COMMUNICATIONS IN
ANALYSIS AND GEOMETRY
Volume 17, Number 5, 851–901, 2009

Invariant Heegaard surfaces in manifolds with involutions and the Heegaard genus of double covers

Yo'av Rieck and J. Hyam Rubinstein

Let M be a 3-manifold admitting a strongly irreducible Heegaard surface Σ and $f : M \to M$ an involution. We construct an invariant Heegaard surface for M of genus at most $8g(\Sigma) - 7$. As a consequence, given a (possibly branched) double cover $\pi : M \to N$ we obtain the following bound on the Heegaard genus of N:

$$g(N) \leq 4g(\Sigma) - 3.$$

We also get a bound on the complexity of the branch set in terms of $g(\Sigma)$. If we assume that M is non-Haken, by Casson and Gordon [3] we may replace $g(\Sigma)$ by $g(M)$ in all the statements above.

1. Statements of results

We study the behavior of Heegaard surfaces under (possibly branched) double covers $\pi : M \to N$. It is easy to lift any Heegaard surface of N to a Heegaard surface of M and see that the Heegaard genus of the cover is bounded above: $g(M) \leq 2g(N) + b - 1$. Here $g(\cdot)$ denotes the genus of a surface or the Heegaard genus of a 3-manifold and b is the bridge index of the branch set with respect to a minimal genus Heegaard surface for N. (This upper bound easily generalizes to any p-fold cover $\pi : M \to N$, provided that the branch set is a 1-manifold: $g(M) \leq pg(N) + (p - 1)b - 1$; see, for example, [2, Chapter 11].)

For the converse, we need a strongly irreducible Heegaard surface for M, say Σ. Since any double cover is regular, it is given as the quotient under an involution $f : M \to M$. (The involution f is easy to describe: send any point $p \in M$ to the other point $q \in M$ that projects to that same point under π; if no such q exists leave p fixed.) Using an invariant version of Cerf theory we get Σ to intersect $f(\Sigma)$ "nicely," and then use $\Sigma \cup f(\Sigma)$ to construct a complex C with a list of useful properties (Properties 9.2). C is used to construct an invariant Heegaard surface for M and bound its genus; the projection of this surface gives the Heegaard surface for N, and estimating its genus we get a linear upper bound for the genus of the quotient manifold in terms of $g(\Sigma)$.

We now give the precise statements of our main results.

Remark 1.1. As is well-known, Casson and Gordon [3] proved that if M is an irreducible, non-Haken manifold then any minimal genus Heegaard surface for M is strongly irreducible. Thus if M is non-Haken we can replace $g(\Sigma)$ by $g(M)$ in all the statements below.

Theorem 1.1 (Invariant Heegaard surface). *Let M be an irreducible, orientable, a-toroidal, non-Seifert fibered manifold of Heegaard genus at least two admitting an orientation preserving involution f and a strongly irreducible Heegaard surface Σ.*

Then M has an invariant Heegaard surface of genus at most $8g(\Sigma) - 7$. Moreover, each handlebody obtained by cutting M open along this surface is invariant.

Theorem 1.2 (Genus of double covers). *Let M be an irreducible, orientable, a-toroidal, non-Seifert fibered manifold of Heegaard genus at least two admitting a strongly irreducible Heegaard surface Σ. Let N be an orientable manifold and $\pi : M \to N$ a double cover. Then we have*

$$g(N) \leq 4g(\Sigma) - 3.$$

Using the invariant Heegaard surface for M constructed in Theorem 1.1 we obtain a bound on the complexity of the branch set. This bound is given in terms of the bridge number of the branch set with respect to the Heegaard surface for N given in Theorem 1.2, i.e., the projection of the invariant Heegaard surface for M. The definition of bridge number with respect to a Heegaard surface is given in Definition 12.1 (for a detailed discussion see, for example, [12] or [10]). We prove:

Theorem 1.3. *Let M be an irreducible, orientable, a-toroidal, non-Seifert fibered manifold of Heegaard genus at least two admitting a strongly irreducible Heegaard surface Σ. Let N be an orientable manifold and $\pi : M \to N$ be a double cover. Denote the bridge index of the branch set with respect to the surface found in Theorem 1.2 by b. Then we have*

$$b \leq 8g(\Sigma) - 6.$$

For proving Theorem 1.1 we study the intersection of strongly irreducible Heegaard surfaces, that is, the intersection of Σ and its image under the involution $f(\Sigma)$. However, our work can be applied for any two strongly irreducible Heegaard surfaces $\Sigma_1, \Sigma_2 \subset M$. We say that two embedded surfaces *intersect essentially* if their intersection is transverse and every curve of $\Sigma_1 \cap \Sigma_2$ is essential in both surfaces. Rubinstein and Scharlemann studied the intersection of strongly irreducible Heegaard surfaces; we build on their work and prove:

Theorem 1.4. *Let M be an irreducible, orientable, a-toroidal, non-Seifert fibered manifold of Heegaard genus at least two. Suppose that either M admits two strongly irreducible Heegaard surfaces Σ_1 and Σ_2 or a strongly irreducible Heegaard surface Σ and an orientation preserving involution f. Then we have:*

(1) *Σ_1 and Σ_2 can be isotoped to intersect essentially and so that every component of M cut open along $\Sigma_1 \cup \Sigma_2$ is a handlebody.*

(2) *Σ can be isotoped so that Σ and $f(\Sigma)$ intersect essentially and so that every component of M cut open along $\Sigma \cup f(\Sigma)$ is a handlebody.*

Theorem 1.4 follows quite easily from Theorem 7.1 (page 873) which is based on and improves results of Rubinstein and Scharlemann, see Remark 7.1. We do not state this theorem here to avoid terminology that had not yet been introduced.

Another tool used in the proof of Theorem 1.1 is the creation of an invariant complex $\mathsf{C} \subset M$ fulfilling a list of properties described in Properties 9.2. Complexes fulfilling Properties 9.2 are called *an-annular complexes.* Properties 9.2 imply that C has the following structure: it is constructed from a finite collection of disjointly embedded tori (say $\{T_i\}_{i+1}^n$) bounding disjointly embedded solid tori (say $\{V_i\}_{i+1}^n$) and a collection of disjointly embedded compact (but not closed) surfaces of *negative* Euler characteristic[1] with their boundary on the tori T_i; all the boundary components form essential curves on the tori. Properties 9.2 bound the Euler characteristic of C and state that M cut open along C consists of handlebodies. We refer the reader to Section 9 for a precise description of C and Properties 9.2, and the statement and proof of Theorem 9.1 where we prove the existence of C.

Remark. Section 9 is based on [17] where Rieck proved the existence of an-annular complexes in manifolds admitting two distinct strongly irreducible Heegaard surfaces.

[1] Getting the Euler characteristic of this surface to be negative (as opposed to non-positive) is the main challenge of the construction and the reason for the name *an-annular*.

Naturally, the solid tori $\{V_i\}_{i+1}^n$ can be viewed as an equivariant link in M. Not every link in M can be realized in this way and we ask which links are (Question 9.1).

This article is written in sections whose order, for the most part, reveals the logic of the proof. It is outlined in the next section.

2. Outline

Section 3: Background material, notation, etc.

Section 4: We give examples of higher order covers to demonstrate where our techniques fail to generalize. We also give examples that show the difficulty in finding invariant reductions of various types (reducing sphere, weak reductions and destabilizations).

Section 5: We give a description of Heegaard functions (our version of sweepouts) and define the Graphic. Because of the invariance requirement the Graphic cannot be assumed to be generic and this is rectified in Proposition 5.1 that shows that the behavior of the Graphic is essentially the same as the behavior of generic graphics.

Section 6: We isotope a strongly irreducible Heegaard surface $\Sigma \subset M$ to intersect its image under the involution in a compression free way, i.e., Σ and its image provide no compressions for each other, yet their intersection contains an essential curve. In the end of this section we construct the generic interval, and isotopy of the Heegaard surface and its image that has the properties needed for Section 7.

Section 7: Using the generic interval we ensure Σ is chopped up completely by its image and a set of compressing disks for the image. We also eliminate inessential simple closed curves of intersection between Σ and its image (that is, we isotope Σ to intersect its image essentially and spinally). We also show that if a manifold M admits two strongly irreducible surfaces Σ_1 and Σ_2 (but not necessarily an involution), then Σ_1 and Σ_2 can be isotoped to intersect essentially and spinally.

Section 8: Proof of Theorem 1.4.

Section 9: We consider Σ union its image as a complex and modify it to get rid of undesired annuli, proving (Theorem 9.1) existence of the complex C fulfilling Properties 9.2.

Section 10: Using this complex we create an invariant Heegaard surface for M and estimate its genus, thus proving Theorem 1.1.

Section 11: Using this Heegaard surface and the Equivariant Disk Theorem we get a Heegaard surface for the quotient thus proving Theorem 1.2.

Section 12: Using the bounded genus invariant surface found in Section 10 we bound the complexity of the branch set in terms of the genus of M.

3. Background

We work in the smooth and orientable category. By *manifold* we mean a 3-dimensional compact manifold without boundary. We follow standard notation for 3-manifolds: $\operatorname{int} X$ is the interior of X, $\operatorname{cl} X$ is the closure of X, ∂X is the boundary of X, etc. See [7] or [9] for basic definitions. We refer the reader to [20] for a detailed discussion about Heegaard splittings. We assume that our manifold M is not a Seifert fibered space. We note that for Seifert fibered spaces results far more refined than ours are known, e.g., for S^3 the positive solution of the Smith Conjecture [14], Hodgson and Rubinstein's work about lens spaces [8], Boileau and Otal's work about small Seifert fibered spaces, [1], and Scott's work about Haken Seifert fibered spaces [22].

We further assume that our manifold contains a strongly irreducible Heegaard surface, i.e., M has a Heegaard surface for which any two compressing disks on opposite sides intersect. By Haken [6] our manifold is irreducible, i.e., every 2-sphere embedded in M bounds a ball. This condition is not vacuous: Casson and Gordon's [3] seminal work show that for irreducible non-Haken manifolds every minimal genus (indeed, any irreducible) Heegaard surface is strongly irreducible. (Non-Haken manifolds are not the only manifolds that contain strongly irreducible Heegaard splittings; see [11] for manifolds admitting both weakly reducible and strongly irreducible minimal genus Heegaard splittings.) We note, however, that constraints are imposed on the cover and not on the manifold being covered, where no additional constraints apply.

