GENERATING THE GENUS $g+1$ GOERITZ GROUP OF A GENUS g HANDLEBODY

MARTIN SCHARLEMANN

ABSTRACT. A specific set of $4g + 1$ elements is shown to generate the Goeritz group of the genus $g + 1$ Heegaard splitting of a genus g handlebody. These generators are consistent with Powell's proposed generating set for the Goeritz group of the genus $g + 1$ splitting of S^3. There are two proofs: one using purely classical techniques and one using thin position.

1. INTRODUCTION

Following early work of Goeritz [Go], the *genus g Goeritz group* of the 3-sphere can be described as the isotopy classes of orientation-preserving homeomorphisms of the 3-sphere that leave the genus g Heegaard splitting invariant. Goeritz identified a finite set of generators for the genus 2 Goeritz group; that work has been recently updated, extended and completed, to give a full picture of the group (see [Sc], [Ak], [Cho]). Goeritz' set of generators was extended by Powell [Po] to a set of generators for all higher genus Goeritz groups, but his proof that the generators suffice contained a gap [Sc]. The finite set of elements that Powell proposed as generators for the full Goeritz group remains a very plausible set, though a proof remains elusive.

One intriguing aspect of the problem is that Gabai's powerful technique of thin position [Ga] is available for objects in S^3, such as Heegaard splitting surfaces (see [ST2]), but the technique was not known to Powell. In addition, one can imagine structuring a proof by induction on the "co-genus" k of Heegaard splittings of a handlebody: any genus g Heegaard splitting of a genus $g - k$ handlebody H gives rise to a genus g splitting of S^3, by identifying H with one of the handlebodies in the standard genus $g - k$ splitting of S^3. In that context, Powell's conjecture would suggest a natural set of generators for the genus g Goeritz group of a genus $g - k$ handlebody (see Section 5 for the definition). As k ascends we eventually have a set of generators for the genus g Goeritz group of the genus 0 handlebody B^3 (or, equivalently,

S^3). With that strategy in mind, here we verify Powell's conjecture for the first and easiest case, namely co-genus 1. Rephrasing slightly, we exhibit, for any genus g handlebody H, a certain finite set of elements that generates the genus $g+1$ Goeritz group $G(H, \Sigma)$ of H. Combining the results of Theorems 3.2 and 5.1, to which we refer for notation, we show:

Theorem 1.1. *The Goeritz group $G(H, \Sigma)$ of the genus g handlebody H is generated by $4g+1$ elements, namely $2g$ generators of the subgroup $\mathfrak{A}_{\{E_1,...,E_g\}}$ and $2g + 1$ generators of the subgroup $\mathfrak{F}_{E_0}$.*

We will give two proofs that these generators suffice: the first is along classical lines (i. e. without thin position) and the second uses thin position. Both arguments are given in a slightly different setting – the isotopies are of an unknotted arc in the handlebody, rather than a Heegaard surface – but the connection between the two is explained in Section 5.

2. Embedding an unknotted arc in a ball

For M, N smooth manifolds, let $\mathrm{Emb}(M, N)$ denote the space of smooth proper embeddings of M into N. Let $\mathrm{Emb}_0(I, B^3) \subset \mathrm{Emb}(I, B^3)$ denote the path-component consisting of those embeddings for which the image is an unknotted arc. There is a natural fibration $\mathrm{Emb}_0(I, B^3) \to \mathrm{Emb}(\partial I, \partial B^3)$ whose fiber is $\mathrm{Emb}_0(I, B^3 \ rel \ \partial I)$ [Pa]. Following Hatcher's proof of the Smale conjecture, this fiber is contractible [Ha2, Appendix (6)], so in particular $\pi_1(\mathrm{Emb}_0(I, B^3)) \cong \pi_1(\mathrm{Emb}(\partial I, \partial B^3))$. The space $\mathrm{Emb}(\partial I, \partial B^3)$ is the configuration space $F_2(S^2)$ of ordered pairs of points in the sphere; its fundamental group is the pure braid group of two points in S^2, which is trivial. Hence $\mathrm{Emb}_0(I, B^3)$ is simply connected. By taking each element of $\mathrm{Emb}_0(I, B^3)$ to its image in B^3 we get a natural map $\mathrm{Emb}_0(I, B^3) \to \mathrm{Unk}(I, B^3)$, the space of unknotted arcs in B^3; its fiber is the space of automorphisms of the interval I, which consists of two contractible components, representing orientation preserving and orientation reversing automorphisms of the interval. Combining these two observations we discover that the natural map from $\mathrm{Unk}(I, B^3)$ to the configuration space $C_2(S^2)$ of unordered pairs of points in S^2 induces an isomorphism between the respective fundamental groups. Note that $\pi_1(C_2(S^2))$ is commonly called the full braid group $B_2(S^2)$. We conclude that $\pi_1(\mathrm{Unk}(I, B^3)) \cong B_2(S^2) \cong \mathbb{Z}_2$, [Bi, Theorem 1.11].

Now suppose P is a connected planar surface in $\partial(B^3)$ and $\mathrm{Unk}_P(I, B^3)$ is the space of all unknotted arcs in B^3 whose end points lie in P. Exactly the same argument as above shows that $\pi_1(\mathrm{Unk}_P(I, B^3)) \cong$

$B_2(P)$, where the latter is the full braid group of two points in P. It is straightforward to identify a set of generators for $B_2(P)$. Begin with the ordered configuration space $C_2(P)$ and project to the first point $x_0 \in P$. The map is a fibration $p : C_2(P) \to C_1(P) = P$ whose fiber is $C_1(P - \{point\}) = P - \{point\}$ [FN]. Since P is connected and $\pi_2(P)$ is trivial, it follows that $\pi_1(C_2(P))$ is an extension of $\pi_1(P)$ by $\pi_1(P - \{point\})$ and each of these groups admits a well-known collection of generators, one for each boundary component of P. Namely, for each boundary component choose a loop from the base point that is parallel in $P - \{point\}$ to that component. One of these generators is redundant in $\pi_1(P)$; all are needed in $\pi_1(P - \{point\})$. To complete this set of generators to a set of generators for $B_2(P)$, add an isotopy of the pair of points that interchanges the pair.

These rather abstract descriptions translate to this concrete description of a set of generators for $\pi_1(\mathrm{Unk}_P(B^3)) \cong B_2(P)$: Let α be a short arc in P; its endpoints x_0, x_1 will be the pair of points whose motion we are describing. Half rotation of α around its center, exchanging its ends is one generator for $B_2(P)$; call it the *rotor* ρ_0. Let $c_1, ..., c_p$ be the boundary components of P and for each c_i choose a loop γ_i in P that passes through x_1 and is parallel in $P - x_0$ to c_i. Choose these loops so that they intersect each other or α only in the point x_1 (see Figure 1). For each $1 \le i \le p$ let ρ_i be an isotopy that moves the entire arc α through a loop in P parallel to γ_i and back to itself. This defines each ρ_i up to multiples of the rotor ρ_0, so the subgroup $\mathfrak{F}$ of $B_2(P)$ defined as that generated by $\rho_i, 0 \le i \le p$ is in fact well-defined. Call $\mathfrak{F}$ the *freewheeling* subgroup. $\mathfrak{F}$ is an extension of $\pi_1(P)$ by $\mathbb{Z}$.

A second subgroup $\mathfrak{A} \subset \pi_1(\mathrm{Unk}_P(B^3)) \cong B_2(P)$, called the *anchored* subgroup, is defined as those elements which keep the "anchor" end x_0 of α fixed as the other end follows a closed path in $P - \{x_0\}$ that begins and ends at x_1. It corresponds to the fundamental group of the fiber $P - \{point\}$ in the above fibration. More concretely, for each $1 \le i \le p$ let $\mathfrak{a}_i$ denote the element determined by keeping x_0 fixed and moving x_1 around the loop γ_i. The subgroup $\mathfrak{A}$ is generated by the $\mathfrak{a}_i$; it includes any even power of the rotor ρ_0, via the relation (in Figure 1) $\mathfrak{a}_1\mathfrak{a}_2...\mathfrak{a}_\mathfrak{p} = \rho^2$.

The fibration above shows that together $\mathfrak{A}$ and $\mathfrak{F}$ generate the group $\pi_1(\mathrm{Unk}_P(B^3))$, so $\{\rho_i, 0 \le i \le p\}$ and $\{\mathfrak{a}_i, 1 \le i \le p\}$ together constitute a set of generators for $\pi_1(\mathrm{Unk}_P(B^3))$.
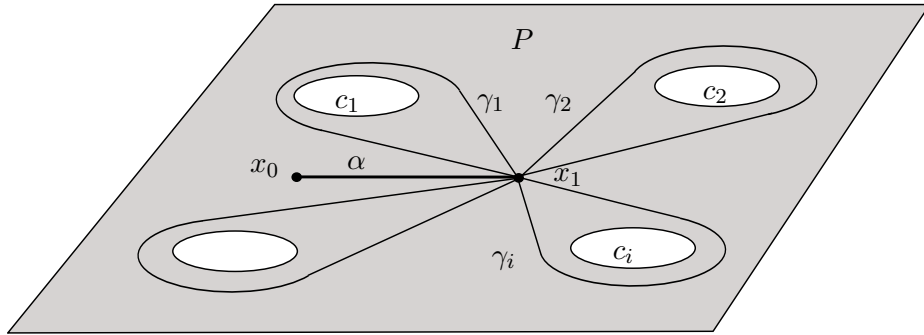
FIGURE 1. Generating $\mathfrak{A} \subset B_2(P)$

3. UNKNOTTED ARCS IN A HANDLEBODY

The goal of this section is to extend this analysis to describe, for a genus g handlebody H, a fairly natural set of generators for the fundamental group of the space $\mathrm{Unk}(I, H)$ of unknotted properly embedded arcs in H.

We begin with a basepoint for $\mathrm{Unk}(I, H)$, i. e. a fixed choice of unknotted arc in H. This is facilitated by viewing H as the product of a planar surface with I: Let Q be a disk D from which g disks $D_1, ..., D_g$ have been removed. Picture the disks D_i as laid out in a horizontal row in D, with a vertical arc $\beta_i, 1 \leq i \leq g$ descending from each $\partial D_i \subset \partial Q$ to ∂D. Further choose a point $x \in inter(Q) - \cup_i \beta_i$ to the left of the disks D_i and connect it to ∂D by a horizontal arc β_0. See Figure 2. Then $Q \times I$ is a handlebody in which $x \times I$ is an unknotted arc I_0 in H with end points $x_i = x \times \{i\}, i = 0, 1$. Furthermore, the g disks $E_i = \beta_i \times I \subset H, i = 1, ..., g$ constitute a complete collection of meridian disks for H. That is, the complement in H of a regular neighborhood $\eta(\cup_{i=1}^{g} E_i)$ of $\cup_{i=1}^{g} E_i$ is a 3-ball B^3 which intersects ∂H in a planar surface P. The boundary of P has $2g$ components, two copies of each $\partial E_i, i = 1, ..., g$.

