian μ_l, say, then α is a longitudinal arc in the punctured annulus A. That is, $\alpha \cup \mu_l$ is a core curve of A. On the other hand, the outermost disk of D cut off by μ must be a meridional wave in T_0, so it too must also be based at μ_l. The complement of the two arcs, one meridional and the other longitudinal, is then a punctured disk in T_0 containing just the puncture μ_r. But then no wave could be based at μ_r, and there would be more ends of Γ on μ_l than on μ_r, an impossibility. We deduce that α has one or both ends on ∂A. Notice that if both ends of α lie on ∂A then they must lie on the same component ($\partial_0 A$, say) of ∂A (since they are connected by an arc in $P - H$) and then the subdisk A_0 of A cut off by α contains at least one puncture (or it would be inessential) but not both (else Γ would intersect $\partial_0 A$ more often than it intersects $\partial_1 A$).

It remains to show that α intersects μ^t. We will show that if it does not, it can be used to make K thinner. Suppose that α is disjoint from μ^t and so lies in the twice-punctured disk $A - \mu^t$. Consider first the case in which α has one end on μ_r (say) and other end on ∂A and let K_- denote the subarc of K that lies between the maximum p and the meridian μ_r. $W - \mu^t$ is a 3-ball which is further cut by P into two 3-balls; let W_- denote the one that lies below P. It's easy to see that α is parallel in W_- to K_-. Via that parallelism and the disk E_0, K_- can be moved to lie entirely in the plane P. (Afterwards, $K_- - W = \beta$). This cancels a minimum with a maximum, thereby thinning K. Similarly, if both ends of α lie on (necessarily the same component of) ∂A, then E_0 basically presents a lower cap that separates K_- from the other components of $K - P$ that lie below P. Within the ball bounded by the lower cap and a subdisk of P, K_- can again be moved to lie in P. From this contradiction, we conclude that α intersects μ^t, establishing the Claim.

Following the Claim, we have two cases to consider, corresponding to the two types of arcs given by the Claim. Both arguments will use the 4-punctured sphere $\Sigma = T_0 - \mu^t$ bounded by μ_l, μ_r and two copies ${\mu^t}_\pm$ of μ^t. We briefly recount some of its properties. An outermost disk D' of D cut off

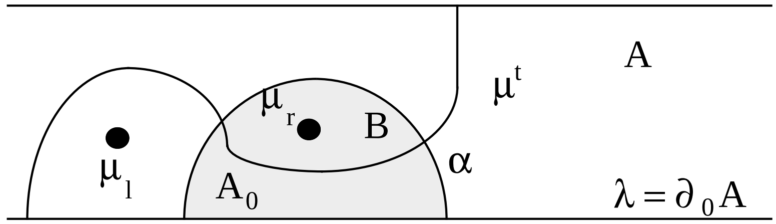

FIGURE 10

by $\mu^t \cup \mu$ exhibits a wave of Γ in Σ; that is, $\partial D' \cap \Sigma$ is an arc with both ends at one of the meridians. If it is based at meridian μ_l an end count shows that there is a wave based at μ_r and vice versa. Similarly if a wave is based at one of $\mu^t{}_\pm$ there is a wave based at the other. Furthermore, any two waves in Σ must have the same slope. The arc $\lambda \cap \Sigma$ is a path joining the two copies $\mu^t{}_\pm$; we can use it to establish slope $0 = 0/1$ in Σ. In these terms, a restatement of the Lemma is the claim that the wave determined by D' is based at one of $\mu^t{}_\pm$ and is disjoint from λ.

Suppose first that the arc α given by the Claim has both ends on a boundary component $\partial_0 A$ of A and cuts off from A a disk A_0 containing a single puncture μ_r (say). Once $\alpha \cap \mu^t$ is minimized by isotopy, an outermost arc of μ^t in A_0 cuts off a bigon B containing μ_r and bounded by subarcs of μ^t and α. (See Figure 10.) An outermost arc of Γ in B (possibly the subarc of α in ∂B) is a wave of Γ in Σ based at $\mu^t{}_\pm$ that is disjoint from λ, since B is. It follows that all the waves in Σ, including that determined by D', have the same property.

Suppose finally that only one end of α lies on λ and the other end lies on μ_l (say). Since α intersects μ^t, the end segments of α in Σ have these properties: One end, η^+, connects one of $\mu^t{}_\pm$ to μ_l and is disjoint from λ. The other end connects λ to one of $\mu^t{}_\pm$ essentially; let η^- denote the segment of $\Gamma \cap \Sigma$ that contains this end. Since one of $\eta^\pm$ intersects λ and one does not, they have different slopes in Σ. (See Figure 11.) Observe that if waves are based on two boundary components of a 4-punctured sphere, the only disjoint arc that can have a different slope than the waves is an arc that connects the bases of the waves. Since only one end of η^+ can be the base of a wave, it follows that η^- must connect the two bases of the waves. Since one end of η^- lies on one of $\mu^t{}_\pm$ this means that the waves must be based at $\mu^t{}_\pm$, and η^- runs between $\mu^t{}_\pm$. Finally, the slope of the wave must be that of η^+, so the wave is disjoint from λ. □

Corollary 3.4. *Suppose S is a splitting sphere for $H = \eta(K \cup \gamma)$ and let $D = S \cap H$. Suppose in a thin positioning of $K \cup \gamma$ the tunnel γ is a level eyeglass at a maximum (minimum) p of K. Then the ends of γ at p may be*

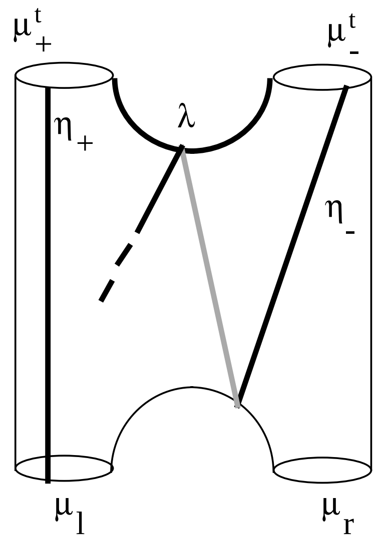

FIGURE 11

slid slightly down (up) K so that, with respect to the resulting meridians $\mu^{\pm}$ of $K-\gamma$, ρ is finite.