Suppose that $p : M \to N$ is a double cover. Since all double covers are regular (including branched double covers), there exists $f : M \to M$ an involution on M, so that p is given by the natural projection $M \to M/(f) \cong N$. A subset $S \subset M$ is called *invariant* if $f(S) = S$. $S \subset M$ is called *equivariant* if S is either invariant or disjoint from its own image, i.e., either $f(S) = S$ or $f(S) \cap S = \emptyset$. We use the notation $N(S)$ to mean a normal neighborhood. When discussing an invariant (resp. equivariant) subset of M, we use $N(S)$ to denote an invariant (resp. equivariant) normal neighborhood.

Since all manifolds are assumed to be orientable, f is orientation preserving.

The first half of this paper deals with the intersection of embedded surfaces. We follow the terminology used in [18]. In particular, a tangency between two embedded surfaces can have one of two forms: a *center* (modeled on the intersection of $z = 0$ and $z = x^2 + y^2$) or saddle (modeled on the intersection of $z = 0$ and $z = x^2 - y^2$ at the origin).

4. Examples

Now an example. We consider Solv manifolds (definition below, see also [21]) since they cover each other generously and in [5] Cooper and Scharlemann gave a complete classification of their Heegaard surfaces.

Definition 4.1. A 3-manifold is called *Solv* if it is a torus bundle over S^1 with Anosov monodromy, i.e., the monodromy has infinite order and no power of it has fixed point in $\pi_1(T^2)$.

Given a Solv manifold M (say with monodromy ϕ) and a positive integer n, the Solv manifold with monodromy ϕ^n (denoted M_n) is an n-fold cover of M. This cover is as nice as one could hope for: cyclic (in particular regular) and unbranched. As it is our goal to get invariant Heegaard surfaces (Theorem 1.1), it is interesting to consider a minimal genus Heegaard surface for M_n, say Σ. Σ and its image under a generator of the action of $\mathbb{Z}/(n)$ are Heegaard surfaces of genus 2 or 3. By picking M_n correctly, Cooper and Scharlemann [5] show that Σ and its image under the generator of the action are isotopic. But for a cyclic group action invariance under a generator implies invariance under the entire group; may we conclude that the surface is invariant?

No. For if it were, for larger and larger values of n we would get surfaces of genus 2 or 3 that are invariant under the free action of a cyclic group of arbitrarily high order. This of course cannot be, since the quotient surface would have fractional Euler characteristic (so a surface invariant under a group action of high order must have high genus). Σ "equals" its image in the sense of "up-to-isotopy," but this isotopy cannot be realized invariantly and therefore it does not provide us with a surface that is truly invariant under the group action. Moreover, it is an easy exercise to get Σ to be disjoint from its image. Yet during the isotopy that takes Σ to its image the two are no longer each other's images. This phenomenon occurs in a very simple setting as well: consider S^1 double covering itself. The preimage of a point is two points which are isotopic to each other but the isotopy cannot be realized invariantly. This example demonstrates that generalizing this work for higher order coverings will not be a straightforward task but will require a new ingredient, perhaps the degree of the cover.

We use Solv manifolds to provide one more example. Suppose M is a Solv manifold of genus two, $\Sigma \subset M$ a minimal genus Heegaard surface, and M_n its n-fold cyclic cover (as above). Since M is irreducible Σ is strongly irreducible. Lifting Σ to M_n we get a Heegaard surface of genus $n+1$, say Σ_n. By Cooper and Scharlemann [5] the only irreducible Heegaard surfaces for M_n are minimal genus Heegaard surfaces. Therefore Σ_n destabilizes $n-1$ or $n-2$ times, which gives many distinct collections of reducing, weakly reducing and destabilizing disks for Σ_n. However, strong irreducibility of Σ implies that no such reducing set can be made equivariant under the cyclic group action. Therefore, either on at least one side the disks are not equivariant, or the disks are equivariant on both sides but in the projection of the disks to M every pair of disks on opposite sides of Σ intersect at least twice (note that the image of disjoint curves on Σ_n may intersect more than once).

5. Heegaard Functions and the Graphic

In this section we introduce the basic set up, beginning with the following definition that formalizes the basic tool we use for studying Heegaard surfaces. It is equivalent to the notion of *sweepout*, as defined in [18]. (Much of the material in this section is not new but is included here for our work in the equivariant setting.)

Definition 5.1. Let M be a manifold. A smooth function $h : M \to [-\infty, \infty] = \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$ is called a *Heegaard function* if the following hold:

(1) $h^{-1}(-\infty)$ and $h^{-1}(\infty)$ are graphs;

(2) $h|_{h^{-1}(\mathbb{R})}$ has no critical points.

Condition (2) implies that for any $t \in \mathbb{R}$ $h^{-1}(t)$ is a smooth surface and its genus is independent of t. In fact, any two such surfaces (say $h^{-1}(s)$ and $h^{-1}(t)$, with $s < t$) are parallel and the region defining the parallelism is given by $h^{-1}([s,t])$.

Definition 5.2.

(1) A surface $\Sigma \subset M$ is called a *Heegaard surface* if it is of the form $h^{-1}(0)$ for some Heegaard function h.

(2) A *spine* for a Heegaard surface Σ is a (disconnected) embedded graph of the form $h^{-1}(-\infty) \sqcup h^{-1}(\infty)$.

We will often start with a Heegaard surface, and then consider a Heegaard function that gave rise to it, i.e., we shall start with Σ and consider h as in Definition 5.2. This function will be called a "corresponding Heegaard function". It is by no means unique, nor is the spine.

Let M be a manifold and $f : M \to M$ an involution. Let Σ be a Heegaard surface for M, and h a corresponding Heegaard function. When studying Σ and $f(\Sigma)$ the Heegaard function corresponding to $f(\Sigma)$ we will use is $h \circ f$ (note that $f = f^{-1}$).

We will use a Cerf theoretic argument, which requires the construction of the Graphic. The Graphic is based on a 2-parameter family of surfaces, i.e., the assignment of two surfaces for every point in the parameter square $\{(s,t) : s,t \in [-\infty,\infty]\}$, denoted $(s,t) \mapsto (F_1(s,t), F_2(s,t))$. The Graphic itself is the subset of points corresponding to surfaces that do not intersect transversely. See [18] or [16] for further details about the Graphic, or Cerf's original work [4]. (We give a more detailed description of the Graphic below.)

In our case, given h a Heegaard function for M and f an involution on M, we start with the assignment: $(s,t) \mapsto (\Sigma_s, f(\Sigma_t))$ where $\Sigma_s = h^{-1}(s)$ and $f(\Sigma_t) = f(h^{-1}(t))$. Note that on the diagonal $\{s = t\}$ the involution exchanges the two surfaces. This assignment is not necessarily generic and therefore no niceness properties of the Graphic can be assumed (not even one dimensionality). To that end, we modify the surfaces. First, and most importantly, we modify the surfaces along the diagonal, as in [8]. Via perturbation, we impose the following two conditions on h: the spines $h^{-1}(\pm\infty)$ are disjoint from their images, and zero must be a regular value of $h - h \circ f$.[2] Since both conditions are generic after imposing the first condition on the spines we can impose the second without ruining the first. In [8] it was shown that these conditions imply the following:

Conditions 5.1.

(1) *For all but finitely many values of t the intersection of Σ_t and $f(\Sigma_t)$ is transverse. Points that correspond to non-transverse intersection are called critical. At a critical point exactly one of the following holds:*

(2) *Σ_t and $f(\Sigma_t)$ intersect in a single non-degenerate critical point that is fixed by f (then t is called a simple critical point).*

---

[2]This condition is quite natural: we are interested in the intersection of the surface $h^{-1}(t)$ with its image; therefore we are forced to look at points where h and $h \circ f$ have the same value.

(3) Σ_t *and* $f(\Sigma_t)$ *intersect at a pair of non-degenerate critical points that are exchanged by* f *(then* t *is called a double critical point).*

Before modifying the 2-parameter family off the diagonal to get the surfaces to be as generic as we can, let us explain what we mean by genericity. It is a local property, i.e., given a point (s_0, t_0) it only depends on surfaces for (s, t) close to (s_0, t_0). In particular we can impose it on an open set, which we shall do (the complement to the diagonal, to be precise). The condition is this: for a dense open set of points (s, t) the surfaces intersect transversely. The points where the surfaces intersect non-transversely fall into three categories.

Edges: One-dimensional sets in the parameter square, with finitely many components each homeomorphic to an interval. The points of the edges are those that correspond to pairs of surfaces having exactly one critical point.

Vertices of valence four: Finitely many points in the parameter square that correspond to pairs of surfaces with exactly two critical points. Each valence four vertex is the endpoint of exactly four edges, more precisely, two pairs of edges where each pair corresponds to a tangency between the two surfaces. We can also consider each pair of edges as one long edge, pasting them together at the vertex. Then the vertex is the point where the two edges cross each other transversely.

Death–birth vertices: Finitely many vertices of valence two. As they play no role whatsoever in this work, so we do not describe them here.

So a generic Graphic forms a finite graph embedded in the parameter square. We now prepare the Graphic: starting with $(s, t) \mapsto (\Sigma_s, f(\Sigma_t))$, on one side of the diagonal (say $s > t$) we perturb the surfaces to be generic. We may do so without changing the diagonal: we take any generic perturbation of the surfaces at $s > t$, say given by $(s, t) \mapsto (F_{1,r}(s, t), F_{2,r}(s, t))$ so that at $r = 0$ we have our original assignment, and we pick a perturbation given by $(s, t) \mapsto (F_{1,r'}(s, t), F_{2,r'}(s, t))$, with r' a function of (s, t) that limits on zero as (s, t) approaches the diagonal. If all the above is done generically, we have an assignment that is generic at $s > t$, fulfills Conditions 5.1 on the diagonal, and is continuous on $s \geq t$. For $s < t$, we set $F_1(s, t) = F_2(t, s)$ and $F_2(s, t) = F_1(t, s)$. (In other words, we perturb the surfaces in the domain $s < t$ in the exact same way we did in the domain $s > t$.) Note that f exchanges F_1 and F_2: $f(F_1(s, t)) = F_2(t, s)$ and $f(F_2(s, t)) = F_1(t, s)$. Hence f induces the involution $(s, t) \mapsto (t, s)$ on the parameter square and the Graphic is invariant under this involution (in general, this forces double critical points on the diagonal).