The disk $E_0 = \beta_0 \times I$ defines a parallelism between the arc I_0 and an arc $\alpha \subset \partial H$. Such a disk will be called a *parallelism disk* for I_0 and the subarc of its boundary that lies on ∂H will be called a *parallel arc* for I_0. It will be convenient, when considering a pair E_0, E_0' of parallelism disks and corresponding parallel arcs $\alpha_0 = E_0 \cap \partial H, \alpha_0' = E_0' \cap \partial H$ for I_0, to isotope the disks so that they are transverse except where they coincide along I_0, and so that they have disjoint interiors near I_0. (This is done by unwinding E_0' along I_0). A standard innermost disk argument shows that the simple closed curves in $E_0 \cap E_0'$ can be removed by an isotopy that does not move I_0, after which what remains
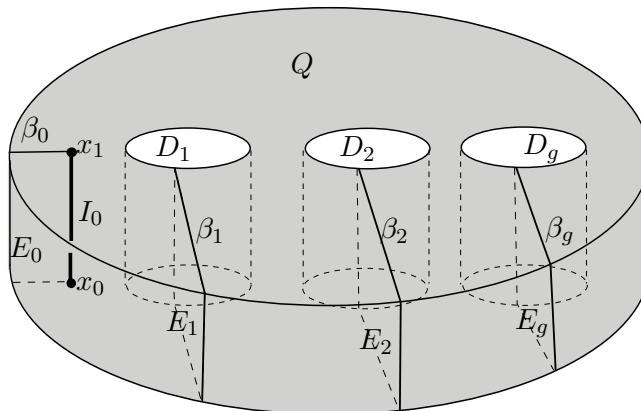
FIGURE 2. A genus g handlebody H with trivial arc I_0

of $E_0 \cap E_0'$ is their common boundary arc I_0 together with a collection of interior arcs whose endpoints are the points of $\alpha_0 \cap \alpha_0'$. Call this a *normal position* for two parallelism disks.

Motivated by the discussion above, we note some obvious elements and subgroups of $\pi_1(\text{Unk}(I, H))$: Since the pair (B^3, P) is a subset of $(H, \partial H)$ there is a natural inclusion-induced homomorphism $\pi_1(\text{Unk}_P(I, B^3)) \to \pi_1(\text{Unk}(I, H))$. For example, a natural picture of the rotor ρ_0 in $\pi_1(\text{Unk}(I, H))$ is obtained by doing a half-twist of I_0 in a 3-ball neighborhood of the disk $\beta_0 \times I$. This is shown on the left in Figure 3. The image of the anchored subgroup $\mathfrak{A} \subset B_2(P)$ in $\pi_1(\text{Unk}(I, H))$ can be defined much like the anchored subgroup in $B_2(P)$ itself: hold the end of I_0 at x_0 fixed while isotoping the end at x_1 so that the whole arc I moves around and back to its original position, never letting the moving I intersect any of the g disks E_i. We denote this subgroup $\mathfrak{A}_{\{E_1,...,E_g\}} \subset \pi_1(\text{Unk}(I, H))$. Two of its generators are shown center and right in Figure 3. There is also a naturally defined *freewheeling* subgroup $\mathfrak{F}_{E_0} \subset \pi_1(\text{Unk}(I, H))$ consisting of those elements represented by a proper isotopy of the disk E_0 through H and back to itself (though perhaps with orientation reversed). Thus again the rotor ρ_0 lies in $\mathfrak{F}_{E_0}$, and the kernel of $\mathfrak{F}_{E_0} \to \pi_1(\partial H)$ is generated by the rotor ρ_0. Since $\pi_1(\partial H)$ is itself generated by $2g$ elements (essentially given by the choice of $\{E_1,...,E_g\}$), $\mathfrak{F}_{E_0}$ is generated by $2g + 1$ elements.

Suppose $\omega \in \pi_1(\text{Unk}(I, H))$ is represented by a proper isotopy $f_t : I \to H$. The isotopy extends to an ambient isotopy of H which we continue to denote f_t; let $\alpha_\omega \subset \partial H$ denote $f_1(\alpha_0)$. Since $f_1(I_0) = I_0$,
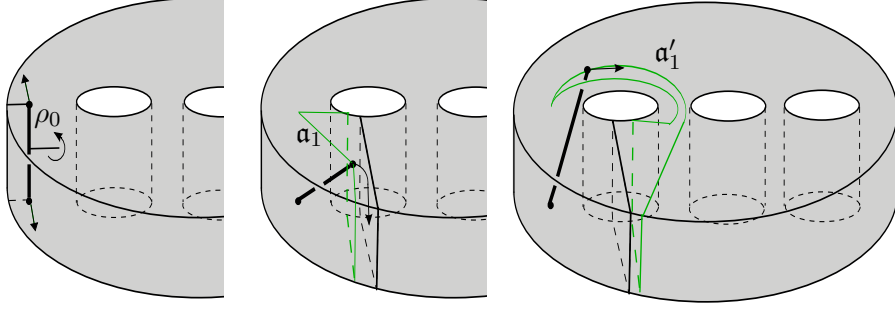
FIGURE 3. The rotor ρ_0 and generators $\mathfrak{a}_1, \mathfrak{a}_1' \in \mathfrak{A}$.

$f_1(E_0)$ is a new parallelism disk, and α_ω is the corresponding parallel arc for I_0 in ∂H.

Lemma 3.1. *The parallel arcs α_0 and α_ω for I_0 are isotopic rel end points in ∂H if and only if $\omega \in \mathfrak{F}_{E_0}$.*

Proof. If $\omega \in \mathfrak{F}_{E_0}$ then by definition $\alpha_\omega = \alpha_0$. On the other hand, if α_ω is isotopic to α_0 rel end points, then we may as well assume $\alpha_\omega = \alpha_0$, for the isotopy from α_ω to α_0 doesn't move I_0. Then, thickening I_0 slightly to $\eta(I_0)$, E_0 and $f_1(E_0)$ are properly embedded disks in the handlebody $H - \eta(I_0)$ and have the same boundary. A standard innermost disk argument shows that then $f_1(E_0)$ may be isotoped rel ∂ (so, in particular, the isotopy leaves I_0 fixed) until $f_1(E_0)$ is disjoint from E_0. Since their boundaries are the same, and the handlebody $H - \eta(I_0)$ is irreducible, a final isotopy can ensure that $f_1(E_0)$ coincides with E_0, revealing that $\omega \in \mathfrak{F}$. $\qquad\square$

A sequence of further lemmas will show:

Theorem 3.2. *The subgroups $\mathfrak{A}_{\{E_1,\ldots,E_g\}}$ and $\mathfrak{F}_{E_0}$ together generate all of $\pi_1(\mathrm{Unk}(I,H))$, so the union of their generators is an explicit set of generators for $\pi_1(\mathrm{Unk}(I,H))$.*

What is perhaps surprising about this theorem is that the subgroups themselves depend heavily on our choice of the disks $\{E_0,\ldots,E_g\}$. Recognizing this dependence, let the combined symbol $\mathfrak{A}\mathfrak{F}_{\{E_0,\ldots,E_g\}}$ denote the subgroup of $\pi_1(\mathrm{Unk}(I,H))$ generated by $\mathfrak{A}_{\{E_1,\ldots,E_g\}}$ and $\mathfrak{F}_{E_0}$.

Lemma 3.3. *Suppose $E_0' \subset H$ is another parallelism disk that lies entirely in $H - \{E_1,\ldots,E_g\}$. Then $\mathfrak{A}\mathfrak{F}_{\{E_0',E_1,\ldots,E_g\}} = \mathfrak{A}\mathfrak{F}_{\{E_0,E_1,\ldots,E_g\}}$.*

Proof. Put the pair E_0, E_0' in normal position. The proof is by induction on the number $|E_0 \cap E_0'|$ of arcs in which their interiors intersect.

If $|E_0 \cap E_0'| = 0$, so the disks intersect only in I_0, then let $E_\cup \subset H - \{E_1, ..., E_g\}$ be the properly embedded disk that is their union. Rotating one end of I_0 fully around a slightly pushed-off copy of $E_\cup$ describes an element $\mathfrak{a} \in \mathfrak{A}_{\{E_1,...,E_g\}}$ for which a representative isotopy carries the disk E_0 to E_0'. See Figure 4. In particular, if f is any element of $\mathfrak{F}_{E_0}$ then the product $\mathfrak{a}^{-1} f \mathfrak{a}$ has a representative isotopy which carries E_0' to itself. Hence $\mathfrak{a}^{-1} f \mathfrak{a} \in \mathfrak{F}_{E_0'}$ so $f \in \mathfrak{A}\mathfrak{F}_{\{E_0', E_1, ..., E_g\}}$. Thus $\mathfrak{F}_{E_0} \subset \mathfrak{A}\mathfrak{F}_{\{E_0', E_1, ..., E_g\}}$, so $\mathfrak{A}\mathfrak{F}_{\{E_0, E_1, ..., E_g\}} \subset \mathfrak{A}\mathfrak{F}_{\{E_0', E_1, ..., E_g\}}$. The symmetric argument shows that $\mathfrak{A}\mathfrak{F}_{\{E_0', E_1, ..., E_g\}} \subset \mathfrak{A}\mathfrak{F}_{\{E_0, E_1, ..., E_g\}}$ and so $\mathfrak{A}\mathfrak{F}_{\{E_0, E_1, ..., E_g\}} = \mathfrak{A}\mathfrak{F}_{\{E_0', E_1, ..., E_g\}}$ in this case.
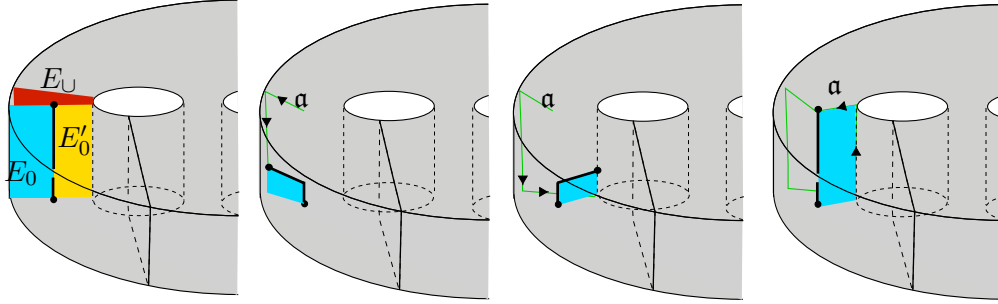

FIGURE 4. $\mathfrak{a} \in \mathfrak{A}$: circling around $E_\cup$ brings E_0 to E_0'.

The argument just given shows that for any pair of parallelism disks $F_0, F_0' \subset H - \{E_1, ..., E_g\}$ which intersect only along I_0, $\mathfrak{A}\mathfrak{F}_{\{F_0', E_1, ..., E_g\}} = \mathfrak{A}\mathfrak{F}_{\{F_0, E_1, ..., E_g\}}$. Suppose inductively that this is true whenever $|F_0 \cap F_0'| \leq k$ and, for the inductive step, suppose that $|E_0 \cap E_0'| = k + 1$. Among all arcs of $E_0 \cap E_0'$, let β be outermost in E_0', so that the subdisk $E^* \subset E_0'$ cut off by β does not contain I_0 in its boundary, nor any other point of E_0 in its interior. Then attaching E^* along β to the component of $E_0 - \beta$ that contains I_0 gives a parallelism disk F_0 that is disjoint from E_0 and intersects E_0' in $\leq k$ arcs. It follows by inductive assumption that $\mathfrak{A}\mathfrak{F}_{\{E_0, E_1, ..., E_g\}} = \mathfrak{A}\mathfrak{F}_{\{F_0, E_1 ..., E_g\}} = \mathfrak{A}\mathfrak{F}_{\{E_0', E_1, ..., E_g\}}$ as required. $\square$

Following Lemma 3.3 there is no loss in dropping E_0 from the notation, so $\mathfrak{A}\mathfrak{F}_{\{E_0, E_1, ..., E_g\}}$ will henceforth be denoted simply $\mathfrak{A}\mathfrak{F}_{\{E_1, ..., E_g\}}$.