Proof. We know from Lemma 2.9 how to find meridians with respect to which ρ is finite: Begin with an outermost disk D_0 of D cut off by μ^t and choose meridians parallel to $\mu^t \cup \partial D_0$ in the solid torus $H - \mu^t = \eta(K)$. Lemma 3.3 tells us precisely what those meridians are: one is μ, the meridian of $K-p$. The other is bounded by the union of an arc disjoint from the highest horizontal longitude λ of the circuit γ_c and an arc that intersects λ once. Hence the boundary of the second meridian has a single maximum, so it can be viewed as simply the vertical meridian of K at p, separating the ends of γ. □

Theorem 3.5. *Suppose S is a splitting sphere for $H = \eta(K\cup\gamma)$ and let $D = S\cap H$. Then γ can be slid and isotoped to some thin positioning of $K\cup\gamma$ so that ρ is finite with respect to the meridians of $K-\gamma$.*

Proof. Put $K\cup\gamma$ in thin position so that γ can be levelled. If γ is an eyeglass, the result follows from Corollary 3.4, so assume γ is a level edge e. Following [GST] we can assume that K is in minimal bridge position and the ends of γ connect the two highest maxima of K. If P_u is a level sphere just below e, then the part of $S^3 - H$ lying above P_u is just a collar $(P_u - H)\times I$, so we may as well assume that the exterior disk $E = S - H$ intersects this product region in bigons whose boundaries each consist of an arc in $P_u - H$ and an arc in $\partial H - P_u$.

Let μ^+ and μ^- be the meridians of the two arcs $K - \gamma$; we know that μ^+, say, cuts off an outermost disk D_0 of $D = S \cap H$. If D_0 intersects the meridian μ^t of the tunnel γ we are done, so suppose that D_0 is disjoint from μ^t. Somewhere below P_u and above the highest minimum of K, there is a generic level sphere P which cuts off both an upper disk E_u and a lower disk E_l from E. Since E_u and D_0 can be made disjoint, it follows that ∂E_u crosses the meridian μ^t of the tunnel at most once.

We are now in a position to apply the argument of [GST, Theorem 5.3], though now in the context that the edge e disjoint from P is γ, not a subarc of K: There cannot be simultaneously an upper cap and a lower cap that have disjoint boundaries in P, or K could be thinned. If there is an upper cap and a disjoint lower disk or a lower cap and a disjoint upper disk, then, as in [GST] we can find such a pair for which the interior of the disk is disjoint from P. Then thin position implies that there cannot be simultaneously an upper cap and a lower disk and, if there is a lower cap which is disjoint from an upper disk (whose interior is now disjoint from P) then we can ensure that the boundary of the upper disk runs across the tunnel, hence exactly once across the tunnel. Similarly, if there is an upper disk which is disjoint from a lower disk then we can ensure that the interior of the upper disk is disjoint from P and is disjoint either from a lower cap or a lower disk whose interior is also disjoint from P. Furthermore, the boundary of the upper disk must be incident to the tunnel, hence run exactly once across the tunnel. But then the upper and lower disks describe how to push the tunnel down to, or even below, the level of at least one minimum. Following [GST, Corollary 6.2, Cases 1 and 2] this implies (almost, see next paragraph) that the tunnel can be pushed down to connect two minima, which we may make to be the lowest minima. Furthermore, this operation lowers by two the number of extrema of K (in our case, the number of minima) found in the interior of the component of $K - \gamma$ containing μ^+, the base of the waves.

The argument then continues; it finally fails when there is only one extremum in the component of $K - \gamma$ containing μ^+. That is, in the argument above, it may finally happen that the two maxima to which the ends of γ are attached have only one minimum lying between them, and it is the component on which μ^+ is found. In this case, the upper disk pushes γ down to P, and then γ together with the segment $\beta \subset (K - P)$ containing the minimum form an unknotted cycle. If any other minimum could be pushed up above this cycle then, again following [GST, Corollary 6.2], all the other minima could, and the cycle would bound a disk disjoint from K, a contradiction. We conclude that the lower disk or cap is in fact a lower disk that pushes β up to P, thereby forming a level cycle. It can be used to slide γ to form a level eyeglass, at which point we appeal to Lemma 3.3. □

4. Pushing the tunnel off of a Seifert surface

The next section will show that if $\rho(K,\gamma) \neq 1$ then γ can be pushed onto a minimal genus Seifert surface for K. Ironically, the first step is to show that γ can be pushed completely off of such a surface, which we will show in this section.

Let K be a knot in S^3 and F be a Seifert surface for K. We say that K is *parallel* in S^3 to an imbedded curve c in F if there is an annulus A imbedded in S^3 such that $A \cap F = \partial A = \partial F \cup c = K \cup c$.