Remark 5.1. We work mostly on the diagonal, where the surfaces are parameterized by a single parameter t, explicitly: $\Sigma_t = F_1(t,t)$ and $f(\Sigma_t) = F_2(t,t)$. For simplicity, while considering points on the diagonal we use the notation Σ_t and $f(\Sigma_t)$.

We conclude this section with the following proposition, which discusses the behavior of the Graphic near the diagonal. When the Graphic intersects the diagonal, this proposition is needed due to lack of transversality. Here and throughout this work we move freely between a point (s,t) of the parameter square and the corresponding surfaces $F_1(s,t)$ and $F_2(s,t)$, and between edges on the Graphic and the corresponding tangencies of $F_1(s,t)$ and $F_2(s,t)$.

Proposition 5.1. *Let (t_0, t_0) be a point on the diagonal that corresponds to surfaces with two critical points, and let S_1 and S_2 be the two curves of the Graphic through it, each corresponding to one of the critical points. Then locally about (t_0, t_0) one of the following holds:*

(1) *S_1 is on one side of the diagonal (except for (t_0, t_0)) and its image S_2 on the other;*

(2) *S_1 and S_2 cross each other.*

Proof. By Conditions 5.1, locally the curve S_1 has only one point on the diagonal. Assume (1) does not occur. Therefore, one of the two curves (say S_1) crosses the diagonal. If $S_1 = S_2$ near (t_0, t_0) then the two critical points are in fact the same (since off the diagonal the Graphic is generic), contrary to our assumption. The proposition follows from the fact that S_2 is the image of S_1 under the involution $(s,t) \to (t,s)$. □

Remarks.

(1) In the first case, while traveling along the diagonal, it is as if we did not encounter a critical point at all, as the diagonal is only tangent to the two but never traverses them. We may ignore such points throughout this work.

(2) In the second case the intersection behaves as if the Graphic is generic.

6. Compression Free Intersection

Following Rubinstein and Scharlemann [18] we define:

Definitions 6.1. Let F_1 and F_2 be surfaces embedded in a 3-manifold intersecting transversely.

(1) A curve of $F_1 \cap F_2$ is called *essential* (resp. *inessential*) if it is essential (resp. inessential) in both F_1 and F_2. A curve of $F_1 \cap F_2$ which is essential on one surface and inessential on the other is called a *compression*.

(2) The intersection of F_1 and F_2 is called *compression free* if no curve of $F_1 \cap F_2$ is a compression.

Recall (Remark 1.1) that if M is non-Haken then any minimal genus Heegaard surface is strongly irreducible; hence the theorem below is not vacuous:

Theorem 6.1. *Let M be an irreducible, orientable, a-toroidal, non-Seifert fibered manifold of Heegaard genus at least two with an orientation preserving involution $f : M \to M$. Let Σ be a strongly irreducible Heegaard surface for M, and $h : M \to \mathbb{R}$ a corresponding Heegaard function. Then there exists an interval $(a, b) \subset \mathbb{R}$ so that the following conditions hold:*

(1) *For any regular point $t \in (a, b)$ the intersection of $\Sigma_t = h^{-1}(t)$ with $f(\Sigma_t)$ is compression free, yet contains an essential curve.*

(2) *a is critical, and arbitrarily close to (a, a) there are points corresponding to transverse intersections that are not compression free or are all inessential[3] (similarly for b).*

(3) *For any $t \in (a, b)$, (t, t) has a neighborhood U so that every regular point in U corresponds to a compression free intersection containing an essential curve.*

Proof. Color the handlebody $h^{-1}([-\infty, s])$ purple and $h^{-1}([s, \infty])$ yellow.

For convenience we present the diagonal as an interval by identifying any point (t, t) on the diagonal with $t \in \mathbb{R}$. Subdivide the interval $[-\infty, \infty]$ into *layers* separated by critical points (recall their definition in Conditions 5.1). A component of the parameter square cut open along the Graphic is called a *region*. Note that every layer is contained in exactly one region. Since the intersection pattern between Σ_t and $f(\Sigma_t)$ is independent of the choice of point within a region (resp. a layer), we call it *the intersection of the region* (resp. the layer). We label regions and layers, exactly as in [18]. By construction $F_1(s, t)$ is a small perturbation of Σ_s and therefore the two handlebodies complementary to $F_1(s, t)$ inherit yellow and purple coloring.

[3]These points may be off the diagonal.

Definitions 6.2. Let R be a region of the Graphic and $(s,t) \in R$, with corresponding surfaces $F_1(s,t)$ and $F_2(s,t)$.

(1) R is labeled $\mathbf{P}$ if there exists a disk $D_{\mathbf{P}} \subset F_2(s,t)$ and the following conditions holds:
   (a) The boundary of $D_{\mathbf{P}}$ is an essential curve of $F_1(s,t)$.
   (b) Near its boundary $D_{\mathbf{P}}$ is purple.
   (c) $\operatorname{int} D_{\mathbf{P}} \cap F_1(s,t)$ does not contain essential curves of $F_1(s,t)$.

(2) R is labeled $\mathbf{p}$ whenever the following conditions holds:
   (a) The intersection of $F_2(s,t)$ and the yellow handlebody contains an essential curve of $F_2(s,t)$.
   (b) Every curve of $F_1(s,t) \cap F_2(s,t)$ is inessential.

(3) The labels $\mathbf{Y}$ and $\mathbf{y}$ are defined similarly.

(4) A layer is labeled by the same labels as a region that contains it.

Remark. Rubinstein and Scharlemann define their labels in [18, pp. 1009–1010]. To see that their labels are indeed the same as ours, denote $F_1(s,t)$ by P, $F_2(s,t)$ by Q, the purple handlebody by A and the yellow handlebody by B. Then the labels $\mathbf{P}$, $\mathbf{p}$, $\mathbf{Y}$, and $\mathbf{y}$, correspond to the labels A, a, B, and b in [18] (in the same order). The label X, x, Y and y appearing in [18] are not needed here (essentially, because the surface $f(\Sigma_t)$ is the image of Σ_t and has the same intersection properties).

From [18, Sections 4 and 5] we have the proposition below for regions of the Graphic, that is applicable directly for simple critical points of the diagonal, as layers separated by simple points are contained in regions that share an edge. The goal of this section can be described as extending this proposition to double critical points (we remark that in general this cannot be done, and we will need to use the involution).

Proposition 6.1. *Every region of the Graphic has at most one label. A region labeled* $\mathbf{p}$ *or* $\mathbf{P}$ *cannot share an edge with a region labeled* $\mathbf{y}$ *or* $\mathbf{Y}$. *Therefore every layer has at most one label and a layer labeled* $\mathbf{p}$ *or* $\mathbf{P}$ *cannot be separated by a simple critical point from a layer labeled* $\mathbf{y}$ *or* $\mathbf{Y}$.

In Lemmas 6.1 and 6.2 we study lowercase labels:

Lemma 6.1. *A region R is labeled with a lowercase label if and only if the corresponding surfaces intersect in inessential curves only. In that case,* *except for punctures the surface $F_2(s,t)$ has one color, purple if the label is* **y** *and yellow if the label is* **p***.*

Proof. This information is contained in [18] so we paraphrase it here. If a region has a lowercase label then by definition the intersection consists of entirely inessential curves. Conversely, if the intersection consists of inessential curves only, except perhaps for punctures $F_2(s,t)$ is colored in one color, yellow or purple (resp.). Any essential curve of $F_2(s,t)$ can be isotoped off the punctures, showing the label is **p** or **y** (resp.). □

Lemma 6.2. *There does not exist a critical level t_0 separating a layer labeled* **y** *from a layer labeled* **p***.*

Proof. Suppose for contradiction there exists a critical level t_0 that separates a layer l_y labeled **y** from a layer l_p labeled **p**. By Proposition 6.1, t_0 is a double critical point. It is easy to see that **p** does not change when crossing centers, hence t_0 is a double saddle. By Lemma 6.1, for $t_y \in l_y$ exactly one component of $f(\Sigma_{t_y}) \cap h^{-1}([-\infty, t_y])$ (say F) contains all of $f(\Sigma_{t_y})$ except, perhaps, for punctures. After the double saddle, no essential curve of $f(\Sigma_{t_p})$ is contained in $h^{-1}([-\infty, t_p])$. Since crossing each saddle changes $f(\Sigma_{t_y}) \cap h^{-1}([t_y, \infty])$ by adding or removing a single 1-handle, we see that all of F (except perhaps for punctures) was moved out of $h^{-1}([-\infty, t_y])$ and into $h^{-1}([t_p, \infty])$ by two 1-handles; hence, F is a punctured torus (resp. pair of pants) and Σ_t is a torus (resp. sphere), contradicting the assumptions of Theorem 6.1. □

In Lemma 6.3 we study uppercase labels:

Lemma 6.3. *There does not exit a critical level t_0 separating a layer labeled* **Y** *from a layer labeled* **P***.*

Proof. Assume for contradiction t_0 is a critical level separating a layer labeled **P** (say l_p) from a layer labeled **Y** (say l_y). For convenience of presentation, we assume that l_p is below t_0 and l_y is above t_0 (the other case is treated by taking ϵ below to be a small negative number). By Proposition 6.1 t_0 is a double critical point. As above, it is easy to see that uppercase labels do not change when crossing centers; hence we may assume that t_0 corresponds to two saddles (say S_1 and S_2). From Definition 6.2 we see that for $t \in l_p$, there exists a compressing disk (say $D_{\mathbf{P}} \subset f(\Sigma_t)$) giving rise to the label **P**.

Similarly, for $t \in l_y$ there exists a compressing disk $D_{\mathbf{Y}}$ giving the label $\mathbf{Y}$. We consider two cases:

Case 1. One of the saddles does not destroy one of the disks ($D_{\mathbf{Y}}$ *or* $D_{\mathbf{P}}$): Leaving the diagonal and crossing the edge we hypothesized not to destroy one of the disks, we see that a region labeled $\mathbf{Y}$ is adjacent to a region labeled $\mathbf{P}$ along an edge of the Graphic, contradicting Proposition 6.1.