Lemma 3.4. *Suppose $E_* \subset H$ is a disk in $H - (I_0 \cup E_1 \cup ... \cup E_g\}$, so that $\{E_*, E_2, ..., E_g\}$ is a complete set of meridian disks for H. Then $\mathfrak{A}\mathfrak{F}_{\{E_1, E_2, ..., E_g\}} = \mathfrak{A}\mathfrak{F}_{\{E_*, E_2, ..., E_g\}}$. That is, the subgroup $\mathfrak{A}\mathfrak{F}_{\{-, E_2, ..., E_g\}}$ is the same, whether we fill in E_1 or E_*.*

Proof. All $g + 1$ meridian disks $\{E_*, E_1, E_2, ..., E_g\}$ are mutually disjoint and all are disjoint from the proper arc I_0. Although an arbitrary

parallelism disk for I_0 may intersect $\{E_*, E_1, E_2, ..., E_g\}$, a standard
innermost-disk, outermost-arc argument can be used to diminish the
number of components of intersection until a parallelism disk E_0 is
found that is completely disjoint from $\{E_*, E_1, E_2, ..., E_g\}$. Use E_0 to
define both $\mathfrak{AF}_{\{E_1,E_2,...,E_g\}}$ and $\mathfrak{AF}_{\{E_*,E_2,...,E_g\}}$. It follows that $\mathfrak{F}_{E_0}$ is
a subgroup of both $\mathfrak{AF}_{\{E_1,E_2,...,E_g\}}$ and $\mathfrak{AF}_{\{E_*,E_2,...,E_g\}}$. Hence it suf-
fices to show that $\mathfrak{A}_{\{E_*,E_2,...,E_g\}} \subset \mathfrak{AF}_{\{E_1,E_2,...,E_g\}}$ and $\mathfrak{A}_{\{E_1,E_2,...,E_g\}} \subset$
$\mathfrak{AF}_{\{E_*,E_2,...,E_g\}}$. We will prove the latter; the former follows by a sym-
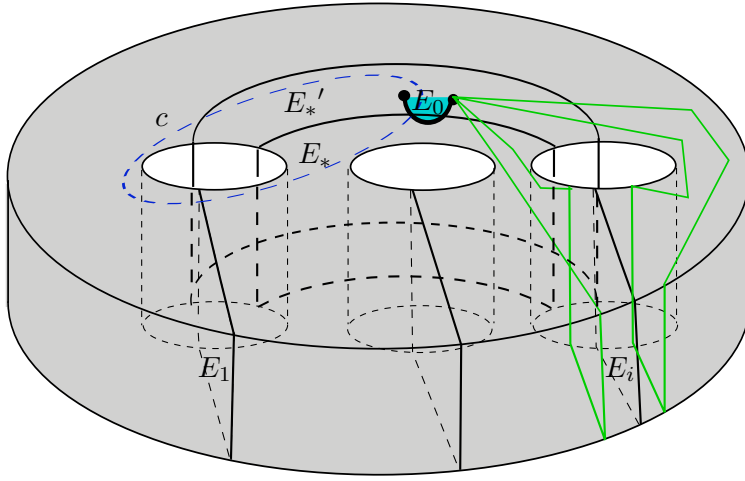metric argument.

FIGURE 5

Extend a regular neighborhood of $\cup_{i=2}^g E_i$ to a regular neighborhood
Y of $\cup_{i=1}^g E_i$ and Y^* of $E_* \cup (\cup_{i=2}^g E_i)$. The disk E_* is necessarily sep-
arating in the ball $B^3 = H - Y$ and, since $H - Y^*$ is also a ball, it
follows that the two sides of E_1 in ∂B^3 lie in different components of
$B^3 - E_*$. Put another way, there is a simple closed curve c in ∂H which
is disjoint from $\{E_2, E_3, ..., E_g\}$ but intersects each of E_1 and E_* in a
single point. Let $f \in \mathfrak{F}_{E_0}$ be the element represented by isotoping E_0
around the circle c in the direction so that it first passes through E_1
and then through E_*.

The image of E_*, after the isotopy f is extended to H, is a disk E_*'
that is isotopic to E_* in B^3 but not in $B^3 - I_0$. The isotopy need not
disturb the disks $\{E_2, E_3, ..., E_g\}$. Put another way, there is a collar
between E_* and E_*' in B^3, a collar that contains both the trivial arc
I_0 and the parallelism disk E_0 but is disjoint from $\{E_2, E_3, ..., E_g\}$. See
Figure 5. As before, let P denote the planar surface $\partial H - Y$, that is
the planar surface obtained from ∂H by deleting a neighborhood of the
meridian disks $\{E_1, E_2, ..., E_g\}$. Modeling the concrete description of

generators of $\mathfrak{A} \subset B_2(P)$ given via Figure 1, the definition of $\mathfrak{A}_{\{E_1,...,E_g\}}$ begins with a collection of loops $\gamma_i, \gamma_i' \subset P, 1 \leq i \leq g$ so that all the loops are mutually disjoint, except in their common end points at x_1; for each i, one of γ_i and γ_i' is parallel in P to each of the two copies of ∂E_i in ∂P; each loop is disjoint from ∂E_0 except at x_1; and (what is new) each loop intersects only one end of the collar that lies between ∂E_* and $\partial E_*'$. Then $2g$ generators $\mathfrak{a}_i, \mathfrak{a}_i', 1 \leq i \leq g$ of $\mathfrak{A}_{\{E_1,...,E_g\}}$ are represented by isotopies obtained by sliding the endpoint x_1 of I_0 around the loops γ_i and γ_i' respectively.

If γ_i (resp γ_i') is one of the loops disjoint from E_*, then $\mathfrak{a}_i$ (resp $\mathfrak{a}_i'$) also lies in $\mathfrak{A}_{\{E_*,...,E_g\}}$. If, on the other hand, γ_i (resp γ_i') is one of the loops that is disjoint from E_*', then $f\mathfrak{a}_i f^{-1}$ (resp $f\mathfrak{a}_i' f^{-1}$) is represented by an isotopy of I_0 that is disjoint from E_*. Moreover, since an isotopy of I_0 representing f doesn't disturb the disks $\{E_2, E_3, ..., E_g\}$, the isotopy of I_0 representing $f\mathfrak{a}_i f^{-1}$ (resp $f\mathfrak{a}_i' f^{-1}$) is disjoint from these disks as well. That is, each such $f\mathfrak{a}_i f^{-1}$ (resp $f\mathfrak{a}_i' f^{-1}$) lies in $\mathfrak{A}_{\{E_*,...,E_g\}}$. Hence in all cases, $\mathfrak{a}_i$ (resp $\mathfrak{a}_i'$) lies in $\mathfrak{A}\mathfrak{F}_{\{E_*,E_2,...,E_g\}}$, so $\mathfrak{A}_{\{E_1,...,E_g\}} \subset \mathfrak{A}\mathfrak{F}_{\{E_*,E_2,...,E_g\}}$. $\square$

It is well-known that in a genus g handlebody any two complete collections of g meridian disks can be connected by a sequence of complete collections of meridian disks so that at each step in the sequence a single meridian disk is replaced with a different and disjoint one. See, for example, [Wa, Theorem 1]. It follows then from Lemma 3.4 that the subgroup $\mathfrak{A}\mathfrak{F}_{\{E_1,E_2,...,E_g\}}$ is independent of the specific collection of meridian disks, so we can simply denote it $\mathfrak{A}\mathfrak{F}$.

Theorem 3.5. *The inclusion $\mathfrak{A}\mathfrak{F} \subset \pi_1(\mathrm{Unk}(I,H))$ is an equality.*

Proof. Begin with a parallelism disk E_0 for I_0, with α_0 the arc $E_0 \cap \partial H$ connecting the endpoints x_0 and x_1. Suppose $\omega \in \pi_1(\mathrm{Unk}(I,H))$ is represented by a proper isotopy $f_t : I \to H$ extending to the ambient isotopy $f_t : H \to H$. Adjust the end of the ambient isotopy so that E_0 and $f_1(E_0)$ are in normal position. Let $\alpha_\omega \subset \partial H$ denote the image $f_1(\alpha_0)$, an arc in ∂H that also connects x_0 and x_1. The proof is by induction on $|\alpha_0 \cap \alpha_\omega|$, the number of points in which the interiors of α_0 and α_ω intersect. The number is always even, namely twice the number $|E_0 \cap f_1(E_0)|$ of arcs of intersection of the disk interiors.

If $|\alpha_0 \cap \alpha_\omega| = 0$, so α_0 and α_ω intersect only in their endpoints at x_0 and x_1, then the union of α_0 and α_ω is a simple closed curve in ∂H that bounds a (possibly inessential) disk $E_\cup$. The disk $E_\cup$ properly contains I_0 and is properly contained in H. Choose a complete collection of meridian disks $\{E_1, ..., E_g\}$ for H that is disjoint from $E_\cup$. Then an

isotopy of one end of I_0 completely around $\partial E_\cup$ (pushed slightly aside) represents an element $\mathfrak{b}$ of $\mathfrak{A}_{\{E_1,...,E_g\}}$ and takes α_ω to α_0. See the earlier Figure 4. It follows from Lemma 3.1 that the product $\omega\mathfrak{b}$ is in $\mathfrak{F}_{E_0}$, so $\omega \in \mathfrak{A}\mathfrak{F}$.

For the inductive step, suppose that any element in $\pi_1(\mathrm{Unk}(I,H))$ whose corresponding isotopy carries some I_0-parallel arc α to an arc that intersects α in k or fewer points is known to lie in $\mathfrak{A}\mathfrak{F}$. Suppose also that $|\alpha_0 \cap \alpha_\omega| = k + 2$. Let E^* be a disk in $f_1(E_0) - E_0$ cut off by an outermost arc β of $E_0 \cap f_1(E_0)$ in $f_1(E_0)$. Then attaching E^* along β to the component of $E_0 - \beta$ that contains I_0 gives a parallelism disk F_0 that is disjoint from E_0 and intersects α_ω in $\leq k$ points. The union of F_0 and E_0 along I_0 is a properly embedded disk $E_\cup$ and, as usual, there is an isotopy representing an element η of $\mathfrak{A}\mathfrak{F}$ that carries F_0 to E_0. Now apply the inductive assumption to the product $\eta\omega$: The isotopy corresponding to $\eta\omega$ carries the arc $F_0 \cap \partial H$ to α_ω. It follows by inductive assumption then that $\eta\omega \in \mathfrak{A}\mathfrak{F}$. Hence also $\omega \in \mathfrak{A}\mathfrak{F}$, completing the inductive step. $\square$

Theorem 3.2 is then an obvious corollary, and provides an explicit set of generators for $\pi_1(\mathrm{Unk}(I,H))$.

4. CONNECTION TO WIDTH

Suppose $f_t : I \to H$ is a proper isotopy from I_0 back to itself, representing an element $\omega \in \pi_1(\mathrm{Unk}(I,H))$. Put f_t in general position with respect to the collection Δ of meridian disks $\{E_1,...E_g\}$ so f_t is transverse to Δ (in particular, $f(\partial I) \cap \Delta = \emptyset$) at all but a finite number $0 < t_0 \leq t_1 \leq ... \leq t_n < 1$ of values of t, which we will call the critical points of the isotopy f_t. Let $c_i, 1 \leq i \leq n$ be any value so that $t_{i-1} \leq c_i \leq t_i$ and define $w_i = |f_{c_i}^{-1}(\Delta)| = |f_{c_i}(I) \cap \Delta|$ to be the width of I at c_i. The values $w_i, 1 \leq i \leq n$ are all independent of the choice of the points $c_i \in (t_{i-1}, t_i)$ since the value of $|f_t(I) \cap \Delta|$ can only change at times when f_t is not transverse to Δ.