Lemma 4.1. *Let K be a knot in S^3 and F be a minimal genus Seifert surface for K. Suppose K is parallel in S^3 to a curve c in F. Then $K = \partial F$ is parallel to c in F.*

Proof. Let A be the annulus giving the parallelism between K and c. Let $\eta(A)$ be a neighborhood of A containing a neighborhood of K. Since A is an annulus, we can think of $\eta(A)$ as being a ribbon-like neighborhood of K itself. In the complement of $\eta(A)$, the remnant of F is a possibly disconnected surface $\overline{F}$, with three (preferred) longitudinal boundaries on the boundary of $\eta(A)$. If $\overline{F}$ is disconnected (corresponding to the case in which c is separating) then one of the components of $\overline{F}$ is a Seifert surface for K. Since it cannot be of lower genus than F, the other component must be an annulus, defining a parallelism between K and c in F, as required.

Suppose c is non-separating. Then zero-framed surgery on K yields a manifold M and "caps off" $\overline{F}$; call the capped-off surface $\overline{F}'$. A capped-off version F' of the Seifert surface F also imbeds in M and $\overline{F}'$ and F' represent the same homology class in M. Since $genus(\overline{F}')$ is less than $genus(F')$, it follows from work of Gabai [Ga, Corollary 8.3] that genus K is less than genus F, a contradiction. □

Proposition 4.2. *Suppose K is a knot and γ is an unknotting tunnel for K. Then γ may be slid and isotoped until it is disjoint from some minimal genus Seifert surface for K.*

Proof. First choose a minimal genus Seifert surface F and slide and isotope γ, doing both so as to minimize the number of points of intersection between γ and F. The slides and isotopies may leave γ as either an edge or an eyeglass. (In the latter case, let γ_a be the edge in γ and γ_c be the circuit.) We aim to show that $\gamma \cap F = \emptyset$.

Suppose to the contrary that after the slides and isotopies $\gamma \cap F$ is non-empty. Let E be an essential disk in the handlebody $S^3 - \eta(K \cup \gamma)$ chosen to minimize the number $|E \cap F|$ of components in $E \cap F$. $|E \cap F| > 0$ for otherwise the incompressible F would lie in a solid torus, namely (a component of) $S^3 - \eta(K \cup \gamma \cup E)$, and so be a disk. Furthermore, since F is incompressible, we can assume that $E \cap F$ consists entirely of arcs.

Let e be an outermost arc of $E \cap F$ in E, cutting off a subdisk E_0 of E. The arc e is essential in $F - \gamma$, for otherwise we could find a different essential disk intersecting F in fewer components. Let $f = \partial(E_0) - e$, an arc in $\partial\eta(K \cup \gamma)$ with each end either on the longitude $\partial F \subset \partial\eta(K)$ or a meridian disk of γ corresponding to a point of $\gamma \cap F$.

- If a meridian of γ is incident to exactly one end of f, then we can use E_0 to describe a simple isotopy of γ which reduces the number of intersections between γ and F.
- If *no* meridian of γ is incident to an end of f, then both ends of f lie on $\partial F \subset \partial\eta(K)$. If the interior of f runs over γ we are done, for f is disjoint from F. If the interior of f lies entirely in $\partial\eta(K)$ and e is essential in F then E_0 would be a boundary compressing disk for F, contradicting the minimality of *genus*(F). If the interior of f lies entirely in $\partial\eta(K)$ and e is *inessential* in F, then the disk D_0 it cuts off from F necessarily contains points of γ (since e is essential in $F - \gamma$). But then replacing D_0 by E_0 would lower $F \cap \gamma$.

The only remaining possibility is that both ends of f lie on the same meridian of γ. In this case, e forms a loop in F and the ends of f adjacent to e both run along the same subarc γ_0 of γ. Since f is disjoint from F, γ_0 either terminates in $\partial\eta(K)$ or γ is an eyeglass and γ_0 terminates in the interior vertex of γ.

If γ_0 terminates in an end of γ in $\eta(K)$ then, since the interior of f is disjoint from F, f must intersect $\partial\eta(K)$ in either an inessential arc in the torus or in a longitudinal arc. The former case is impossible, since if the disk bounded by the inessential arc did not contain the other end of γ then it could be isotoped away and $E \cap F$ reduced. If the disk did contain the other end of γ, then ∂E would cross one end of γ more often than the other, an impossibility. It follows that f intersects the torus $\partial\eta(K)$ in a longitudinal arc. Then $\eta(\gamma_0 \cup E_0)$ is a thickened annulus A, defining a parallelism in S^3 between K and the loop e on F. By Lemma 4.1 that means the loop e is parallel to ∂F. Substituting A for the annulus between e and ∂F in F would create a Seifert surface for K with fewer intersections with γ, a contradiction.

So γ is an eyeglass and γ_0 terminates in the interior vertex of γ. If, nonetheless, the interior of f intersects $\partial\eta(K)$ this means that f traverses the edge $\gamma_a \subset \gamma$ so γ_a is disjoint from F. In that case, we can just repeat the argument above, absorbing $\eta(\gamma_a)$ into $\eta(K)$. So we can assume that f lies entirely on $\partial\eta(\gamma)$. Now the component Q of $\partial\eta(\gamma) - F$ on which f lies is either a punctured torus (if F is disjoint from γ_c) or a pair of pants. In the former case, consider the Seifert surface F' obtained from F by removing the meridian disk μ_f of $F \cap \eta(\gamma)$ on which the ends of f lie and substituting

Q. F' is of one higher genus than F, and intersects γ in one fewer point. Surgery to F' using E_0 reclaims the minimal genus without introducing another intersection point. Thus we get a contradiction to our choice of F.