Case 2. Each saddle S_1 *and* S_2 *destroys both* $D_{\mathbf{Y}}$ *and* $D_{\mathbf{P}}$: Let $\gamma = \partial D_{\mathbf{P}}$. As we approach t_0 the curve γ limits on both saddles, or we would be in Case 1. Since the saddles are involutes of each other, $f(\gamma)$ must limit on both saddles as well. There are two subcases:

Subcase 2a: $\gamma = f(\gamma)$. Since $\gamma = \partial D_{\mathbf{P}}$ and $D_{\mathbf{P}} \subset f(\Sigma_t)$, we see that $\partial f(D_{\mathbf{P}}) = f(\partial D_{\mathbf{P}}) = f(\gamma) = \gamma$. But $f(D_{\mathbf{P}}) \subset f(f(\Sigma_t)) = \Sigma_t$. Therefore γ is inessential in Σ_t, contradicting Definition 6.2.

Subcase 2b: $\gamma \neq f(\gamma)$. Since $f(\gamma)$ must limit on both saddles as well, γ limits on each saddle once only.

Let $\epsilon > 0$ be a small number. Let v be a non-vanishing vector field along γ that is everywhere transverse to γ and is tangent to $\Sigma_{t_0-\epsilon}$. We may assume v points towards S_1 as ϵ tends to zero (else we reverse it). The curve γ is called *untwisted* if v points towards S_2 as ϵ tends to zero, *twisted* otherwise (i.e., if v points away from S_2 as ϵ tends to zero); note that this is independent of choice of v. Likewise, $f(\gamma)$ may be twisted or untwisted; when examining twistedness of $f(\gamma)$ we regard it as a curve on $\Sigma_{t_0-\epsilon}$, not on $f(\Sigma_{t_0-\epsilon})$. We show that γ is twisted if and only if $f(\gamma)$ is: suppose γ is untwisted. Near S_1 and S_2 we can view $\Sigma_{t_0-\epsilon}$ as a flat disk in $z = -\epsilon$, and $f(\Sigma_{t_0-\epsilon})$ as a small piece of $z = x^2 - y^2$ forming a little arch. We denote the handlebodies given by M cut open along $f(\Sigma_{t_0-\epsilon})$ by H_1 and H_2. By construction v is tranverse to $f(\Sigma_{t_0-\epsilon})$ and after renaming H_1 and H_2 if necessary we may assume that v points out of H_1. Let u be the vector field along $f(\gamma)$ that points out of H_1 and is tangent to $\Sigma_{t_0-\epsilon}$. By assumption, near S_1 and S_2 v points towards the saddles or "into" the arches; hence H_1 is above both arches and u points into the arches as well. Thus u points towards both saddles, showing that $f(\gamma)$ is untwisted. Thus γ is twisted implies that $f(\gamma)$ is twisted and similarly we see that $f(\gamma)$ is twisted implies that γ is twisted, as required.

After crossing S_1 γ and $f(\gamma)$ form a single curve (say β) and after crossing S_2 this curve breaks up into two involute curves, say α and $f(\alpha)$. As we approach t_0 from above the boundary of $D_{\mathbf{Y}}$ limits on both saddles. Thus the boundary of $D_{\mathbf{Y}}$ is α or $f(\alpha)$.

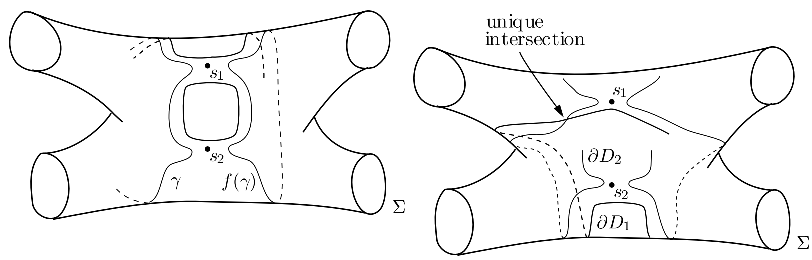

Figure 1: The reductions: untwisted (left) and twisted (right).

Assume first both γ and $f(\gamma)$ are untwisted and let v (resp. u) be a vector field along γ (resp. $f(\gamma)$) pointing towards both saddles. Pushing γ (resp. $f(\gamma)$) slightly along v (resp. u) we obtain a curve γ' (resp. γ''). Performing both saddle crossings on γ' and γ'' we get two curves that are isotopic to α and $f(\alpha)$ and are disjoint from γ and $f(\gamma)$. See figure 1. Although the disks $D_{\mathbf{P}}$ and $D_{\mathbf{Y}}$ may intersect the Heegaard surface in their interior, by definition of uppercase labels any such curve of intersection is trivial in the Heegaard surface and (as noted in [18]) the boundary of $D_{\mathbf{P}}$ bounds a purple meridian disk and the boundary of $D_{\mathbf{Y}}$ bounds a yellow meridian disk. We conclude that Σ_{t_0} weakly reduces, contradicting our assumption.

Next assume that both γ and $f(\gamma)$ are twisted. Let v and u be vector fields pointing towards S_1 and away from S_2. Again, we push γ along v and $f(\gamma)$ along u obtaining γ' and γ''. Performing the saddle operation S_1 on γ' and γ'', we obtain the curve β and see that β is disjoint from γ and $f(\gamma)$. Performing the saddle operation S_2 on β we obtain two curves (say α' and α''). Since γ and $f(\gamma)$ separate β from S_2, the curves obtained are not disjoint from γ. However, it is easy to see directly that $|\alpha' \cap \gamma| = 1$ and $|\alpha'' \cap \gamma| = 1$. As above, γ bounds a purple meridian disk and either α' or α'' is isotopic to $\partial D_{\mathbf{Y}}$ and hence bounds a yellow meridian disk. We conclude that Σ_t destabilizes, contradicting our assumptions. □

In Lemma 6.4 we study mixed labels:

Lemma 6.4. *There does not exit a critical level t_0 separating a layer labeled* $\mathbf{y}$ *from a layer labeled* $\mathbf{P}$ (*and similarly for* $\mathbf{Y}$ *and* $\mathbf{p}$).

Proof. Assume for contradiction t_0 is a critical level separating a layer labeled $\mathbf{P}$ (say l_p) from a layer labeled $\mathbf{y}$ (say l_y). For convenience we assume l_p is below l_y (the other case is treated by taking ϵ below to be a small negative number). For $t \in l_p$ there exists a compressing disk (say $D_{\mathbf{P}}$) that gives rise to the label $\mathbf{P}$. As in Lemma 6.3 we may assume that t_0 corresponds to two involute saddles, say S_1 and S_2. We consider two cases:

Case 1: One of the saddles does not destroy $D_{\mathbf{P}}$. This is identical to Lemma 6.3(1). We may assume from now on this is not the case.

Case 2: Both S_1 and S_2 destroys $D_{\mathbf{P}}$. Let $\epsilon > 0$ be sufficiently small. For $t_0 - \epsilon$, let $\gamma = \partial D_{\mathbf{P}}$. As ϵ approaches zero the curve γ limits on both saddles, or we would be in case 1. Since the saddles are involutes of each other, $f(\gamma)$ must limit on both saddles as well. There are two subcases:

Subcase 2a: $\gamma = f(\gamma)$. This is identical to Lemma 6.3(2a). We may assume from now on this is not the case.

Subcase 2b: $\gamma \neq f(\gamma)$. On $\Sigma_{t_0+\epsilon}$ all curves of intersection are inessential. Denote the region containing l_y by R_y (similarly R_p) and the region we get to after crossing S_1 out of R_y by R. Since R shares an edge R_y and another edge with R_p, by Proposition 6.1 R is unlabeled. Fix $(s,t) \in R$. If $F_1(s,t) \cap F_2(s,t)$ consists entirely of inessential simple closed curves then by Lemma 6.1 R has a lowercase label, contradiction. Suppose that $F_1(s,t) \cap F_2(s,t)$ contains curves that are essential in $F_1(s,t)$ (the other case is symmetric). Thus we see that crossing S_1 a single inessential curve of intersection becomes two simple closed curves that are both essential in $F_1(s,t)$ say α' and α''. Note that α' is parallel to α'' in $F_1(s,t)$ and all other curves of $F_1(s,t) \cap F_2(s,t)$ are inessential in $F_1(s,t)$.

By symmetry of α' and α'', there are four possibilities when crossing S_2 out of R into R_p:

(1) S_2 does not involve α' or α''.

(2) S_2 connects α' to a simple closed curve that is inessential in $F_1(s,t)$.

(3) S_2 connects α' to itself.

(4) S_2 connects α' to α''

We conclude the proof of Lemma 6.4 by reducing (1)–(4) above to previous subcases: in (1), no curve of $\Sigma_{t_0-\epsilon} \cap f(\Sigma_{t_0-\epsilon})$ involves both saddles hence this is in fact Case 1 above, contradiction. In (2), let β be the curve obtained from α' and an inessential simple closed curve after crossing S_2. Then β is the unique curve of $\Sigma_{t_0-\epsilon} \cap f(\Sigma_{t_0-\epsilon})$ involving both saddles. In (3) the curve α' splits into two curves and exactly one of the two involves both saddles. In (4) the curves α' and α'' become a single curve. Thus, in (2)–(4) there is a unique curve of $\Sigma_{t_0-\epsilon} \cap f(\Sigma_{t_0-\epsilon})$ that involves both saddles; since both γ and

$f(\gamma)$ involve both saddles, we conclude that $\gamma = f(\gamma)$ and we are in fact in Subcase 2a. With this contradiction we conclude the proof of Lemma 6.4. $\square$

Combining Proposition 6.1 and Lemmas 6.2, 6.3 and 6.4 we get:

Proposition 6.2. *A layer labeled* **y** *or* **Y** *cannot be adjacent to a layer labeled* **p** *or* **P**.