Definition 4.1. *The width $w(f_t)$ of the isotopy f_t is $max_i\{w_i\}$. The width $w(\omega)$ of $\omega \in \pi_1(\mathrm{Unk}(I,H))$ is the minimum value of $w(f_t)$ for all isotopies f_t that represent ω.*

Corollary 4.2. *For any $\omega_1, \omega_2 \in \pi_1(\mathrm{Unk}(I,H))$, the width of the product $w(\omega_1\omega_2) \leq max\{w(\omega_1), w(\omega_2)\}$.*

It follows that, for any $n \geq 0$, the set of elements of $\pi_1(\mathrm{Unk}(I,H))$ of width no greater than n constitutes a subgroup of $\pi_1(\mathrm{Unk}(I,H))$. For example, if $n = 0$ the subgroup is precisely the image of the inclusion-induced homomorphism $\pi_1(\mathrm{Unk}_P(I,B^3)) \to \pi_1(\mathrm{Unk}(I,H))$ defined at

the beginning of Section 3, a subgroup that includes the anchored subgroup $\mathfrak{A}_\Delta$. It is easy to see that the width of any element in $\mathfrak{F}_{E_0}$ is at most 1, so it follows from Corollary 4.2 and Theorem 3.2, that the width of any element in $\pi_1(\mathrm{Unk}(I, H)) = \mathfrak{A}\mathfrak{F}_{\{E_0, E_1, ..., E_g\}}$ is at most 1.

The fact that every element in $\pi_1(\mathrm{Unk}(I, H))$ is at most width 1 suggests an alternate path to a proof of Theorem 3.2, a path that would avoid the technical difficulties of Lemmas 3.3 and 3.4: prove directly that any element of width 1 is in $\mathfrak{A}\mathfrak{F}_{\{E_0, E_1, ..., E_g\}}$ and prove directly that any element in $\pi_1(\mathrm{Unk}(I, H))$ has width at most 1 (say by thinning a given isotopy as much as possible) . We do so below. Neither argument requires a change in the meridian disks $\{E_1, ..., E_g\}$. The first argument, that any element of width 1 is in $\mathfrak{A}\mathfrak{F}_{\{E_0, E_1, ..., E_g\}}$, is the sort of argument that might be extended to isotopies of unknotted graphs in H, not just isotopies of the single unknotted arc I, just as other thin position arguments have been extended to graphs (see [ST1], [ST2]). The second argument, which shows that the thinnest representation of any element in $\pi_1(\mathrm{Unk}(I, H))$ is at most width 1, seems difficult to generalize to isotopies of an arbitrary unknotted graph, because the argument doesn't directly thin a given isotopy, but rather makes use of Lemma 3.1 in a way that may be limited to isotopies of a single arc.

Proposition 4.3. *If an element $\omega \in \pi_1(\mathrm{Unk}(I, H))$ has $w(\omega) = 1$ then $\omega \in \mathfrak{A}\mathfrak{F}_{\{E_0, E_1, ..., E_g\}}$*

Proof. The proof requires a series of lemmas.

Lemma 4.4. *It suffices to consider only those $\omega \in \pi_1(\mathrm{Unk}(I, H))$ represented by isotopies $f_t : I \to H$ with exactly two critical points.*

Proof. Suppose $w(\omega) = 1$ and $f_t : I \to H$ is an isotopy realizing this width. Let $0 < t_0 \leq t_1 \leq ... \leq t_n < 1$ be the critical points of the isotopy and, as defined above, let w_i be the width of I during the ith interval. Since each w_i is either 0 or 1 and $w_i \neq w_{i-1}$ it follows that the value of w_i alternates between 0 and 1. Since $I_0 \cap \Delta = \emptyset$ the value of $w_1 = 1$. During those intervals when the width is 0, $f_t(I)$ is disjoint from Δ, so the isotopy f_t can be deformed, without altering the width, so that at some time t during each such interval, $f_t(I) = I_0$. Thereby ω can be viewed as the product of elements, for each of which there is an isotopy with just two critical points. $\square$

So we henceforth assume that $\omega \in \pi_1(\mathrm{Unk}(I, H))$ is represented by an isotopy f_t so that $n = 1$ and there are just two critical points during the isotopy, t_0 and t_1. If either point were critical because of a tangency between $f_{t_i}(I)$ and Δ, then the value of w_i would change by

2 as the tangent point passed through Δ, and this is impossible by the assumption in this case. We conclude that t_0 and t_1 are critical because they mark the point at which f_t moves a single end of I through Δ, say through the meridian disk $E_1 \subset \Delta$. It is possible that at t_0 and t_1 the same end of I moves through E_1 (so I does not pass completely through E_1), or possibly different ends of I move through E_1 (when I does pass completely through E_1). We will refer to these two types of isotopies as a *bounce* and a *pass*.

Pick a generic value $t_0 < t' < t_1$. There is a natural way to associate to the isotopy f_t two disks, D_0, D_1, called *tracking disks* for f_t with these properties (see Figures 6 and 7):

- Each D_i is embedded in H and has interior disjoint from $f_{t'}(I)$.
- Each ∂D_i consists of the end-point union of three arcs: an arc in E_1, an arc in ∂H and the subarc J_i of $f_{t'}(I)$ lying between the point $p = E_1 \cap f_{t'}(I)$ and the end of $f_{t'}(I)$ that passes through E_1 at $t_i, i = 0, 1$.
- The subarcs of ∂D_0 and ∂D_1 that lie on E_1 coincide. Along this arc α the disks D_i meet transversally.
- $D_0 \cap D_1$ consists of a collection of properly embedded arcs, each ending in a point on ∂H.

Note that, by definition, $J_0 = J_1$ for a bounce and $f_{t'}(I) = J_0 \cup J_1$ for a pass. Here is the construction:

Recall that $f_{t_0}(I)$ intersects E_1 exactly in a single endpoint. Since the isotopy is generic, at time $t_0 + \epsilon$, there is a small disk D_0 on the side of E_1 opposite to $f_{t_0}(I)$ so that ∂D_0 is the union of an arc in E_1, an arc in ∂H and the small end-segment of $f_{t_0+\epsilon}(I)$. For values of t between $t_0 + \epsilon$ and t', $f_t(I)$ remains transverse to E_1 and is divided by E_1 into two segments, each of which is properly isotoped in $H - \Delta$ during the interval $[t_0 + \epsilon, t']$. Apply the isotopy extension theorem to this isotopy in $H - \Delta$. At the end of the isotopy, the small end-segment of $f_{t_0+\epsilon}(I)$ has become $J_0 \subset f_{t'}(I)$ and D_0 remains disjoint from the other interval $f_{t'}(I) - J_0$ because they were disjoint at the start of the isotopy. Similarly construct D_1 and J_1 using the isotopy extension theorem in the interval $[t', t_1 - \epsilon]$, but starting at the right end-point $t_1 - \epsilon$. The disks D_0 and D_1 each intersect E_1 in a single arc; it is straightforward to isotope the disks further so that the arcs coincide, the disks are transverse (including along the arc $\alpha = \partial D_i \cap E_1$ and also along $J_0 = J_1$ in the case of a bounce). An innermost disk argument ensures that all components of intersection are proper arcs.
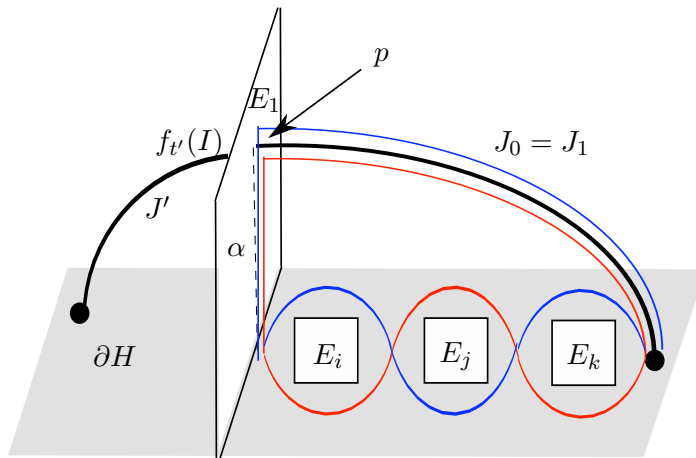
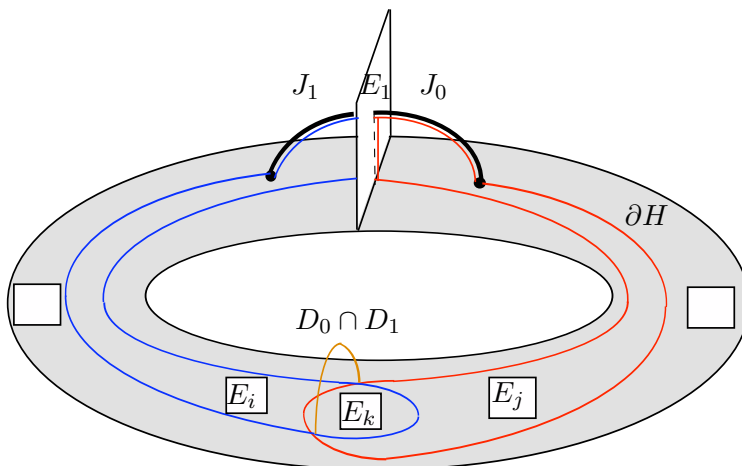FIGURE 6. Tracking disks: ∂D_0 (red) and ∂D_1 (blue)

FIGURE 7. ∂D_0 (red), ∂D_1 (blue), $D_0 \cap D_1$ (orange)

Definition 4.5. *Minimize the number of arc components in $D_0 \cap D_1$ via isotopies that fix each subarc $\partial D_i - \partial H$. The number $|D_0 \cap D_1|$ of intersection arcs is called the complexity of the isotopy f_t.*

Lemma 4.6. *If there is a parallelism disk D for I so that the interiors of both $f_{t_0}(D)$ and $f_{t_1}(D)$ are disjoint from the meridian disks Δ, then $\omega \in \mathfrak{A}\mathfrak{F}_{\{E_0, E_1, \ldots, E_g\}}$.*

Proof. Replace the isotopy $f_t, 0 \leq t \leq t_0$ by an isotopy that carries E_0 to $f_{t_0}(D)$ in $H - \Delta$ and replace the isotopy $f_t, t_1 \leq t \leq 1$ by an isotopy that carries $f_{t_1}(D)$ to E_0 in $H - \Delta$. The resulting isotopy carries E_0 back to itself, and therefore represents an element of $\mathfrak{F}_{E_0}$. It has been

obtained by pre- and post-multiplying ω by elements whose represent-
ing isotopies keep I disjoint from Δ and therefore lie in $\mathfrak{AF}_{E_0,\dots,E_g}$. It
follows that $\omega \in \mathfrak{AF}_{E_0,\dots,E_g}$. $\hspace{1cm}$ $\square$

Corollary 4.7. *If f_t is a pass of complexity zero, then $\omega \in \mathfrak{AF}_{\{E_0,E_1,\dots,E_g\}}$.*

Proof. Since f_t is a pass, $f_{t'}(I) = J_0 \cup J_1$. Since the complexity is zero,
the union D of the disks D_0 and D_1 along α is embedded. Hence D is
a parallelism disk for $f_{t'}(I)$. By construction, the isotopy f_t extends to
an ambient isotopy of D so that the interiors of both $f_{t_0}(D)$ and $f_{t_1}(D)$
are disjoint from the meridian disks Δ. Now apply Lemma 4. $\hspace{1cm}$ $\square$

Lemma 4.8. *If f_t is a bounce of complexity zero, then ω can also
be represented by the product of two passes of complexity zero. Hence
$\omega \in \mathfrak{AF}_{\{E_0,E_1,\dots,E_g\}}$.*

Proof. Since f_t is a bounce, $J_0 = J_1$. Since the complexity is zero, the
union D of the tracking disks D_0 and D_1 along their common boundary
subarc is an embedded disk D in H which contains $J_0 = J_1$ and is
disjoint from the complementary subarc $J' = f_{t'}(I) - J_i$. Furthermore
D is disjoint from Δ except along the arc α where it is tangent to E_1.