If Q is a pair of pants a similar argument works: Since F is incompressible, it follows that the loop e bounds a disk in F. Since f is essential, that disk contains exactly one of the other two meridian disks (call it μ_e) of γ in F that correspond to boundary components of Q. Remove the meridian disks μ_f and μ_e from F and attach instead an annulus that runs parallel to the subarc of γ (containing the interior vertex) that has ends at μ_e and μ_f. This creates a Seifert surface F' of genus one greater than F, but having one fewer intersection point with γ. Now do surgery on F' using E_0, deriving the same contradiction as before. □

5. Pushing the tunnel onto a Seifert surface

In this section we will show that if $\rho(K,\gamma) \neq 1$ then γ can be pushed onto a minimal genus Seifert surface for K. The first step was taken in the previous section: In general, γ can be pushed completely off of some such surface. A difficulty is that, after the slides used to push the tunnel off the Seifert surface, we no longer know that the resulting meridians of $K - \gamma$ are the ones by which we defined $\rho(K,\gamma)$ above. Our strategy will be to use the meridians $\mu^{\pm}$ by which we defined ρ, but at this cost: On the twice-punctured torus $\partial\eta(K) - \gamma$ we no longer can assume that ∂F is a standard longitude. All we know is that it is a curve that is isotopic in the *unpunctured* torus to the standard longitude. The point of the following lemma, is that this situation is not a serious obstacle to further analysis.

Lemma 5.1. *Suppose K is a knot with unknotting tunnel γ, $H = \eta(K \cup \gamma)$, and $K_0 \subset \partial H$ is a curve in the twice-punctured torus $\partial H - \eta(\gamma)$ which, in the unpunctured torus $\partial\eta(K)$ is isotopic to a standard longitude. Suppose α is an arc in ∂H so that $\alpha \cap K_0 = \partial\alpha$ and α traverses $\eta(\gamma)$ once. Then α is an unknotting tunnel for K_0 and the pair (K_0, α) is equivalent (by slides and isotopies) to the pair (K, γ).*

Proof. Since α traverses γ once, we can shrink α, dragging along its end points in K_0 until α is just a spanning arc of the annulus $\partial H \cap \eta(\gamma)$. (At this point, we can identify α with γ but we cannot yet identify K_0 with K.) K_0 is a possibly complicated curve lying on $T = \partial\eta(K)$ and K_0 is incident to $\alpha = \gamma$ at the ends of γ.

Using a collar $T \times I$ of T in $\eta(K)$ isotope $K_0 \subset T$ until it is a standard longitude lying on the boundary of the smaller tubular neighborhood $\eta_- = \eta(K) - (T \times I)$ of the core K. Extend the isotopy to an ambient isotopy of T, i. e. a self-homeomorphism of $T \times I \subset \eta(K)$. This extends the ends of α as a 2-braid through $T \times I$; call the extended arc α_+. The construction shows that $\eta_- \cup \eta(\alpha_+)$ is isotopic in H to $\eta(K_0 \cup \alpha)$. Now make the braid trivial by absorbing it into η_-. (This translates into slides of the ends of α_+ on $\partial\eta_-$). Afterwards, $H - (\eta_- \cup \eta(\alpha_+))$ is just a collar of ∂H. □

Theorem 5.2. *Suppose K is a knot with unknotting tunnel γ and S is a splitting sphere for the handlebody $H = \eta(K \cup \gamma)$ with $\rho(K, \gamma, S) \neq 1$. Suppose further that F is an incompressible Seifert surface for the knot K and that F is disjoint from γ. Then γ can be slid and isotoped until it lies on F.*

Proof. Let K_0 be the copy $F \cap \partial H$ of K in ∂H. Consider the hemispheres $E = S - H$ and $D = S \cap H$. By definition of ρ there are meridians μ^+ and μ^- for $\eta(K) \subset H$ that realize the slope ρ. Then, in particular, there are subarcs of $\partial E = \partial D$ that are waves based at one of these meridians. (Warning: we know little about how these meridians intersect K_0.) If the exterior disk E were disjoint from F, then F would lie in a solid torus obtained by compressing H to the outside along E, contradicting the assumption that F is incompressible. So $F \cap E \neq \emptyset$. We can isotope ∂E and ∂F to have minimal intersection and then remove any closed components of $F \cap E$ since F is incompressible.

Let E_0 be an outermost disk of E cut off by F. Then ∂E_0 consists of two arcs, α lying on ∂H and β lying on F. We may assume that β is essential in F, for otherwise the subdisk of F it cuts off, together with E_0, would again give an essential disk in $S^3 - H$ that is disjoint from F. An important observation is that the ends of α lie on ∂F and are *incident to the same side of F*. That is, if ∂F is normally oriented, the orientation points into (say) α at both ends of α.

It is also true that α must cross the meridian μ^t of γ at least once. For otherwise, E_0 would give a ∂-compression of F to $\partial\eta(K)$, contradicting the fact that F is an incompressible (hence ∂-incompressible) Seifert surface for K. If α crosses μ^t exactly once, then, following Lemma 5.1, α is equivalent to γ, so E_0 provides a way of isotoping γ to F, completing the proof. Hence it suffices to show:

Claim: Any subarc of $\partial E \cap \partial H$ whose interior is disjoint from ∂F and whose ends lie on the same side of ∂F crosses μ^t at most once.

Proof of claim: Continue to denote this subarc by α. Let Σ denote the four-punctured sphere obtained by cutting open ∂H along the meridians μ^+ and μ^- of K.

Case 1: ∂F intersects Σ in loops as well as arcs.

Say the loops are based at μ_l^+ and μ_r^+ (see Figure 12).