Next we prove (cf. [18, Corollary 6.2]):

Proposition 6.3. *Let R be a region of the Graphic and $(s,t) \in R$. Let R' denote the image of R under $(s,t) \to (t,s)$. Then the intersection of $F_1(s,t)$ and $F_2(s,t)$ is compression free and contains an essential curve if and only if R and R' are both unlabeled.*

Similarly, let l be a layer and $t \in l$. Then the intersection of Σ_t with $f(\Sigma_t)$ is compression free yet contains an essential curve if and only if l is unlabeled.

Proof. Suppose R and R' are both unlabeled. By Lemma 6.1 the absence of lowercase label in R implies that $F_1(s,t) \cap F_2(s,t)$ contains a curve that is essential in $F_1(s,t)$ or $F_2(s,t)$ (or both). Let γ be such a curve. Assume (for contradiction) that γ is essential in $F_1(s,t)$ but not in $F_2(s,t)$ and denote by $D \subset F_2(s,t)$ the disk γ bounds in $F_2(s,t)$. Consider all curves of $D \cap \Sigma_t$ that are essential in $F_1(s,t)$. An innermost such curve shows that R has an uppercase label, contradiction. Next assume (for contradiction) that γ is essential in $F_2(s,t)$ but not in $F_1(s,t)$. Then $f(\gamma)$ is essential in $F_1(t,s)$ but not in $F_2(t,s)$ implying an upper case label for R', again contradicting our assumption. Hence the absence of uppercase labels in R and R' implies that every curve of $F_1(s,t) \cap F_2(s,t)$ is essential or inessential in both surfaces, i.e., the intersection is compression free.

The converse is similar and we outline it here. Suppose R or R' is labeled. If R has a lowercase label then for $(s,t) \in R$, $F_1(s,t) \cap F_2(s,t)$ contains only inessential curves by Lemma 6.1. If R' has a lowercase label, then by the same lemma, for $(s,t) \in R'$, $F_1(s,t) \cap F_2(s,t)$ contains only inessential curves; since $F_1(t,s) = f(F_2(s,t))$ and $F_2(t,s) = f(F_1(s,t))$, applying the involution we see that for $(s,t) \in R$, $F_1(s,t) \cap F_2(s,t)$ contains only inessential curves as well. If R has an uppercase label, then by Definition 6.2 for $(s,t) \in R$ some curve of $F_1(s,t) \cap F_2(s,t)$ is essential in $F_1(s,t)$ and inessential in $F_2(s,t)$. Finally, if R' has an uppercase label than for $(s,t) \in R'$, some curve of $F_1(s,t) \cap F_2(s,t)$ is essential in $F_1(s,t)$ and inessential in $F_2(s,t)$. Applying the involution we see that for $(s,t) \in R$ some curve of $F_1(s,t) \cap F_2(s,t)$ is essential in $F_2(s,t)$ and inessential in $F_1(s,t)$.

The claim for layers follows from the claim for regions since every layer l is contained in a region R for which $R = R'$ and has the same labels as R. $\square$

Let $l_p = (a', a)$ (for some $a, a' \in [-\infty, \infty]$) be the highest layer labeled $\mathbf{p}$ or $\mathbf{P}$. Since for $t << 0$ the label is $\mathbf{p}$ the layer l_p exists and since for $t >> 0$ the label is $\mathbf{y}$ the layer l_p is not the topmost layer; hence $a \in \mathbb{R}$. Let $l_y = (b, b')$ (for some $b, b' \in [-\infty, \infty]$) be the first layer past l_p labeled $\mathbf{y}$ or $\mathbf{Y}$. Since the topmost layer is labeled $\mathbf{y}$ the layer l_y exists. By Proposition 6.2 the layers l_p and l_y cannot be adjacent; hence $a < b$. By choice of l_p, the layers between l_p and l_y are not labeled $\mathbf{p}$ or $\mathbf{P}$, and by choice of l_y they are not labeled $\mathbf{y}$ or $\mathbf{Y}$. Hence all the layers in (a, b) are unlabeled and by Proposition 6.3 the corresponding surfaces have compression free intersection, yet their intersection has an essential curve; this completes the proof of Theorem 6.1(1).

Let t_0 be a point $a < t_0 < b$ and suppose there is a region R of the Graphic adjacent to (t_0, t_0) corresponding to an intersection which is either not compression free, or consists entirely of inessential simple closed curves. Since every regular point $a < t < b$ is unlabeled, (t_0, t_0) is not in the interior of R; hence (t_0, t_0) is a vertex of R. Let R' be the image of R under $(s, t) \to (t, s)$ (note that (t_0, t_0) is a vertex of R' as well). By Proposition 6.3 either R or R' is labeled. If the label at R or R' is $\mathbf{p}$ or $\mathbf{P}$ (resp. $\mathbf{y}$ or $\mathbf{Y}$) we shorten the interval (a, b) by replacing a by t_0 (resp. replacing b by t_0). Repeating this process if necessary, we may assume all the regions near every point of (a, b) are unlabeled, some region near a is labeled $\mathbf{p}$ or $\mathbf{P}$, and some region near b is labeled $\mathbf{y}$ or $\mathbf{Y}$.

This completes the proof of Theorem 6.1. $\square$

Note that the proof gave us a little more than we bargained for: we have control over the labels appearing near a and b. We need to improve the intersection from "compression free" to "essential." This is achieved in the next section; for the remainder of this section we follow a technique of [18] to control inessential curves appearing in (a, b). Fix $\epsilon > 0$ small enough so that distance between any two critical points is greater than 2ϵ. Denote the critical points in $(a - \epsilon, b + \epsilon)$ by $a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b$. Consider the following embedding of the interval $(a - \epsilon, b + \epsilon)$ in the parameter square, denoted δ: send $t \in [a, b]$ to (t, t), send $(a - \epsilon, a]$ into a region labeled $\mathbf{p}$ or $\mathbf{P}$ that is adjacent to (a, a) and send $[b, b + \epsilon)$ into a region labeled $\mathbf{y}$ or $\mathbf{Y}$. More specifically, let R_1 be the region containing (a, t_1). If a is a single critical point we embed $(a - \epsilon, a]$ in the diagonal. If a is a double critical point and there is a region labeled $\mathbf{p}$ or $\mathbf{P}$ that is adjacent to R_1 we embed $(a - \epsilon, a]$ in that region. Note that in that case moving from $(a - \epsilon, a)$ to (a, t_1) we cross only one of the critical points at a while tangent to the other, which we may ignore. Finally, if a is a double critical point and both regions adjacent to R_1 are unlabeled we embed $(a-\epsilon, a]$ in the diagonal. $(b, b+\epsilon)$ is treated similarly.

δ gives an isotopy of two surfaces denoted $F_1(t)$ and $F_2(t)$ ($(a-\epsilon < t < b+\epsilon)$). We label points of $(a-\epsilon, b+\epsilon)$ according to the intersection of $F_1(t)$ and $F_2(t)$, as in Definition 6.2.

Proposition 6.4. *With the hypothesis of Theorem* 6.1 *there are families of surfaces $F_1(t)$, $F_2(t)$ (for $t \in (a-\epsilon, b+\epsilon)$) with the following properties:* (1) *$F_1(t)$ is isotopic to Σ_t.* (2) *For $t \in [a, b]$, $f(F_1(t)) = F_2(t)$.* (3) *(a, b) has a neighborhood N in the parameter square so that every regular point in N is unlabeled.* (4) *Arbitrarily close to a (resp. b) there is a region labeled* **p** *or* **P** *(resp.* **y** *or* **Y**).

Furthermore, we may assume that for any regular point $t \in (a-\epsilon, b+\epsilon)$ the intersection of $F_1(t)$ with $F_2(t)$ contains at most one invariant inessential simple closed curve or a pair of involute inessential simple closed curves. When it does, the layer containing t is bounded by a single or double saddle on one side and a single or double center on the other side, and the intersection in the adjacent layers contains no inessential curves.

Proof. The proof is based on [18]; we need to verify that it works in the invariant setting. We induct on the number of critical points in $(a-\epsilon, b+\epsilon)$ that involve an inessential simple closed curve. Below, we modify $F_1(t)$ and $F_2(t)$ by removing inessential simple closed curves via disk swaps or introducing inessential simple closed curves via fixed or involute centers. By definition an inessential simple closed curve bounds a disk on both F_1 and F_2 and therefore its image is an inessential simple closed curve as well; thus we may perform the disk swap invariantly. It is easy to see that this does not change labels. We note that we never change compressions or essential simple closed curves of $F_1 \cap F_2$, hence there is a natural bijection between these curves before and after the modification and we may talk of "the same curves" and "the same saddles".

We begin with $(a-\epsilon, a+\epsilon)$. Assume first that the label at $a-\epsilon$ is **p**. By isotopy (equivariant if $(a-\epsilon, a)$ was mapped to the diagonal) we remove all inessential curves of $F_1(t) \cap F_2(t)$ for $t \in (a-\epsilon, a-\frac{1}{2}\epsilon)$. At a a single invariant inessential simple closed curve (resp. two involute inessential simple closed curves) get pinched to form two or three essential curves (resp. two essential curves). We introduce a new critical point at $a-\frac{1}{2}\epsilon$ as follows: if at a two involute inessential curves get pinched, the critical point at $a-\frac{1}{2}\epsilon$ corresponds to involute centers where the necessary pair of curves appear. If at a a single invariant inessential curve gets pinched we create this curve using an invariant center at $a - \frac{1}{2}\epsilon$. Note that this can be done: the invariant inessential curve bounds two disks from F_1 and F_2 that (using an innermost disk argument) we may assume are disjoint. These disks bound an invariant ball that by the Brouwer Fixed Point Theorem contains a fixed point of f. Isotoping $F_1(t)$ and $F_2(t)$ to that fixed point we create an invariant center (we will often use this construction in the proof of this proposition). This defines the isotopy for $t \in (a - \epsilon, a + \epsilon)$. If the label at $(b, b + \epsilon)$ is $\mathbf{y}$, we modify $(b - \epsilon, b + \epsilon)$ similarly. Note that $(a - \epsilon, a + \epsilon)$ and $(b - \epsilon, b + \epsilon)$ fulfill the requirements of Proposition 6.4.