We know that $f_{t'}(I)$ is ∂-parallel in H; a parallelism disk can be
isotoped so that the arc of intersection with Δ that has its endpoint
at $p \in E_1$ is α. A standard innermost-disk, outermost-arc argument in
$(\Delta \cup D) - \alpha$ will then find a parallelism disk that intersects the entire
collection of disks $D \cap \Delta$ in just α. The half E' of this parallelism
disk that is incident to J' can be used to define an isotopy of the arc
J' across E_1. When this isotopy, followed by its inverse, are inserted
into f_t at time t', the resulting isotopy is the result of a deformation
of f_t and so still represents ω, but the new isotopy is the product of
two passes. Since $\partial E'$ intersects each ∂D_i only along α, $D_0 \cup E'$ and
$D_1 \cup E'$ are embedded disks, so each pass has complexity zero. $\hspace{0.5cm}$ $\square$

Lemma 4.9. *A bounce of complexity n can be written as the product
of two bounces, one of complexity zero and the other of complexity less
than n.*

Proof. The assumption is that the tracking disks D_0 and D_1 for f_t
intersect in n arc components, each with its endpoints on $\partial D_i \cap \partial H$. A
standard outermost-arc argument on the n curves of intersection gives
a third disk E' such that $\partial E'$ coincides with the boundaries of the
D_i away from ∂H, E' is entirely disjoint from D_0, and E' intersects
D_1 in $n - 1$ or fewer arcs. The disk E' can be used to define an
isotopy of the arc $J_0 = J_1$ across E_1. When this isotopy, followed
by its inverse, are inserted into f_t at time t', the resulting isotopy is

the result of a deformation of f_t and so still represents ω, but the new isotopy is the product of two bounces each determined by one of $D_0 \cup E'$ or $E' \cup D_1$. The former is an embedded disk, so that bounce has complexity zero. Since $|D_1 \cap E'| \le n-1$, the bounce determined by $E' \cup D_1$ has complexity at most $n-1$. $\qquad\square$

Lemma 4.10. *A pass of complexity n can be written as the product of a bounce of complexity zero and a pass of complexity less than n.*

Proof. The proof is much like that of Lemma 4.9: Using an outermost (in D_1) arc of intersection, a disk E' is found so that $\partial E' = J_0 \cup \alpha$, E' is disjoint from D_0, and E' intersects D_1 in at most $n-1$ arcs. The disk E' can be used to define an isotopy of the arc J_0 across E_1. When the isotopy, followed by its inverse, are inserted into f_t at time t', the resulting isotopy is the result of a deformation of f_t and so still represents ω, but the new isotopy is the product of a bounce determined by $D_0 \cup E'$ and a pass determined by $E' \cup D_1$. The former is an embedded disk, so the bounce has complexity zero. The pass determined by $E' \cup D_1$ has complexity at most $n-1$. $\qquad\square$

Proposition 4.3 now follows by induction on the complexity of f_t. $\quad\square$

Proposition 4.11. *Any element $\omega \in \pi_1(\mathrm{Unk}(I, H))$ has a representative isotopy which is of width at most 1 with respect to the collection of meridian disks Δ.*

Proof. Let D_0 be any disk of parallelism for I_0 that is disjoint from Δ, let f_t be an isotopy of I that represents ω, so in particular $f_1(I_0) = I_0$. Extend f_t to an ambient isotopy of H, and let $D_1 = f_1(D_0)$ be the final position of D_0 after the isotopy. Call D_1 the *terminal disk* for the isotopy. It follows from Lemma 3.1 that, up to a product of elements of $\mathfrak{F}_{D_0}$, each of which has width 1 with respect to Δ, ω is represented by any isotopy of I_0 back to itself that has the same terminal disk D_1. The proof will be by induction on $|D_1 \cap \Delta|$; we assume this has been minimized by isotopy, so in particular all components of intersection are arcs with end points on the arc $\alpha = \partial D_1 \cap \partial H$.

Given the disk D_1, here is a useful isotopy, called a D_1-*sweep*, that takes D_0 to D_1. Pick any p point on α, and isotope D_0 in the complement of Δ so that it becomes a small regular neighborhood N_p of p in D_1. (Do not drag D_1 along.) This isotopy carries I_0 to an arc $I_p = \partial N_p - \partial H$ properly embedded in D_1. Then stretch N_p in D_1 until it fills all of D_1, so we view N_p as sweeping across all of D_1. The combination of the two isotopies carries I_0 back to itself and takes D_0 to D_1 so, up to a further product with width one isotopies, we can assume that this combined isotopy represents ω (via Lemma 3.1).

It will be useful to have a notation for this two stage process: let g be the isotopy in the complement of Δ that takes D_0 to N_p and let s be the isotopy that sweeps N_p across D_1. We wish to determine the width of the sequence $g * s$ of the two isotopies. Since the isotopy g is disjoint from Δ, much depends on the D_1-sweep s of N_p across D_1. In particular, if the number of intersection arcs $|D_1 \cap \Delta| \leq 1$ it is obvious how to arrange the sweep so the width is at most 1. So henceforth we assume that $|D_1 \cap \Delta| \geq 2$.
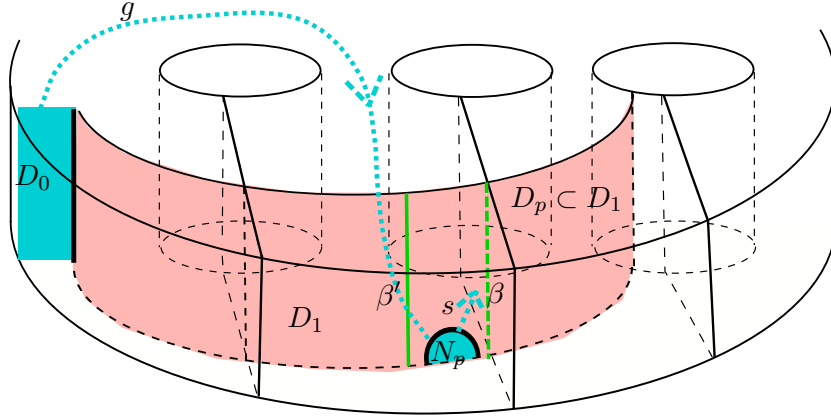

FIGURE 8

Of all arcs in $D_1 \cap \Delta$, let β be one that is outermost in Δ. That is, the interior of one of the disks that β cuts off from Δ is disjoint from D_1. Pick the point p for the D_1-sweep to be near an end-point of β in ∂D_1, on a side of β that contains at least one other arc. (Since $|D_1 \cap \Delta| \geq 2$ there is at least one other arc.) Let $\beta' \subset D_1$ be an arc in $D_1 - \Delta$ with an end at p and which is parallel to β and let D_p be the subdisk of D_1 cut off by β' that does not contain $I_0 \subset \partial D_1$. See Figure 8. Now do the sweep s in two stages: first sweep N_p across D_p until it coincides with D_p, then complete the sweep across D_1. See Figure 9 Denote the two stages of the sweep by $s = s_1 * s_2$. The isotopy s_1 carries I_p to the arc β'; exploiting the fact that β is outermost in Δ there is an obvious isotopy h (best imagined in Figure 8) that carries β' back to I_p, an isotopy that is disjoint from Δ and from D_1. Finally, deform the given isotopy $g * s = g * s_1 * s_2$ whose width we seek, to the isotopy $g * s_1 * h * \overline{g} * g * \overline{h} * s_2$, which can be written as the product $(g*s_1*h*\overline{g})*(g*\overline{h}*s_2)$, of two isotopies, each representing an element of $\pi_1(\mathrm{Unk}(I, H))$. The first isotopy has terminal disk containing exactly the arcs of $D_p \cap \Delta$ and the second has terminal disk containing the other arcs $D_1 \cap \Delta$. By construction of β' each set is non-empty so each

terminal disk intersects Δ in fewer arcs than D_1 did. By inductive hypothesis, each isotopy can be deformed to have width at most 1 so, by Corollary 4.2, $w(\omega) \leq 1$. $\square$
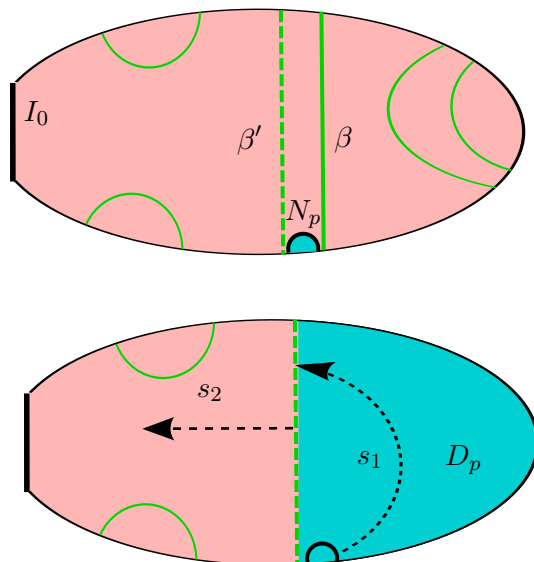

FIGURE 9

5. CONNECTION TO THE GOERITZ GROUP

Suppose H is a genus $g \geq 1$ handlebody, and Σ is a genus $g+1$ Heegaard surface in H. That is, Σ splits H into a handlebody H_1 of genus $g+1$ and a compression-body H_2. The compression-body H_2 is known to be isotopic to the regular neighborhood of the union of ∂H and an unknotted properly embedded arc $I_0 \subset H$ (see [ST1, Lemma 2.7]).

Let $\mathrm{Diff}(H)$ be the group of diffeomorphisms of H and $\mathrm{Diff}(H, \Sigma) \subset \mathrm{Diff}(H)$ be the subgroup of diffeomorphisms that take the splitting surface Σ to itself. Following [JM], define the Goeritz group $G(H, \Sigma)$ of the Heegaard splitting to be the group consisting of those path components of $\mathrm{Diff}(H, \Sigma)$ which lie in the trivial path component of $\mathrm{Diff}(H)$. So, a non-trivial element of $G(H, \Sigma)$ is represented by a diffeomorphism of the pair (H, Σ) that is isotopic to the identity as a diffeomorphism of H, but no isotopy to the identity preserves Σ.

Theorem 5.1. *For H a handlebody of genus ≥ 2,*

$$G(H, \Sigma) \cong \pi_1(\mathrm{Unk}(I, H)).$$

For H a solid torus, there is an exact sequence

$$1 \to \mathbb{Z} \to \pi_1(\mathrm{Unk}(I, H)) \to G(H, \Sigma) \to 1.$$

In either case, the finite collection of generators of $\pi_1(\mathrm{Unk}(I, H))$ described above is a complete set of generators for $G(H, \Sigma)$.