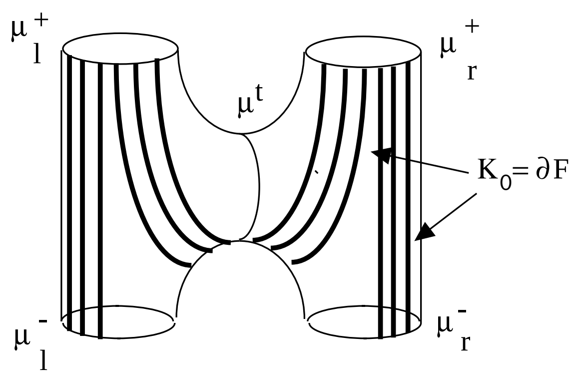

FIGURE 12

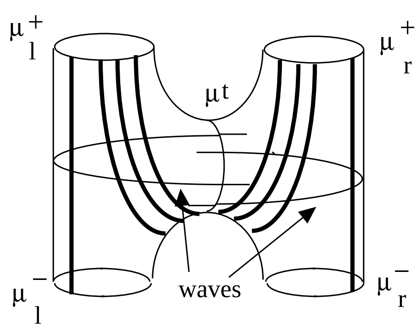

FIGURE 13

We first note that we may as well assume that α lies entirely in Σ. Indeed, since ρ is finite, there are waves of $\partial E = \partial D$ based at μ^t that lie in Σ and are on opposite sides of μ^t (see Figure 13). It follows that α cannot cross μ^- except to terminate at K_0; if that is the case, then α is disjoint from μ^t as required. For the same reason, we can assume that any component of $\alpha - \mu^{\pm}$ with an end on μ^+ must lie in the component Σ^t of $\Sigma - \partial F$ that contains μ^t, for any other component can be isotoped out through μ^+. But the segments $\mu_r^+ \cap \Sigma^t$ and $\mu_l^+ \cap \Sigma^t$ are disjoint, for otherwise the two adjacent loops of $\partial F \cap \Sigma$ based at μ_l^+ and μ_r^+ would form a simple closed curve, which is impossible.

So now assume that $\alpha \subset \Sigma$. (We say α is *short*.) Since ∂F intersects Σ in loops as well as arcs, then to intersect μ^t at all, α must lie in the annulus lying between the two outermost loops. (See Figure 14.) This annulus has μ^t as its core (since ∂F is disjoint from μ^t) and any essential path in the annulus intersects μ^t at most once.

Case 2: ∂F intersects Σ only in arcs.

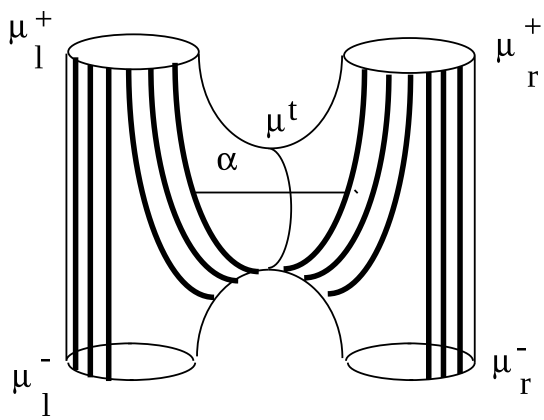

FIGURE 14

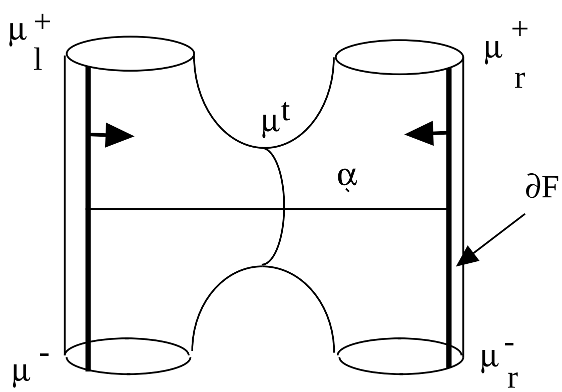

FIGURE 15

If ∂F intersects Σ only in arcs, then (considering the torus $H - \mu^t$) it must be in precisely two arcs, both of infinite slope (connecting μ_l^+ to μ_l^- and μ_r^+ to μ_r^-.) If $\alpha \subset \Sigma$, the argument is the same as above, using the annulus $\Sigma - \partial F$ (see Figure 15).

So finally suppose α is not contained in Σ and suppose with no loss that the waves of $\Gamma = S \cap \Sigma$ are based at μ_l^- and μ_r^-. Any arc of $\Gamma = S \cap \Sigma$ with an end on either μ_l^+ or μ_r^+ will then have a fixed slope, and since the slope ρ is not $+1 = -1 \in \frac{\mathbb{Q}}{2\mathbb{Z}}$ it will intersect ∂F in its interior. Moreover, if the normal orientation induced by that of F points towards μ_l^+ on an arc with an end on μ_l^+ it will point away from μ_r^+ on an arc with an end on μ_r^+ (see Figure 16). Hence we conclude that α cannot cross μ^+.