Next assume that the label at $a - \epsilon$ is $\mathbf{P}$. In that case, after removing all inessential simple closed curves from the intersection at $(a - \epsilon, a)$ the curves giving rise to compressions still exist. The critical point at a is a (single or double) saddle that destroys these curves. In the case of a single saddle, this saddle cannot join an inessential simple closed curve to a compression or the labels would not change; therefore the (one or two) curves involved in a were not removed and we may cross a. In the case of a double saddle $(a - \epsilon, a + \epsilon)$ is embedded in the diagonal. The two saddles involve one, two or three distinct curves. If only one curve is involved it is the compression which was not removed and we may cross a. If two curves are involved, at least one is a compression and the other is the image of the compression (note that a compression is never invariant) and hence is also a compression; again we may cross a. If three curves are involved, one is the compression. Crossing only one of the two saddles we arrive at a region adjacent to R_a, which by construction of the embedding of $(a - \epsilon, b + \epsilon)$ we know is unlabeled. Hence the curve attached to the compression is not inessential, and similarly the third curve is not inessential. In this case too we may cross a. In all cases, we constructed the family $F_1(t), F_2(t)$ for $t \in (a - \epsilon, a + \frac{1}{2}\epsilon)$. If no inessential curves appeared at a we extend this family to $(a - \epsilon, a + \epsilon)$ without a change; otherwise, there is either a single invariant inessential curve or a pair of involute inessential curves. We remove them using centers at $a + \frac{1}{2}\epsilon$, symmetrically to the construction of inessential curves described in the previous paragraph. This describes the modification at $(a - \epsilon, a + \epsilon)$ in this case. If the label at $(b, b + \epsilon)$ is $\mathbf{Y}$, we modify $(b - \epsilon, b + \epsilon)$ similarly. Note that $(a - \epsilon, a + \epsilon)$ and $(b - \epsilon, b + \epsilon)$ fulfill the requirements of Proposition 6.4. This concludes the base case of the induction.

By the inductive hypotheses, we suppose the isotopy in $(a - \epsilon, t_i + \epsilon)$ (for some critical $t_i \leq b$) fulfills the requirements of Proposition 6.4; moreover, by construction $F_1(t_i+\epsilon)\cap F_2(t_i+\epsilon)$ contains no inessential curves. The possibilities when passing from $t_0-\epsilon$ to $t_0+\epsilon$ are (note that $t_i\in(a,b]$ hence every curve at $a<t<t_i$ is either essential or inessential but not a compression):

(1) A single center (resp. double center) in which a one (resp. two) inessential simple closed curve is created or destroyed.

(2) A single saddle (resp. double saddle) in which one (resp. two) inessential simple closed curve is attached to split off from another curve (which may or may not be inessential).

(3) A saddle or double saddle in which one inessential simple closed curve becomes essential curves or compressions.

(4) A double saddle in which two inessential simple closed curves become two essential curves or compressions.

(5) All curves involved in t_0 are essential.

In (1) and (2) the critical points are unnecessary, as they do not change the pattern of essential curves. Therefore we may ignore these critical points and continue past t_i. As above, in (3) we create the invariant simple closed curve at $t_i-\frac{1}{2}\epsilon$ and in (4) we create the two involute simple closed curves at $t_i-\frac{1}{2}\epsilon$. In (5) there is nothing to do.

The surfaces $F_1(t_{i-1}+\epsilon)$, $F_2(t_{i-1}+\epsilon)$ and $F_1(t_i-\epsilon)$, $F_2(t_i-\epsilon)$ are obtained from the original surfaces by removing all inessential simple closed curves of intersection. Hence these surfaces are isotopic and we may extend the isotopy across $[t_{i-1}+\epsilon,t_i-\epsilon]$. Continuing in this way we finally arrive at an isotopy of $(a-\epsilon,t_{n-1}+\epsilon)$ that can be extended to $(b-\epsilon,b+\epsilon)$ across $[t_{n-1}+\epsilon,b-\epsilon]$, proving the proposition. $\square$

For $(s,t)\in(a-\epsilon,b+\epsilon)\times(a-\epsilon,b+\epsilon)$ we construct the parameter square by setting $F_1(s,t)=F_1(s)$ and $F_2(s,t)=F_2(t)$; the involution exchanges the surfaces along the diagonal only for $t\in[a,b]$. We perturb the parameter square fixing the diagonal to be generic as we did in Section 5. By construction δ is the diagonal. We perturb δ to obtain the *generic interval*: we move δ slightly off a and b to be transverse to the Graphic; similarly, near a double critical point (say t_i) we replace $(t_i-\epsilon,t_i+\epsilon)$ (for some tiny ϵ) by a small semicircle in $[a,b]\times[a,b]$ that avoids the double point (see figure 2). By construction the generic interval is parameterized by $t\in(a-\epsilon,b+\epsilon)$, starting at a region labeled $\mathbf{p}$ or $\mathbf{P}$, going through unlabeled regions to a region labeled $\mathbf{y}$ or $\mathbf{Y}$.

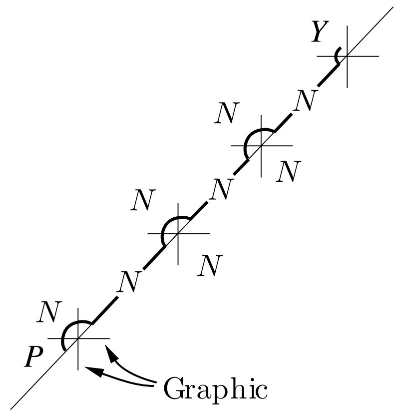

Figure 2: The Generic Interval.

7. Essential, Spinal Intersection

Definitions 7.1.

(1) Let S be a surface, and $K \subset S$ be an embedded graph. We say that K *contains a spine* of S if no component of S cut open along K contains a simple closed curve that is essential in S.

(2) Let F_1, $F_2 \subset M$ be embedded surfaces, and Δ_2 a set of compressing disks for F_2. Suppose $F_1 \cap F_2$, $F_1 \cap \Delta_2$ and $F_1 \cap \partial\Delta_2$ are all transverse. We say that F_1 intersects $F_2 \cup \Delta_2$ *spinally* if $F_1 \cap (F_2 \cup \Delta_2)$ contains a spine of F_1.

(3) Let F_1, $F_2 \subset M$ be embedded surfaces. We say that the intersection of F_1 and F_2 is *spinal* if there exists some set of compressing disks for F_2 fulfilling condition (2) above or disks for F_1 fulfilling the same condition with the indices exchanged; for convenience we always assume that compressing disks are for F_2.[4]

In this section we prove Theorem 7.1, which is a combination of two theorems (one for a manifold admitting two strongly irreducible Heegaard splittings and the other for a manifold admitting a strongly irreducible Heegaard splitting and an involution).

[4]We often use (3) without any mention of the specific set of compressing disks. We shall see that this makes sense, particularly in light of Lemma 8.1 which shows the global nature of essential, spinal intersection.

Recall (Remark 1.1) that if M is non-Haken then any minimal genus Heegaard surface is strongly irreducible; hence the theorem below is not vacuous:

Theorem 7.1. *Let M be an irreducible, orientable, a-toroidal, non-Seifert fibered manifold of Heegaard genus at least two. Suppose that either M admits two strongly irreducible Heegaard surfaces Σ_1 and Σ_2 or a strongly irreducible Heegaard surface Σ and an orientation preserving involution f. Then we have:*

(1) *Σ_1 and Σ_2 can be isotoped to intersect essentially and spinally.*

(2) *Σ can be isotoped so that Σ and $f(\Sigma)$ intersect essentially and spinally.*

Remark 7.1. In [18] Rubinstein and Scharlemann prove a result very close to (1) above: they show that Σ_1 and Σ_2 can be isotoped so that their intersection is compression free, spinal and contains at most one inessential simple closed curve. (If we remove the inessential curve of intersection we may lose spinality, so Theorem 7.1(1) does not follow.) However, their result is not quite strong enough for our purpose: in the next section we prove that if Σ_1 intersects Σ_2 essentially and spinally then M cut open along $\Sigma_1 \cup \Sigma_2$ consists of handlebodies. Existence of an inessential curve of intersection allows for "knotted handles" and hence M cut open along $\Sigma_1 \cup \Sigma_2$ may not consists of handlebodies; it is quite easy to construct such examples.

Proof of Theorem 7.1. The method for finding a point that corresponds to spinal intersection is given in [18, Proposition 6.5] where it is shown that given an interval transverse to the Graphic, starting in a region labeled **p** or **P** and ending in a region labeled **y** or **Y** (such as the generic interval constructed in the previous section) there exists a set of compressing disks for one of the two surfaces (say Δ_2 for F_2) so that no component of M cut open along $F_2 \cup \Delta_2$ is adjacent to itself,[5] and for some regular point t in that interval the intersection of $F_1(t)$ with $F_2(t) \cup \Delta_2(t)$ contains a spine of $F_1(t)$. In this section, we show that this point can be found on the diagonal and that the surfaces corresponding to this point may be assumed to intersect essentially. Note that since the generic interval gives an ambient isotopy of F_2, it provides an isotopy for $\Delta_2 = \Delta_2(t)$ as well ((t) is suppressed throughout this section).

[5]We need this property for quoting claims from [18] but we will not refer to it directly.

We note that proving (1) of the theorem requires finding a point that corresponds to spinal intersection in a layer that corresponds to essential intersection, while (2) requires in addition that this point is on the diagonal. We will concentrate on (2) in this section and (1) will follow from the argument here and the isotopy constructed in [18] that has all the properties of the generic interval. From here on, we will not refer to (1) directly.

Definition 7.1. Let F_1, F_2 and Δ_2 be as above. A point on the generic interval is called *regular* if the intersections $F_1 \cap F_2$, $F_1 \cap \partial\Delta_2$, and $F_1 \cap \Delta_2$ are all transverse, *critical* otherwise.

After a small perturbation of Δ_2 (if necessary) we may assume there are only finitely many critical points. The intervals obtained by cutting the generic interval open along the critical points are called *sublayers*.

The following lemma provides conditions to preserve spinality near saddles. A saddle move is similar to a boundary compression, and crossing a saddle is equivalent to isotoping one of the surfaces across a disk (say δ) so that $\partial\delta = (\delta \cap F_1) \cup (\delta \cap F_2)$, where $\delta \cap F_1$ and $\delta \cap F_2$ are two arcs meeting at their endpoints. We say that δ *defines* the saddle. While the interior of δ is disjoint from F_1 and F_2, it may intersect Δ_2. An arc of $\Delta_2 \cap \delta$ has two endpoints, either both on F_1, or one on F_1 and one on F_2, or both on F_2. We say that these arcs are of *type 1–1, 1–2, 2–2* (respectively). See figure 3.