Proof. This is a special case of [JM, Theorem 1], but there the fundamental group of the space $\mathcal{H}(H, \Sigma) = \mathrm{Diff}(H)/\mathrm{Diff}(H, \Sigma)$ takes the place of $\pi_1(\mathrm{Unk}(I, H))$. So it suffices to show that there is a homotopy equivalence between $\mathcal{H}(H, \Sigma)$ and $\mathrm{Unk}(I, H)$. We sketch a proof:

Fix a diffeomorphism $e : I \to I_0 \subset H$, for I_0 as above a specific unknotted arc in H. It is easy to see that for any other embedding $e' : I \to H$ with $e'(I)$ unknotted, there is a diffeomorphism $h : H \to H$, so that $he = e'$. It follows that the restriction to I_0 defines a surjection $\mathrm{Diff}(H) \to \mathrm{Emb}_0(I, H)$. Since any automorphism of I_0 extends to an automorphism of H, this surjection maps the subgroup $\mathrm{Diff}(H, I_0)$ (diffeomorphisms of H that take I_0 to itself) onto the space of automorphisms of I_0. It follows that $\mathrm{Unk}(I, H) = \mathrm{Emb}_0(I, H)/\mathrm{Diff}(I)$ has the same homotopy type as $\mathrm{Diff}(H)/\mathrm{Diff}(H, I_0)$.

Suppose for the compression-body H_2 in the Heegaard splitting above we take a fixed regular neighborhood of $\partial H \cup I_0 \subset H$. Let $\mathrm{Diff}(H, H_2, I_0)$ denote the subgroup of $\mathrm{Diff}(H, H_2)$ that takes I_0 to itself.

Lemma 5.2. *The inclusion $\mathrm{Diff}(H, H_2, I_0) \subset \mathrm{Diff}(H, H_2)$ is a homotopy equivalence.*

Proof. First note that the path component $\mathrm{Diff}_0(H, H_2)$ of $\mathrm{Diff}(H, H_2)$ containing the identity is contractible, using first [EE] on ∂H_2 and then [Ha1] on the interiors of H_1 and H_2. Any other path component of $\mathrm{Diff}(H, H_2)$ is therefore contractible, since each is homeomorphic to $\mathrm{Diff}_0(H, H_2)$. It therefore suffices to show that any path component of $\mathrm{Diff}(H, H_2)$ contains an element of $\mathrm{Diff}(H, H_2, I_0)$. Equivalently, it suffices to show that any diffeomorphism of H_2 is isotopic to one that sends I_0 to I_0. A standard innermost disk, outermost arc argument shows that H_2 contains, up to isotopy, a single ∂-reducing disk D_0; once D_0 has been isotoped to itself and then the point $I_0 \cap D_0$ isotoped to itself in D_0, the rest of I_0 can be isotoped to itself, using a product structure on $H_2 - \eta(D_0) \cong \partial H \times I$. □

Lemma 5.3. *The inclusion $\mathrm{Diff}(H, H_2, I_0) \subset \mathrm{Diff}(H, I_0)$ is a homotopy equivalence.*

Proof. Such spaces have the homotopy type of CW-complexes. (See [HKMR, Section 2] for a discussion of this and associated properties.)

So it suffices to show that, for any

$$\Theta : (B^k, \partial B^k) \to (\mathrm{Diff}(H, I_0), \mathrm{Diff}(H, H_2, I_0)), k \geq 1,$$

there is a pairwise homotopy to a map whose image lies entirely inside $\mathrm{Diff}(H, H_2, I_0)$. The core of the proof is a classic argument in smooth topology; a sketch of the rather complex argument is given in the Appendix. Note that the level of analysis that is required is not much deeper than the Chain Rule in multivariable calculus.[1]   □

A diffeomorphism of H that takes Σ to itself will also take H_2 to itself, since H_2 and H_1 are not diffeomorphic. Hence $\mathrm{Diff}(H, \Sigma) = \mathrm{Diff}(H, H_2)$ and

$$\mathcal{H}(H, \Sigma) = \mathrm{Diff}(H)/\mathrm{Diff}(H, \Sigma) = \mathrm{Diff}(H)/\mathrm{Diff}(H, H_2).$$

Lemma 5.2 shows that the natural map

$$\mathrm{Diff}(H)/\mathrm{Diff}(H, H_2, I_0) \to \mathrm{Diff}(H)/\mathrm{Diff}(H, H_2)$$

is a homotopy equivalence and Lemma 5.3 shows that the natural map

$$\mathrm{Diff}(H)/\mathrm{Diff}(H, H_2, I_0) \to \mathrm{Diff}(H)/\mathrm{Diff}(H, I_0)$$

is a homotopy equivalence. Together these imply that $\mathcal{H}(H, \Sigma) = \mathrm{Diff}(H)/\mathrm{Diff}(H, H_2)$ is homotopy equivalent to $\mathrm{Diff}(H)/\mathrm{Diff}(H, I_0)$, which we have already seen is homotopy equivalent to $\mathrm{Unk}(I, H)$.   □

---

[1]But in some ways the argument in the Appendix is just a distraction: First of all, we do not need the full homotopy equivalence to show that the fundamental groups of $\mathcal{H}(H, \Sigma)$ and $\mathrm{Unk}(I, H)$ are isomorphic (but restricting attention just to the fundamental group wouldn't really simplify the argument). Secondly, we could have, from the outset, informally viewed each isotopy of I in H described in the proof above to be a proxy for an isotopy of a thin regular neighborhood H' of $\partial H \cup I$; the difference then between isotopies of H' and isotopies of the compression body H_2 is easily bridged, requiring from the Appendix only Lemma A.7 and following.

APPENDIX A. MOVING A DIFFEOMORPHISM TO PRESERVE H_2.

A conceptual sketch of the proof of Lemma 5.3 can be broken into six steps. It is useful to view $\Theta : (B^k, \partial B^k) \to (\mathrm{Diff}(H, I_0), \mathrm{Diff}(H, H_2, I_0))$ as a family of diffeomorphisms $h_u : (H, I_0) \to (H, I_0)$ parameterized by $u \in B^k$ so that for u near ∂B^k, $h_u(H_2) = H_2$. By picking a particular value $u_0 \in \partial B^k$ and post-composing all diffeomorphisms with $h_{u_0}^{-1}$ we may, with no loss of generality, assume that $\Theta(B^k)$ lies in the component of $\mathrm{Diff}(H, I_0)$ that contains the identity. Here are the six stages:

(1) (Pairwise) homotope Θ so that for each u, h_u is the identity on ∂H.
(2) Further homotope Θ so that for some collar structure $\partial H \times I$ near ∂H and for each $u \in B^k$, $h_u|(\partial H \times I)$ is the identity.
(3) Further homotope Θ so that after the homotopy there is a neighborhood of I_0 and a product structure $I_0 \times \mathbb{R}^2$ on that neighborhood so that for all $u \in B^k$, $h_u|(I_0 \times \mathbb{R}^2)$ commutes with projection to I_0 near I_0.
(4) Further homotope Θ so that for all $u \in B^k$, $h_u|(I_0 \times \mathbb{R}^2)$ is a linear (i. e. GL_n) bundle map over I_0 near I_0.
(5) Further homotope Θ so that for all $u \in B^k$, $h_u|(I_0 \times \mathbb{R}^2)$ is an orthogonal (i. e. O_n) bundle map near I_0. There is then a regular neighborhood $H' \subset H_2$ of $\partial H \cup I_0$ so that for all $u \in B^k$, $h_u(H') = H'$.
(6) Further homotope Θ so that for the specific regular neighborhood H_2 of $\partial H \cup I_0$ and for each $u \in B^k$, $h_u(H_2) = H_2$.

For the first stage, let $\Theta_{\partial H} : B^k \to \mathrm{Diff}(\partial H, \partial I_0)$ be the restriction of each h_u to ∂H. The further restriction $\Theta_{\partial H}|\partial B^k$ defines an element of $\pi_{n-1}(\mathrm{Diff}(\partial H, \partial I_0))$ in the component containing the identity $h_{u_0}|\partial H$. The space $\mathrm{Diff}(\partial H, \partial I_0)$ is known to be contractible (see [EE], [ES], or [Gr]). Hence $\Theta_{\partial H}|\partial B^k$ is null-homotopic. The homotopy can be extended to give a homotopy of $\Theta_{\partial H} : B^k \to \mathrm{Diff}(\partial H, \partial I_0)$ and then, by the isotopy extension theorem, to a homotopy of $\Theta : B^k \to \mathrm{Diff}(H, I_0)$, after which each $h_u|\partial H$ is the identity.

The homotopy needed for the second stage is analogous to (and much simpler than) the sequence of homotopies constructed in stages 3 to 5, so we leave its construction to the reader.

Stages 3 through 5 might best be viewed in this context: One is given a smooth embedding $h : I \times \mathbb{R}^2 \to I \times \mathbb{R}^2$ that restricts to a diffeomorphism $I \times \{0\} \to I \times \{0\}$ and which is the identity near $\partial I \times \mathbb{R}^2$. One hopes to isotope the embedding, moving only points near $I \times \{0\}$ and away from $\partial I \times \mathbb{R}^2$ so that afterwards the embedding

is an orthogonal bundle map near $I \times \{0\}$. Moreover, one wants to do this in a sufficiently natural way that a B^k-parameterized family of embeddings gives rise to a B^k-parameterized family of isotopies. The relevant lemmas below are proven for general $\mathbb{R}^n$, not just $\mathbb{R}^2$, since there is little lost in doing so. (In fact the arguments could easily be further extended to families of smooth embeddings $D^\ell \times \mathbb{R}^n \to D^\ell \times \mathbb{R}^n, \ell > 1$).

Stage 3: Straightening the diffeomorphism near I_0

Definition A.1. *A smooth isotopy $f_t : I \times \mathbb{R}^n \to I \times \mathbb{R}^n$ of an embedding $f_0 : I \times \mathbb{R}^n \to I \times \mathbb{R}^n$ is* allowable *if it has compact support in $(\text{int } I) \times \mathbb{R}^n$. That is, the isotopy is fixed near $\partial I \times \mathbb{R}^n$ and outside of a compact set in $I \times \mathbb{R}^n$.*

Definition A.2. *A smooth embedding $f : I \times (\mathbb{R}^n, 0) \to I \times (\mathbb{R}^n, 0)$* commutes with projection to I *if for all $(x,y) \in I \times \mathbb{R}^n$, $p_1 f(x,y) = f(x,0) \in I$.*

Definition A.3. *For A a square matrix or its underlying linear transformation, let $|A|$ denote the operator norm of A, that is*

$$|A| = max\{|Ax| : x \in \mathbb{R}^n, |x| = 1\}$$
$$= max\{\frac{|Ax|}{|x|} : 0 \neq x \in \mathbb{R}^n\}$$

.

Let $\phi : [0, \infty) \to [0, 1]$ be a smooth map such that $\phi([0, 1/2]) = 0$, $\phi : (1/2, 1) \to (0, 1)$ is a diffeomorphism, and $\phi([1, \infty)) = 1$. Let b_0 be an upper bound for $s\phi'(s)$; for example ϕ could easily be chosen so that $b_0 = 4$ suffices. For any $\epsilon > 0$ let $\phi_\epsilon : [0, \infty) \to [0, 1]$ be defined by $\phi_\epsilon(s) = \phi(\frac{s}{\epsilon})$.

Then

$$\phi_\epsilon([\epsilon, \infty)) = 1$$

and, for any $\epsilon > 0$ and any $s \in [0, \infty)$,

$$s\phi'_\epsilon(s) = s\phi'(\frac{s}{\epsilon})\frac{1}{\epsilon} = \frac{s}{\epsilon}\phi'(\frac{s}{\epsilon}) \leq b_0.$$

Thus b_0 is an upper bound for all $s\phi'_\epsilon(s)$.