It is as easy to rule out the possibility that α crosses μ^-. An arc of Γ with one end on μ_l^- may have slope ρ or have slope ρ' with $\Delta(\rho, \rho') = 1$. That is,

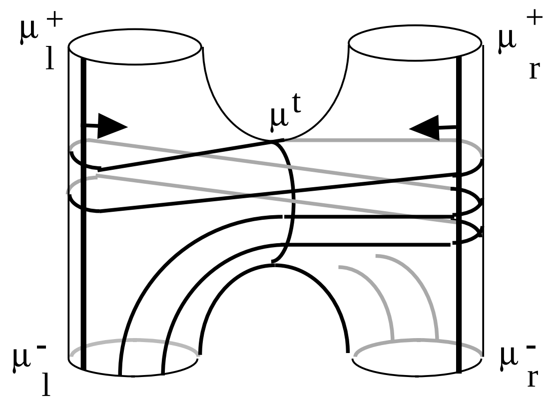

FIGURE 16

fixing a meridian $\mu^{\perp}$ so that $\rho_{\perp} = p/q, q > |p|$, p odd, we could have that $\rho'_{\perp} = r/s$ with $|ps - rq| = 1$. The arc could not have its other end on μ^{+} so r is even, hence s is odd and $r/s \neq \pm 1$.

Note that $|p/q - r/s| = 1/|qs|$ whereas both $|p/q|$ and $1 - |p/q| \geq 1/|q| \geq 1/|qs|$. In words, p/q is at least as close to r/s as it is to 0 or ± 1. Hence $|r/s| < 1$ and either $r/s = 0$ (e. g. when $p/q = 1/k$) or r/s has the same sign as p/q. So any subarc of Γ with an end on μ_l^- or μ_r^- is either disjoint from ∂F (if $\rho'_{\perp} = 0$) or it first intersects ∂F on the same side as an arc with slope ρ. In particular, a subarc of $S \cap \partial H$ that intersects μ^- and intersects ∂F precisely in its endpoints, necessarily ends on opposite sides of ∂F, and so cannot be α. $\square$

The proof of Case 1 in Theorem 5.2 did not require the assumption that $\rho \neq 1$. In particular, we have the following corollary:

Corollary 5.3. *Suppose K is a knot with unknotting tunnel γ and S is a splitting sphere for the handlebody $H = \eta(K \cup \gamma)$. Suppose further that F is an incompressible Seifert surface for the knot K and that F is disjoint from γ. If either of the meridians $\mu^{\pm}$ of H, chosen via 2.9 so that ρ is finite, intersects $K_0 = F \cap \partial H$ in more than one point, then γ is isotopic to an arc on F.*

More significantly:

Corollary 5.4. *Suppose K is a knot with unknotting tunnel γ, S is a splitting sphere for the handlebody $H = \eta(K \cup \gamma)$, genus$(K) = 1$ and $\rho(K, \gamma, S) \neq 1$. Then K is a 2-bridge knot.*

Proof. By Proposition 4.2 we may assume that γ is disjoint from some genus one Seifert surface F. Theorem 5.2 shows that we can then isotope γ onto F, necessarily as an essential arc. Then $F-\eta(\gamma)$ is an incompressible annulus A whose ends comprise a non-simple (because of A) tunnel number one link L. (The core of L's unknotting tunnel is the dual arc to γ in the rectangle $\eta(\gamma)\cap F$.) This implies, via [EU], that each component of L is unknotted. This then implies via [HR] or [ST] that the figure 8 graph obtained from $K\cup\gamma$ by crushing γ to a point v can be isotoped to lie in a plane. This finally implies that K is a 2-bridge knot, with $\partial\eta(v)$ the bridge sphere. □

This establishes the following conjecture of Goda-Teragaito ([GT]) in the case that $\rho\neq 1$, and without the assumption that K is hyperbolic.

Conjecture 5.5 (Goda-Teragaito)**.** *A knot that is genus one, has tunnel number one, and is not a satellite knot is a* 2*-bridge knot.*

The verification of the remaining case, when $\rho = 1$ and K is hyperbolic, will be discussed elsewhere. Note that Matsuda [Ma] has verified the conjecture for all knots which are 1-bridge on the uknotted torus, i. e. those with a $(1,1)$-decomposition.

6. A SAMPLE CALCULATION

Let $T\subset S^3$ be an unknotted torus and $K\subset T$ be a torus knot in T. Let γ be a spanning arc for the annulus $T-K$. γ is an unknotting tunnel for K since $S^3-\eta(K\cup\gamma)$ is a handlebody, namely the union of the interior and the exterior of T along a disk in T. In this section we will show that $\rho(K,\gamma)=1$. We will then use this calculation to construct examples of knots and tunnels with ρ taking any value in $\mathbb{Q}/2\mathbb{Z}$.

To understand $\eta(K\cup\gamma)$ we will regard it as a bicollar of the punctured torus $T\cap\eta(K\cup\gamma)=\eta_T$ and consider its lift $\tilde{\eta}\times I$ to the universal cover $\underline{U}=\mathbb{R}^2\times I$ of $T\times I$. Here is a back-handed way of doing that. Since $\overline{L}=T-\eta_T$ is a disk, $\mathbb{R}^2-\tilde{\eta}$ is a $\mathbb{Z}^2=\mathbb{Z}\times\mathbb{Z}$ lattice of disks. We can then regard $\tilde{\eta}$ as the complement of the slightly fattened lattice $L=\eta(\mathbb{Z}^2)\subset R^2$.

Using this picture, it is easy to describe how a lift of the meridian μ^t of the tunnel intersects R^2. Begin by considering an arc μ connecting the lattice point $(0,0)$ with the lattice point (m,n), where $m,n>0$ are relatively prime. Then the complement of all translates of μ in $\mathbb{R}^2-L$ will be the complement of all translates of the line $my=nx$, namely an infinite collection of bands, each with slope n/m. Each of these bands can also be described as a lift of $\eta_T(K)$ to $\mathbb{R}^2$, if K is the (n,m) torus knot. Thus we see: if K is the (n,m) torus knot, $\mu\times I\subset\tilde{\eta}\times I$ is a lift of the meridian disk μ^t of γ to U.