Lemma 7.1. *Let R_1 and R_2 be adjacent regions in the Graphic, and suppose the critical point separating R_1 from R_2 is a saddle. Suppose further that $\Delta_2 \cap \delta$ contains no type 1–1 arcs.*

Then the intersection in R_1 is spinal if and only if the intersection in R_2 is.

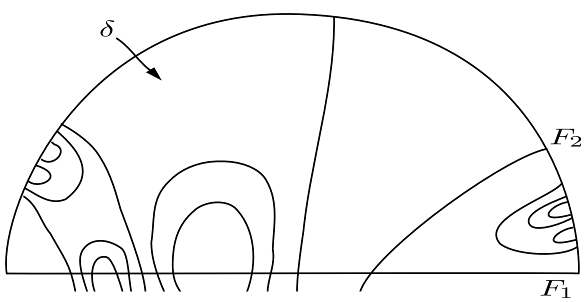

Figure 3: Arcs of $\Delta_2 \cap \delta$ (on δ).

Proof of Lemma 7.1. Assume the intersection is spinal in, say, R_1; we will show it is spinal in R_2 as well. Let F_1, F_2 and Δ_2 be as above, intersecting essentially.

First we show that after isotoping Δ_2 if necessary we may assume that $\Delta_2 \cap \delta$ contains no simple closed curves: let $\gamma \subset \Delta_2 \cap \delta$ be a simple closed curve, chosen to be innermost in δ. By isotopy of Δ_2, we can replace the disk γ bounds in Δ_2 by the disk it bounds in δ, and by small perturbation push this disk off δ. Since γ was chosen innermost in δ, Δ_2 remains embedded. It is easy to see that $|\Delta_2 \cap \delta|$ was reduced by at least one; we need to show that the intersection is still spinal. The only change to $F_1 \cap (F_2 \cup \Delta_2)$ is removing simple closed curves of $F_1 \cap \operatorname{int} \Delta_2$. None of these curves connects to any other component of $F_1 \cap (F_2 \cup \Delta_2)$. (We call such components *isolated simple closed curves.*) If an isolated simple closed curve were essential in F_1, then a parallel copy of it would contradict spinality; hence isolated simple closed curves are inessential in F_1 and removing them from $F_1 \cap (F_2 \cup \Delta_2)$ does not change spinality.

Suppose that $\delta \cap \Delta_2 \neq \emptyset$. Consider a component (say T) of δ cut open along Δ_2 that contains a point of $(\delta \cap F_1) \cap (\delta \cap F_2)$. First assume T contains no 2–2 arcs; recall that by assumption there are no 1–1 arcs. Therefore T is a triangle with a single 1–2 arc on its boundary (see the leftmost triangle in figure 3). We use T to guide an isotopy of Δ_2 that removes the given 1–2 arc from $\delta \cap \Delta_2$, see figure 4. Let α be a simple closed curve on F_1 disjoint from $F_2 \cup \Delta_2$ after the isotopy. It is easy to see that α is homotopic on F_1 to a curve disjoint from $F_2 \cup \Delta_2$ before the isotopy (figure 4 shows $F_1 \cap (F_2 \cup \Delta_2)$ "before and after"; α is not shown). Since we assumed the intersection to be spinal before the isotopy, α must be inessential on F_1. Hence the intersection is spinal after the isotopy. This reduces $|\delta \cap \Delta_2|$.

Next suppose that a component of δ cut open along Δ_2 that contains a point of $(\delta \cap F_1) \cap (\delta \cap F_2)$ does contain 2–2 arcs; we use δ to guide an

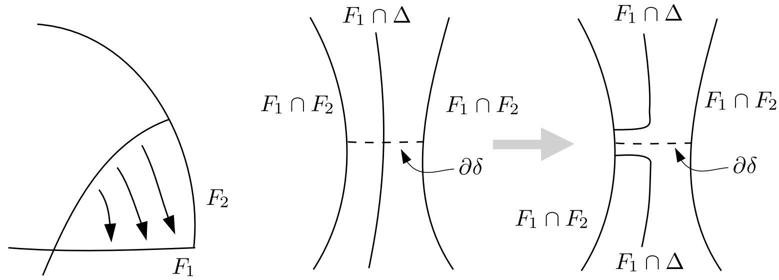

Figure 4: Removing arcs of type 1–2.

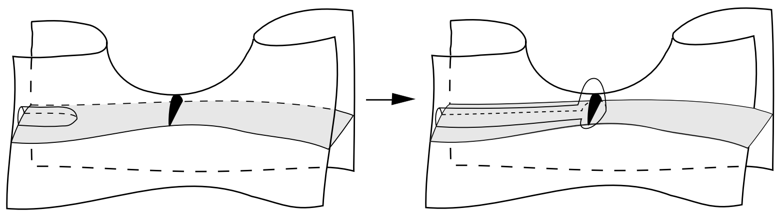

Figure 5: Making the intersection spinal after the saddle.

isotopy of Δ_2 sliding these arcs off δ. As a result of this isotopy, for every 2–2 arc removed a pair of arcs are added to $F_1 \cap (F_2 \cup \Delta_2)$ but nothing is removed, hence the intersection is still spinal. This too reduces $|\delta \cap \Delta_2|$.

Since $|\delta \cap \Delta_2|$ is being reduced this process must terminate; when it does, $\delta \cap \Delta_2 = \emptyset$. We now cross the saddle. After crossing the saddle the pattern of intersection between F_1 and $F_2 \cup \Delta_2$ changes only near the saddle point where two parallel arcs (say horizontal) are replaced by two vertical arcs, denoted v_1 and v_2. Clearly if we add an arc connecting v_1 to v_2 to $F_1 \cap (F_2 \cup \Delta_2)$ the intersection will become spinal. We obtain this by the move shown in figure 5. It is described as follows: after crossing the saddle we obtain a disk similar to δ defining the same saddle from the opposite side; denote this disk δ^{-1}. Then $\delta^{-1} \cap F_1$ is an arc connecting v_1 to v_2. Let V be the component of M cut open along F_2 containing δ^{-1}. Denote the frontier of a neighborhood of δ^{-1} in V by D.[6] $D \cap F_1$ consists of exactly two arcs parallel to $\delta^{-1} \cap F_1$. Let α be an arc on F_2 connecting $\partial\Delta_2$ to ∂D, missing $\partial\Delta_2$ in its interior (i.e., an outermost arc; α may intersect F_1 in its interior). Band-connect sum a disk of Δ_2 to D along α; this changes Δ_2 by an isotopy. After the band sum, $F_1 \cap (F_2 \cup \Delta_2)$ consists of the intersection prior to the band sum union an arc for every point of $\alpha \cap F_1$ union $D \cap F_1$. It is now clear that $F_1 \cap (F_2 \cup \Delta_2)$ contains a spine of F_1, proving Lemma 7.1. $\square$

It follows immediately from Definition 6.2 (labels) that at a region labeled **p** or **P** there is a meridian disk for $F_1(t)$ that is purple near its boundary (i.e., "below" $F_1(t)$) and intersects $F_1(t)$ in curves that are all inessential in $F_1(t)$ (cf. [18, Section 8]). Similarly, in a layer labeled **y** or **Y** there exists a meridian disk for $F_1(t)$ that is yellow near its boundary (i.e., "above" $F_1(t)$). This motivates the following labeling scheme for sublayers:

[6] In other words, D is the disk obtained from two parallel copies of δ^{-1} connected together on the other side of F_1 so that $\partial D \subset F_2$. Note that D is a boundary parallel disk in V.

Definition 7.2. The label *"below"* is used in a sublayer where there is a compressing disk for F_1 that is disjoint from $F_2 \cup \Delta_2$, is below F_1, and intersects $F_1(t)$ (if at all) only in curves that are inessential in $F_1(t)$. The label *"above"* is defined similarly.

Strong irreducibility implies that each sublayer has at most one label and adjacent sublayers cannot be labeled *"below"* and *"above."* Since the generic interval starts with a layer labeled **p** or **P** (and hence with a sublayer labeled *"below"*) and ends with a layer labeled **y** or **Y** (and hence with a sublayer labeled *"above"*) some sublayer is unlabeled. From [18, Section 8] we have:

Proposition 7.1 [18]. *Let t be a regular point. $F_1(t) \cap (F_2(t) \cup \Delta_2(t))$ contains a spine of $F_1(t)$ if and only if t is in an unlabeled sublayer.*

Thus, the generic interval described in Proposition 6.4 contains a point t that corresponds to spinal intersection. However, t may not have all the properties required by Theorem 7.1, specifically:

(1) t may be off the diagonal and is separated from it by centers.

(2) t may be off the diagonal and is separated from it by saddles.

(3) t may be in a sublayer where one or two inessential simple closed curves of intersection exist.

Everything we said until this point is true for *any* isotopy of $\Delta_2(t)$. We exploit this flexibility and design an isotopy of $\Delta_2(t)$ that helps us deal with the three problems listed above, then combine the three cases to prove the theorem.

Case 1: The unlabeled sublayer is separated from the diagonal by centers, denoted b and c (say $b < c$). We may assume that crossing b from left to right an inessential curve appears (otherwise, we reverse t). Then by construction crossing c from left to right another inessential curve appears. Let $a < b$ and $d > c$ be points of the generic interval on the diagonal and $\epsilon > 0$ small enough so that the only critical points of $F_1(t) \cap F_2(t)$ in $(a - \epsilon, d + \epsilon)$ are b and c.