Now fix an $\epsilon > 0$ and define $g_1 : \mathbb{R}^n \to [0, 1]$ by $g_1(y) = \phi_\epsilon(|y|)$. This is a smooth function on $\mathbb{R}^n$ which is 0 on the ball $B_{\epsilon/2}$ and 1 outside of B_ϵ. Linearly interpolating,

$$g_t(y) = (1 - t) + tg_1(y)$$

is a smooth homotopy with support in B_ϵ from the constant function 1 to g_1. Define a smooth homotopy $\lambda_t : \mathbb{R}^n \to \mathbb{R}^n, 0 \le t \le 1$ by $\lambda_t(y) = g_t(y)y$ and note that, by the chain rule, the derivative $D\lambda_t(y)(z) = g_t(y)z + t\phi'_\epsilon(|y|)(\frac{y}{|y|} \cdot z)y$ so $D\lambda_t(y)$ has the matrix

$$g_t(y)I_n + t\frac{\phi'_\epsilon(|y|)}{|y|}yy^*$$

and so satisfies

$$|D\lambda_t(y)| \le g_t(y) + t\phi'_\epsilon(|y|)|y|\frac{|yy^*|}{|y^2|} \le 1 + b_0$$

since the norm of the matrix yy^* is $|y|^2$. (This last point is best seen by taking y to be a unit vector, z to be any other unit vector and observing that $|yy^*z| = |y(y \cdot z)| \le |y(y \cdot y)| = 1$. Note also that the function represented by an expression like $\phi'_\epsilon(|y|)/|y|$ is understood to be 0 when $y = 0$. Since $\phi'_\epsilon(y) \equiv 0$ for y near 0, the function is smooth.)

The central point of the above calculation is only this: $|D\lambda_t(y)|$ has a uniform bound that is independent of t or the value of ϵ that is used in the construction of λ.

Lemma A.4 (Handle-straightening). *Suppose $f : I \times (\mathbb{R}^n, 0) \to I \times (\mathbb{R}^n, 0)$ is an embedding which commutes with projection to I near $\partial I \times \{0\}$. Then there is an allowable isotopy of f to an embedding that commutes with projection to I near all of $I \times \{0\}$. (See Figure 10.)*

Moreover, given a continuous family $f^u, u \in B^k$ of such embeddings so that for u near ∂B^k, f^u commutes with projection to I near $I \times \{0\}$, a continuous family of such isotopies can be found, and the isotopy is constant for any u sufficiently near ∂B^k.

Proof. We will construct an isotopy f_t for a given $f = f_0$ and then observe that it has the properties described. By post-composing with a GL_n bundle map over the diffeomorphism $f^{-1}|I \times \{0\}$, we may as well assume that $f|I \times \{0\}$ is the identity and, along $I \times \{0\}$, $D(p_2f)$ is the identity on each $\mathbb{R}^n$ fiber $\mathbb{R}^n_x$. Hence, at any point $(x, 0) \in I \times \{0\}$, the matrix of the derivative $Df : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R} \times \mathbb{R}^n$ is the identity except for perhaps the last n entries of the first row, which contain the gradient $\nabla(p_1f|R^n_x)$.

For any $\epsilon > 0$, consider the map

$$f_t : I \times \mathbb{R}^n \to I \times \mathbb{R}^n$$

defined near $I \times B_\epsilon$ by

$$f_t\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} p_1f(x, \lambda_t(y)) \\ p_2f(x, y) \end{pmatrix}$$

and fixed at f outside $I \times B_\epsilon$. For any value of t, the derivative of f_t at any $(x, y) \in I \times \mathbb{R}^n$ differs from that of f by multiplying its first row on the right by the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & D\lambda_t(y) \end{pmatrix}.$$

For ϵ small, the first row of Df_t will then be quite close to a row vector of the form

$$\begin{pmatrix} 1 & v^* \end{pmatrix}$$

where $|v^*| \leq (1 + b_0)|\nabla(p_1 f | \mathbb{R}_x^n)|$ and the matrix for Df_t will be quite close to the matrix

$$\begin{pmatrix} 1 & v^* \\ 0 & I_n \end{pmatrix}.$$

So, although we cannot necessarily make v^* small by taking ϵ small, the entries in v^* are at least bounded by a bound that is independent of ϵ, and, by taking ϵ small, the rest of the matrix can be made to have entries arbitrarily close to those of the identity matrix. In particular, for ϵ sufficiently small, Df_t will be non-singular everywhere, and so each f_t will be a smooth embedding. Thus f_t will be an allowable isotopy to a smooth embedding f_1 that commutes with projection to I on $I \times B_{\frac{\epsilon}{2}}$, since on $I \times B_{\frac{\epsilon}{2}}$ we have $p_1 f(x, \lambda_1(y)) = p_1 f(x, 0) = f(x, 0) \in I$.
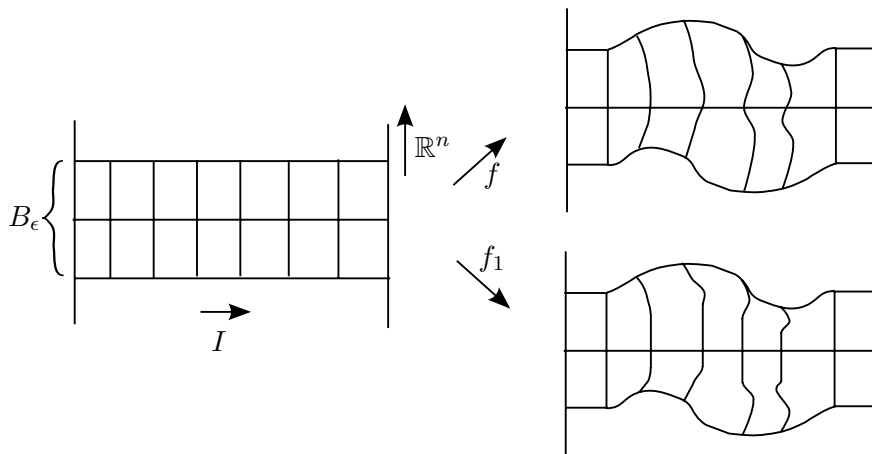

FIGURE 10

The extension to a parameterized family of embeddings $f^u, u \in B^k$ is relatively easy: pick ϵ so small (as is possible, since B^k is compact) so that the above argument works simultaneously on each $f^u, u \in B^k$ and also so small that, for each u near ∂B^k, $I \times B_\epsilon$ lies within the area on which f^u already commutes with projection to I. $\square$

Stage 4: Linearizing the diffeomorphism near I_0

Lemma A.5 ($\mathbf{Diff_n/GL_n}$). *Suppose $f : I \times (\mathbb{R}^n, 0) \to I \times (\mathbb{R}^n, 0)$ is an embedding which commutes with projection to I and which is a GL_n bundle map near $\partial I \times \{0\}$. Then there is an allowable isotopy of f, through embeddings which commute with projection to I, to an embedding that is a GL_n bundle map near $I \times \{0\}$.*

Moreover, given a continuous family $f^u, u \in B^k$ of such embeddings so that for u near ∂B^k, f^u is a GL_n bundle map near $I \times \{0\}$, a continuous family of such isotopies can be found, and the isotopy of f^u is constant for any u sufficiently near ∂B^k.

Proof. For each $x \in I$ consider the restriction $f|\mathbb{R}^n_x$ of f to the fiber $\mathbb{R}^n_x$ over x. By post-composing f with the GL_n bundle map over $f^{-1}|I \times \{0\}$ determined by $D(f|\mathbb{R}^n_x)(0)^{-1}$ we may as well assume that $f|I \times \{0\}$ is the identity and that $D(f|\mathbb{R}^n_x)(0)$ is the identity for each x. There are differentiable maps $\psi_{i,x} : \mathbb{R}^n_x \to \mathbb{R}^n_x, 1 \leq i \leq n$, smoothly dependent on x, so that for $y \in R^n_x$, $f(y) = \Sigma^n_{i=1} y_i \psi_{i,x}(y)$ and $\psi_{i,x}(0) = e_i$, the ith unit vector in $\mathbb{R}^n$ [BJ, Lemma 2.3].

Choose ϵ small and let $\lambda_t : \mathbb{R}^n \to \mathbb{R}^n$ be the homotopy defined above using ϵ. For each $t \in [0,1]$ define $f_t : I \times (\mathbb{R}^n, 0) \to I \times (\mathbb{R}^n, 0)$ as the bundle map for which each $f_t|\mathbb{R}^n_x$ is given by $\Sigma^n_{i=1} y_i \psi_{i,x}(\lambda_t(y))$. When $t = 1$ the function on $B_{\epsilon/2} \subset \mathbb{R}^n_x$ is given by $\Sigma^n_{i=1} y_i e_i$, i. e. the identity. As in Step 1, the bound $|D\lambda_t(y_0)| \leq 1 + b_0$ guarantees that if ϵ is chosen sufficiently small the derivative of $\Sigma^n_{i=1} y_i \psi_{i,x}(\lambda_t(y))$ is close to the identity throughout $I \times B_\epsilon$; hence f_t remains a diffeomorphism for each t.

The extension to a parameterized family of embeddings f^u is done as in Stage 3 (Sub-section A) above. $\square$

Stage 5: Orthogonalizing the diffeomorphism near I_0

Lemma A.6 ($\mathbf{GL_n/O_n}$). *Suppose $f : I \times (\mathbb{R}^n, 0) \to I \times (\mathbb{R}^n, 0)$ is a GL_n bundle map which is an O_n bundle map near $\partial I \times \{0\}$. Then there is an allowable isotopy of f, through embeddings which commute with projection to I, to an embedding that is an O_n bundle map near $I \times \{0\}$.*

Moreover, given a continuous family $f^u, u \in B^k$ of such embeddings so that for u near ∂B^k, f^u is an O_n bundle map near $I \times \{0\}$, a continuous family of such isotopies can be found, and the isotopy of f^u is constant for any u sufficiently near ∂B^k.

Proof. Once again we may as well assume $f|I \times \{0\}$ is the identity and focus on the linear maps $f|R^n_x$.

As a consequence of the Gram-Schmidt orthogonalization process, any matrix $A \in GL_n$ can be written uniquely as the product QT of an orthogonal matrix Q and an upper triangular matrix T with only positive entries in the diagonal. The entries of Q and $T = Q^t A$ depend smoothly on those of A. In particular, by post-composing f with the orthogonal bundle map determined by the inverse of the orthogonal part of $(Df_x)(0)$ we may as well assume that for each $x \in I$, $f|\mathbb{R}_x^n$ is defined by an upper triangular matrix T_x with all positive diagonal entries.

A worrisome example: It seems natural to use the function g_t defined above to interpolate linearly between T_x and the identity, in analogy to the way g_t (via λ_t) was used in Stages 3 and 4. Here this would mean setting

$$f_t(y) = [I_n + g_t(y)(\mathrm{T}_x - I_n)]y \qquad \forall y \in R_x^n.$$

This strategy fails without control on the bound b_0 of $s\phi'(s)$, even in the relevant case $n = 2$, because f_t may fail to be a diffeomorphism. For example, if

$$\mathrm{T}_x = \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}$$

then with the above definition

$$(D(f_1|\mathbb{R}_x^2)_y)(z) = \begin{pmatrix} 1 & g_1(y)r \\ 0 & 1 \end{pmatrix} z + \begin{pmatrix} \phi'(|y|)(ry_2)(y \cdot z)/|y| \\ 0 \end{pmatrix}$$

Choose $z = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and get the vector

$$\begin{pmatrix} 1 + \phi'(|y|)(ry_1y_2)/|y| \\ 0 \end{pmatrix} = \begin{pmatrix} 1 + r\phi'(|y|)|y|\frac{y_1y_2}{y_1^2+y_2^2} \\ 0 \end{pmatrix}$$

For a fixed value of $|y|$, the ratio $\frac{y_1y_2}{y_1^2+y_2^2}$ takes on every value in $[-1/2, 1/2]$. So, unless $\phi'(|y|)|y| < \frac{2}{|r|}$, which could be much smaller than b_0, the vector $(D(f_1|\mathbb{R}_x^2)_y)(z)$ will be trivial for some y. At this value of y, $D(f_1|\mathbb{R}_x^2)_y$ would be singular, so $f_1|\mathbb{R}_x^2$ would not be a diffeomorphism.