In a similar spirit, a vertical arc between adjacent points in $\mathbb{Z}^2$ is the lift of a meridian circle of T, and a horizontal arc between adjacent points is the lift of a longitude. Consider the following simple closed curve σ on $\partial\eta(K\cup\gamma)$ (or rather the lift $\tilde{\sigma}$ of σ to $\tilde{\eta}\times I$): $\tilde{\sigma}$ intersects $R^2\times\{0\}$ in two

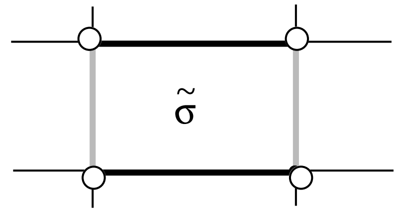

FIGURE 17

adjacent vertical arcs connecting, say, the pair of points $(0,0)$ and $(1,0)$ to the points $(0,1)$ and $(1,1)$. $\tilde{\sigma}$ intersects $R^2 \times \{1\}$ in the two horizontal arcs connecting the pair of points $(0,0)$ and $(0,1)$ to the points $(1,0)$ and $(1,1)$. These two pairs of arcs, one vertical and one horizontal, are then connected to each other by product arcs in $\partial L \times I$. See Figure 17. In particular, the curve $\tilde{\sigma}$ projects to a unit square in R^2.

One can see that σ bounds an essential disk in both $\eta(K \cup \gamma)$ and in its complement. Indeed, it bounds a disk D in $\eta(K \cup \gamma)$ whose lift in $\tilde{\eta} \times I$ projects to the nullhomotopy of the square in the plane. A disk E that σ bounds in $S^3 - \eta(K \cup \gamma)$ can be described as the union of two meridian disks in each solid torus component of $S^3 - \underline{(T \times I)}$ (a total of four disks) each attached along a single arc to a disk in the ball $\overline{L} \times I \subset T \times I$. The disk $E \cap (\overline{L} \times I)$, when projected to I has a single critical point, a saddle.

Now that we have found the tunnel meridian and a splitting sphere $S = D \cup E$, we need to find the preferred meridians of $K - \gamma$, that is, meridians with respect to which $\rho_\perp$ is finite. The following argument is inspired by the proof of [OZ, Lemma 2.2].

Since m, n are relatively prime, there are p, q so that $0 < p < m$ and $0 < q < n$ and $mq - np = 1$. Since

$$det \begin{pmatrix} m & n \\ p & q \end{pmatrix} = 1,$$

it follows that

$$\begin{pmatrix} m & n \\ p & q \end{pmatrix}^{-1}$$

is an integral matrix, so every point of $\mathbb{Z}^2$ is in

$$\begin{pmatrix} m & n \\ p & q \end{pmatrix} \cdot \mathbb{Z}^2.$$

It follows that the parallelogram P with corners

$$\{(0,0), (m-p, n-q), (m,n), (p,q)\}$$

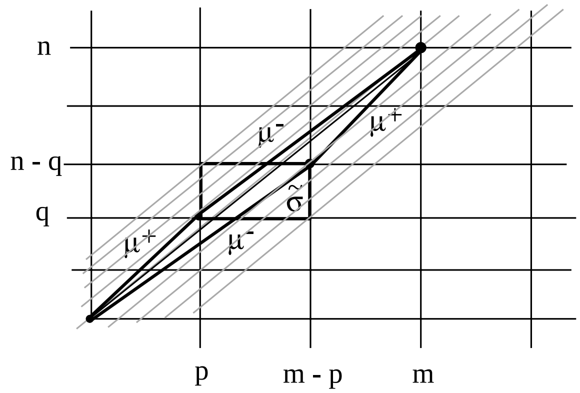

FIGURE 18. The parallelogram P

contains no element of $\mathbb{Z}^2$ in its interior. See Figure 18. In particular, appropriate lifts of the two triangles

$$\Delta((0,0),(m-p,n-q),(m,n))$$

and

$$\Delta((0,0),(p,q),(m,n))$$

tile each band in $\mathbb{R}^2$ that (as was described above) is a universal cover of $\eta_T(K) \subset T$. It follows that the two arcs $\alpha^{\pm}$ in $\mathbb{R}^2$ with ends respectively at $(0,0),(p,q)$ and $(0,0),(m-p,m-q)$, when thickened in U, are lifts of meridians μ^+ and μ^- of $\eta(K)$. It will now be straightforward to show that these are the appropriate meridians for calculating ρ.

We begin with the easy fact that if $a,b,c,d \geq 0$ are integers so that $a > b, c > d$ and $ac - bd = 1$, then $a = c = 1$ and $b = d = 0$. It follows then from $det \begin{pmatrix} m-p & n-q \\ p & q \end{pmatrix} = 1$ that if $m-p > p$ then $n-q \geq q$ and if $m-p < p$ then $n-q \leq q$. Rephrasing this as geometry: the minor diagonal of the parallelogram P, whose ends lie at (p,q) and $(m-p,n-q)$, never has negative slope. It follows that, for each corner (p,q) and $(m-p,n-q)$ (say (p,q)) of P, some lift of the square σ has the property that it intersects P in a triangle, with (p,q) a vertex, and its entire opposite side lying in a single side of P. (See Figure 19.) Translating the plane geometry into the motivating context, the triangle represents a disk D_0, cut off from the interior disk D bounded by σ. D_0 is cut off by the meridian of K represented by the side of P on which the triangle is based. So the two sides of the triangle incident to (p,q) are a wave of ∂D that crosses μ^t, represented by the major diagonal of P, in two points. Thus $\rho(K,\gamma)$ is not only seen to be finite, it is calculated to be $1 \in \mathbb{Q}/2\mathbb{Z}$.