On the other hand, in contrast to the previous two stages, there is no advantage to restricting the support of the isotopy f_t to an ϵ neighborhood of $I \times \{0\}$ since f, as a GL_n bundle map, is independent of scale. So we are free to choose ϕ a bit differently:

Given $\kappa > 0$, let $\phi : [0, \infty) \to [0, 1]$ be a smooth, monotonically nondecreasing map such that $\phi([0, 1]) = 0$, $\phi(s) = 1$ outside some closed interval, and, for all s, $s\phi'(s) < \kappa$. For example, ϕ could be obtained

by integrating a smooth approximation to the discontinuous function on $[0, \infty)$ which takes the value $\kappa/2s$ for $s \in [1, e^{\frac{2}{\kappa}}]$ but is otherwise 0.

Suppose A is any upper triangular matrix with positive entries in the diagonal and $\tau \in [0, 1]$. Then the matrix $A_\tau = (I + \tau(A - I)) = (1 - \tau)I + \tau A$ is invertible, since it also is upper triangular and has positive diagonal entries. Suppose $\kappa_1, \kappa_2 \in [0, \infty)$ satisfy

$$\kappa_1 \leq \frac{1}{sup_\tau |A_\tau^{-1}|}, \qquad\qquad \kappa_2 \leq \frac{1}{|A - I|}.$$

Then, for any $\tau \in [0, 1]$ and any $y, z \neq 0 \in \mathbb{R}^n$,

$$|A_\tau z| \geq \kappa_1 |z|, \qquad\qquad \kappa_2 |(A - I)y| \leq |y|.$$

In particular, if $0 \leq \kappa < \kappa_1 \kappa_2$ then

$$|A_\tau(z) + \kappa \frac{|z|}{|y|}(A - I)y| \geq |A_\tau(z)| - \kappa \frac{|z|}{|y|}|(A - I)y| > \kappa_1 |z| - \kappa_1 \frac{|z|}{|y|}|y| = 0.$$

Apply this to the problem at hand by choosing values κ_1, κ_2 so that for all $x \in [0, 1]$,

$$\kappa_1 \leq \frac{1}{sup_\tau |(T_x)_\tau^{-1}|}, \qquad\qquad \kappa_2 \leq \frac{1}{|T_x - I_n|}.$$

Then choose $\phi : [0, \infty) \rightarrow [0, 1]$ as above so that for all $s \in [0, \infty)$, $s\phi'(s) < \kappa_1 \kappa_2$. Much as in Step 2, define the smooth homotopies $g_t : \mathbb{R}^n \rightarrow [0, 1]$ and $f_t|\mathbb{R}^n_x : \mathbb{R}^n_x \rightarrow \mathbb{R}^n_x$ via

$$g_t(y) = (1 - t) + t\phi(|y|)(y)$$
$$f_t(y) = (I_n + g_t(y)(T_x - I_n)).$$

Then for all $x \in I$ and each $y \neq 0 \neq z \in \mathbb{R}^n_x$,

$$D(f_t|\mathbb{R}^n_x)_y(z) = (I + g_t(y)(T_x - I_n))z + Dg_t(y)(z)(T_x - I_n)y$$
$$= (I + g(y)(T_x - I_n))z + t\phi'(|y|)\frac{y \cdot z}{|y|}(T_x - I_n)y.$$

Now notice that by construction

$$t\phi'(|y|)\frac{y \cdot z}{|y|} \leq \phi'(|y|)|y|\frac{y \cdot z}{|y|^2} < \kappa_1 \kappa_2 \frac{|z|}{|y|}$$

so, applying the argument above to $\tau = g(y)$ and $A = T_x$, we have $|D(f_t|\mathbb{R}^n_x)_y(z)| > 0$. Thus $D(f_t|\mathbb{R}^n_x)$ is non-singular everywhere and so, for all t, f_t is a diffeomorphism, as required.

The extension to a parameterized family $f^u, u \in B^k$ is essentially the same as in the previous two cases, after choosing κ to be less than the infimum of $\kappa_1 \kappa_2$ taken over all $u \in B^k$. Note that for u near ∂B^k, where f^u is already an orthogonal bundle map, each T_x will be the

identity, so, regardless of the value of t or y in the definition $(f_t | \mathbb{R}^n_x) y = (I_n + g_t(y)(T_x - I_n)) y$, the function is constantly the identity. $\square$

Stage 6: From preserving H' to preserving H_2.

The previous stages allow us to define a possibly very thin regular neighborhood $H' \subset H_2$ of $\partial H \cup I_0$ and a relative homotopy of Θ to a map $(B^k, \partial B^k) \to (\text{Diff}(H, H', I_0), \text{Diff}(H, H_2, I_0))$. Continue to denote the result as Θ and invoke this special case of Hatcher's powerful theorem:

Lemma A.7. *Suppose F is a closed orientable surface. Any $\psi : S^k \to \text{Diff}(F \times I, F \times \{0\})$ is homotopic to a map so that each $\psi(u) : F \times I \to F \times I$ respects projection to I. (In fact, unless F is a torus or a sphere, there is a diffeomorphism $f : F \to F$ and a homotopy of ψ to a map so that each $\psi(u)$ is just $f \times id_I$.)*

Proof. The special case in which F is a torus or a sphere is left to the reader; it will not be used. Pick a base point $u_0 \in S^k$ and let $\psi_0 = \psi(u_0) : F \times I \to F \times I$. Take f in the statement of the Lemma to be $\psi_0 | (F \times \{0\}) : F \to F$ and, with no loss of generality, assume this diffeomorphism is the identity. Since $\text{Diff}(F)$ is contractible [EE], the map $\psi_| : S^k \to \text{Diff}(F)$ defined by $\psi_|(u) = \psi(u) | (F \times \{0\})$ can be deformed so that each $\psi_|(u)$ is the identity. The homotopy of $\psi_|$ induces a homotopy of ψ via the isotopy extension theorem.

The map $p_1 \psi_0 : F \times I \to F$ defines a homotopy from $\psi_0 | (F \times \{1\}) : F \to F$ to the identity, and this implies that $\psi_0 | (F \times \{1\})$ is isotopic to the identity. The previous argument applied to diffeomorphisms of $F \times \{1\}$ instead of $F \times \{0\}$ then provides a further homotopy of ψ, after which each diffeomorphism $\psi(u) : F \times I \to F \times I$ is the identity on $F \times \{0, 1\} = \partial(F \times I)$. That is, after the homotopy of ψ, ψ maps S^k enirely into the space of diffeomorphisms of $F \times I$ that are the identity on $\partial(F \times I)$. The lemma then follows from the central theorem of [Ha1]. $\square$

The region $H_2 - \text{int}(H')$ between the regular neighborhoods of $\partial H \cup I_0$ is a collar which can be parameterized by $\Sigma \times I$. With this in mind, apply Lemma A.7 to $\psi = \Theta | \partial B^k : S^{k-1} \to \text{Diff}(H_2 - \text{int}(H'))$, extending the parameterization slightly outside of $H_2 - \text{int}(H')$ via an argument like that in Stage 2 above. We then have a parameterization $\Sigma \times \mathbb{R}$ of a neighborhood of $H_2 - \text{int}(H')$ so that ∂H_2 corresponds to $\Sigma \times \{1\}$, $\partial H'$ corresponds to $\Sigma \times \{0\}$, and so that for each $u \in \partial B^k$, the restriction of $\Theta(u) : H \to H$ to $\Sigma \times \mathbb{R}$ respects projection to $\mathbb{R}$. Let $g_t : \mathbb{R} \to \mathbb{R}$ be a smooth isotopy with compact support from

the identity to a diffeomorphism that takes $1 \in \mathbb{R}$ to 0. Define the isotopy $r_t : \Sigma \times \mathbb{R} \to \Sigma \times \mathbb{R}$ by $r_t(y, s) = (y, g_t(s))$ and extend by the identity to an isotopy of H. The result is an isotopy from the identity to a diffeomorphism that takes H_2 to H'. Then the deformation Θ_t of Θ defined by $\Theta_t(u) = r_t^{-1} \Theta(u) r_t : H \to H$ pairwise homotopes $\Theta : (B^k, \partial B^k) \to (\text{Diff}(H, I_0), \text{Diff}(H, H_2, I_0))$ to a map whose image lies entirely in $\text{Diff}(H, H_2, I_0)$.

REFERENCES

[Ak] E. Akbas, A presentation for the automorphisms of the 3-sphere that preserve a genus two Heegaard splitting, *Pacific J. Math.* **236** (2008) 201222.

[Bi] J. Birman, Braids, links and mapping class groups, Annals of Mathematics Studies **82**. Princeton University Press, 1974.

[BJ] T. Bröcker and J. Jänich, Introduction to differential topology, Cambridge, 1973.

[Cho] Cho, Homeomorphisms of the 3-sphere that preserve a Heegaard splitting of genus two, *Proc. Amer. Math. Soc.* **136** (2008) 11131123

[EE] C. Earle and J. Eells, A fibre bundle description of Teichmüller theory, *J. Differential Geometry* **3** (1969) 1943.

[ES] C. Earle and A. Schatz, Teichmller theory for surfaces with boundary, *J. Differential Geometry* **4** (1970) 169185.

[FN] E. Faddell and L. Neuwirth, Configuration spaces, *Math. Scand.*, **10** (1962) 111–118.

[Ga] D. Gabai, Foliations and the topology of 3-manifolds. III, *J. Differential Geom.* , **26** (1987) 479536.

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

[Gr] A. Gramain, Le type dhomotopie du groupe des diffeomorphismes dune surface compacte, *Ann. Scient. Ec. Norm. Sup.* , **6** (1973) 5366.

[HKMR] S. Hong, J. Kalliongis, D. McCullough, J. H. Ruginstein, Di?eomorphisms of Elliptic 3-Manifolds, to appear.

[Ha1] A.. Hatcher, Homeomorphisms of sufficiently large P^2-irreducible 3-manifolds, *Topology* **11** (1976) 343-347.

[Ha2] A.. Hatcher, A proof of the Smale Conjecture, *Annals of Mathematics*, **117** (1983) 553–607.

[JM] J. Johnson and D. McCullough, The space of Heegaard splittings, arXiv:1011.0702.

[Pa] R. Palais, Local triviality of the restriction map for embeddings, *Comment. Math. Helv.*, **34** (1960) 305–312.

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

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

[ST1] M. Scharlemann, A. Thompson, Heegaard splittings of (surface) x I are standard, *Math. Ann.* **295** (1993) 549–564.

[ST2] M. Scharlemann, A. Thompson, Thin position and Heegaard splittings of the 3-sphere, *J. Differential Geom.* **39** (1994), 343–357.

[Wa] B. Wajnryb, Mapping class group of a handlebody, *Fund. Math.* **158** (1998), 195–228.

Martin Scharlemann, Mathematics Department, University of California, Santa Barbara, CA USA

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