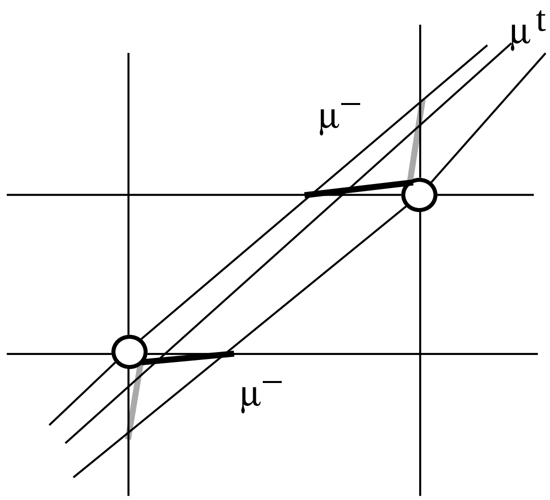

FIGURE 19

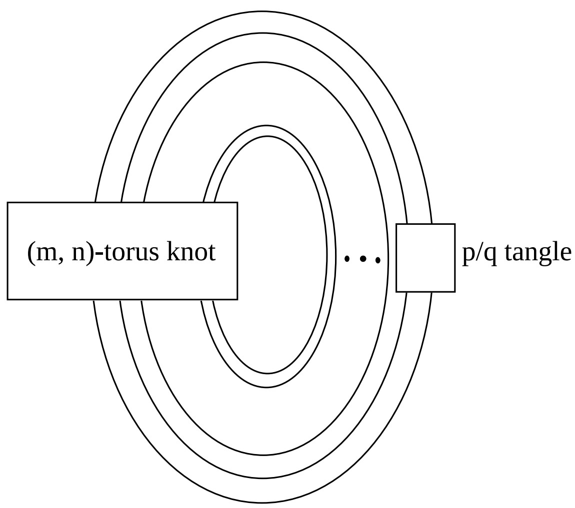

FIGURE 20

It is easy to extrapolate from this calculation to create examples in which ρ can take on any value we like. Namely, replace the two strands of K in a neighborhood of γ by an appropriate rational tangle. See Figure 20.

A torus knot is a simple example of a knot admitting a $(1,1)$-decomposition. That is, it can be written as a 1-bridge knot on an unknotted torus T in S^3 (see [Do], [MS]). Each knot K admitting a $(1,1)$-decomposition has an unknotting tunnel γ' (in fact two of them) best described as the eyeglass obtained by connecting the core γ'_c of a solid torus T bounds to the minimum of the knot $K \subset (T \times I)$ by an ascending arc γ'_a in $T^2 \times I$. (For K a torus knot, γ' is typically different from the tunnel γ above.) An ambitious reader should be able to discover an algorithm which determines, for this eyeglass tunnel γ' and any knot K admitting a $(1,1)$ decomposition, the value of $\rho(K,\gamma')$.

REFERENCES

[Do] H. Doll, *A generalized bridge number for links in* 3*-manifolds*, Math. Ann, **294** (1992) 701–717.

[EU] M. Eudave-Muñoz and Yoshiaki Uchida, *Non-simple links with tunnel number one*, Proc. AMS, **124** (1996) 1567–1575.

[Ga] D. Gabai, *Foliations and the topology of* 3*-manifolds. III*, J. Differential Geom. **26** (1987), no. 3, 479–536.

[GST] H. Goda, M. Scharlemann and A. Thompson, *Levelling an unknotting tunnel*, Geometry and Topology **4** (2000) 243–275.

[GT] H. Goda, M. Teragaito, *Tunnel one genus one non-simple knots*, Tokyo J. Math. **22** (1999) 99–103.

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

[HT] A. Hatcher, W. Thurston, *A presentation for the mapping class group of a closed orientable surface*, Topology **19** (1980), 221–237.

[HR] J. Hempel and L. Roeling, *Free factors of handlebody groups*, unpublished.

[Ma] Hiroshi Matsuda, *Genus one knots which admit* $(1,1)$*-decompositions*, preprint.

[OZ] R. P. Osborne and H. Zieschang, *Primitives in the free group on two generators*, Inventiones Math **63** (1981) 17–24.

[MS] K. Morimoto and M. Sakuma, *On unknotting numbers for knots*, Math. Ann., **289** (1991) 143–167.

[Po] J. Powell, *Homeomorphisms of* S^3 *leaving a Heegaard surface invariant*, Trans. AMS **257** (1980), 193-216.

[Sc1] M. Scharlemann, *Heegaard reducing spheres for the 3-sphere*, Rend. Instit. Mat. Univ. Trieste. **32** (2001), 397-410.

[Sc2] M. Scharlemann, *There are no unexpected tunnel number one knots of genus one*, to appear (cf preprint http://arxiv.org/abs/math.GT/0106017).

[ST] M. Scharlemann and A. Thompson, *Detecting unknotted graphs in* 3*-space*, Jour. Diff. Geom. **34** (1991) 539–560.

MATHEMATICS DEPARTMENT, UNIVERSITY OF CALIFORNIA, SANTA BARBARA, CA 93106, USA

E-mail address: mgscharl@math.ucsb.edu

MATHEMATICS DEPARTMENT, UNIVERSITY OF CALIFORNIA, DAVIS, CA 95616, USA

E-mail address: thompson@math.ucdavis.edu