Fix $t_0 \in (b, c)$. Then there is an arc α_c so that one endpoint of α_c is $\alpha_c \cap F_1(t_0)$, the other is $\alpha_c \cap F_2(t_0)$, and crossing c is equivalent to isotoping $F_1(t_0)$ along α_c and pushing a small disk of $F_1(t_0)$ across $F_2(t_0)$. We say that α_c *defines the center* c. We change the isotopy of $\Delta_2(t)$ as follows: we reparameterize $\Delta_2(t)$ in $(a - \epsilon, b)$ so that all the critical points of $F_1(t) \cap \Delta_2$ and $F_1(t) \cap \partial\Delta_2$ are in $(a - \epsilon, a)$. In $[b, t_0]$ $F_1 \cap \Delta_2$ and $F_1 \cap \partial\Delta_2$ have no critical points. In (t_0, c) we slide Δ_2 off α_c. Thus $F_1 \cap \Delta_2$ has exactly $|\alpha_c \cap \Delta_2|$ critical points, each introducing an isolated simple closed curve to $F_1(t) \cap (F_2(t) \cup \Delta_2(t))$. $F_1 \cap \partial\Delta_2$ has no critical point in (t_0, c). In $[c, d + \frac{1}{2}\epsilon]$ there are no critical points of $F_1(t) \cap \Delta_2(t)$ or $F_1(t) \cap \partial\Delta_2$. In $(d + \frac{1}{2}\epsilon, d + \epsilon)$ we isotope Δ_2 to its original configuration. For a regular value $t \in (a, d)$ the difference between $F_1(t) \cap (F_2(t) \cup \Delta_2(t))$ and $F_1(a) \cap (F_2(a) \cup \Delta_2(a))$ is isolated curves; hence the intersection is spinal in t if and only if it is spinal in a, and we may assume case (1) does not happen. Moreover, we have control over the labels of sublayer: either all sublayers of (a, d) are unlabeled (if the intersection is spinal) or all are labeled, and since adjacent labeled sublayers have the same label we conclude that *either both a and d are both unlabeled or both are labeled and the labels at a and d are the same.*

Case 2: The unlabeled sublayer is separated from the diagonal by saddles. Similar to Case (1) denote the saddles $b < c$ and let $a < b$ and $d > c$ be points of the generic interval on the diagonal, and $\epsilon > 0$ small enough so that the only critical point of $F_1(t) \cap F_2(t)$ in $(a - \epsilon, d + \epsilon)$ are b and c. At a there exist two disks δ_b, δ_c defining the saddle b, c, respectively. (Recall the construction of δ in the paragraph preceeding Lemma 7.1.) Since moving along the diagonal both saddles are crossed simultaneously $\partial\delta_b \cap \partial\delta_c = \emptyset$, and applying a standard innermost disk argument we may assume that $\delta_b \cap \delta_c = \emptyset$. Similar to the proof of Lemma 7.1 we use δ_b and δ_c to guide an isotopy of Δ_2 off δ_b and δ_c. (This changes the sublayers, and since $\Delta_1 \cap \delta_b$ and $\Delta_2 \cap \delta_c$ may have 1–1 arcs we cannot assume the labels do not change.) Isotope Δ_1 in (a, d) so that $F_1 \cap \Delta_2$ and $F_1 \cap \partial\Delta_2$ have no critical points in $[a, d]$. (Thus (a, b) and (c, d) are contained in one sublayer each, and (b, c) is a sublayer.) In $(d, d + \epsilon)$ isotope Δ_2 to its original configuration. By Lemma 7.1 if the sublayer (b, c) is unlabeled so is the sublayer containing (c, d). Hence, if a regular value $t \in (b, c)$ corresponds to spinal intersection so does the regular value d on the diagonal. As in case (1) we have a little more: Lemma 7.1 implies that the sublayer containing (a, b) is unlabeled if and only if the sublayer (b, c) is. As adjacent sublayers have the same labels we again conclude that *either both a and d are unlabeled or both are labeled and the labels at a and d are the same.*

Case 3: The unlabeled sublayer corresponds to an intersection that contains one or two inessential simple closed curves. Let l be a layer containing inessential curves. By Proposition 6.4 l is bounded on one side by a (single or double) center and the other side by a (single or double) saddle. Say the center is at c and the saddle at s. For convenience we assume $c < s$ (the other case is similar), so $l = (c, s)$. Let $\epsilon > 0$ be small enough so that c

and s are the only critical points on the generic interval in $(c-\epsilon, s+\epsilon)$. For convenience, we assume the semicircles of the generic interval have radius $\frac{1}{3}\epsilon$. For $t \in (c, s)$ there are one or two δ disks that define the saddles; if there are two δ disks we may assume (as in case (2)) that they are disjoint. In $(c-\epsilon, c)$ there are one or two α arcs that define the centers. By the construction in Proposition 6.4 the saddles at s involve the inessential curves of (c, s); hence the δ disks cannot be seen in $(c-\epsilon, c)$. However, for $t \in (c-\epsilon, c-\frac{1}{3}\epsilon)$ we can find the trace of the δ disks as disks are disjointly embedded in their interior but not on their boundary. The boundary of each disk consists of four arcs, one on $F_1(t)$, one on $F_2(t)$, and between them two arcs on the α arcs. There are three cases, but their treatment is identical: in the case of a single center and a single saddle we see a single δ disk attached to itself along the single α arc to form an annulus $\mathcal{A}$. In case of a single center and two saddles we see two δ disks attached to each other along a single α arc, each disk forming an embedded annulus (say A_1, A_2), with $A_1 \cap A_2 = \alpha$. We take $\mathcal{A}$ to be $A_1 \cup A_2$. The case of a double center and a single saddle is impossible since there would still be inessential curves in $(s, s+\epsilon)$, contradicting Proposition 6.4. Finally, in case of double center and double saddle, the two δ disks are glued to each other along the two α arcs forming a single annulus $\mathcal{A}$. In all three cases we see an annulus-like complex $\mathcal{A}$ which is homeomorphic to either a circle cross an interval or a wedge of two circles cross an interval, and the arcs α are contained in $\mathcal{A}$ and have the form one or two points cross interval.

We are now ready to describe the isotopy of Δ_2: in $(c-\epsilon, c-\frac{2}{3}\epsilon)$ slide Δ_2 off the α arcs. As before this creates isolated simple closed curves. At $c-\frac{2}{3}\epsilon$ the arcs of $\Delta_2 \cap \mathcal{A}$ come in three flavors, arcs of types 1–1, 1–2, and 2–2, where an arc is of type $i-j$ if it has one boundary component on F_i and the other on F_j. In $(c-\frac{2}{3}\epsilon, c-\frac{1}{3}\epsilon)$ we use $\mathcal{A}$ to guide an isotopy of Δ_2 that removes all 1–1 arcs (the so-called karate-chop). After crossing the centers, the δ disks contain no 1–1 arcs. We isotope Δ_2 in $(c+\frac{1}{3}\epsilon, s-\frac{1}{3}\epsilon)$ to remove the 1–2 and 2–2 arcs. After crossing the saddles near s, we isotope Δ_2 in $(s+\frac{1}{3}\epsilon, s+\epsilon)$ to its original configuration. After this isotopy, if some layer in the semicircle $(c-\frac{1}{3}, c+\frac{1}{3}\epsilon)$ is unlabeled then so is the layer past $c+\frac{1}{3}\epsilon$, as addition of isolated curves at the centers cannot change spinality. If some label in $(c+\frac{1}{3}\epsilon, s+\frac{2}{3}\epsilon)$ is unlabeled then by Proposition 7.1 the region containing $s+\frac{1}{3}\epsilon$ is unlabeled. We conclude that *if some layer in* $(s-\epsilon, c+\epsilon)$ *is unlabeled than some layer in* $(c-\epsilon, c-\frac{1}{3}\epsilon)$ *or in* $(s+\frac{1}{3}, s+\epsilon)$ *is unlabeled.*

For proving Theorem 7.1(1): in [18] Rubinstein and Scharlemann give an isotopy of Σ_1 and Σ_2 with the properties listed in Proposition 6.4 (with no reference to invariance, of course). Theorem 7.1(1) follows from that and the argument in case (3) above.

We combine the three cases to prove Theorem 7.1(2): starting with the generic interval $(a - \epsilon, b + \epsilon)$, we isotope Δ_2 in a neighborhood of any layer that contains an inessential curve as described in case (3) above. Next, given a double critical point not on the boundary of an layer containing inessential curves (say t_0), we isotope Δ_2 near it as described in case (1) or (2) above. The generic interval starts at a sublayer labeled *"below"* and ends at a sublayer labeled *"above"* and is transverse to the Graphic; by Rubinstein and Scharlemann [18] some sublayer of the generic interval is unlabeled, and by cases (1)–(3) above there exists a point on the diagonal corresponding to essential, spinal intersection.

This completes the proof of Theorem 7.1. □

8. M cut open along $\Sigma \cup f(\Sigma)$

This section is devoted to the proof of Theorem 1.4. The proofs of cases (1) and (2) are identical. For simplicity we use the notation Σ_1 and Σ_2 in the proof, (2) follows by setting $\Sigma = \Sigma_1$ and $f(\Sigma) = \Sigma_2$. In Theorem 7.1 we established the existence of an isotopy of Σ_1 and Σ_2 so that the intersection of Σ_1 and Σ_2 is essential and spinal. Theorem 1.4 follows from that and the following lemma that originally appeared in [17]. For completeness we bring it here with its proof.

Lemma 8.1. *Let Σ_1 and Σ_2 be Heegaard surfaces intersecting spinally and essentially. Then the components of M cut open along $\Sigma_1 \cup \Sigma_2$ are handlebodies.*

Proof. Since the intersection is spinal there exists a complete set of compressing disks Δ_2 for one of the surfaces (say Σ_2) so that $\Sigma_1 \cap (\Sigma_2 \cup \Delta_2)$ contains a spine of Σ_1. By definition of spinal intersection, Σ_1 is incompressible in the complement of $\Sigma_2 \cup \Delta_2$. (Note that components of Σ_1 cut open along $\Sigma_2 \cup \Delta_2$ may compress, but any curve of Σ_1 cut open along $\Sigma_2 \cup \Delta_2$ that is compressed is inessential in Σ_1.)

We may assume that $\Sigma_1 \cap \Delta_2$ consists of arcs only: let γ be a simple closed curve in $\Sigma_1 \cap \Delta_2$. Since the intersection is spinal, γ bounds a disk in Σ_1. Passing to an innermost such, we see a disk whose interior intersects neither Δ_2 nor Σ_2 (by essentiality). We now use this disk to isotope Δ_2 and reduce $|\Delta_2 \cap \Sigma_1|$.

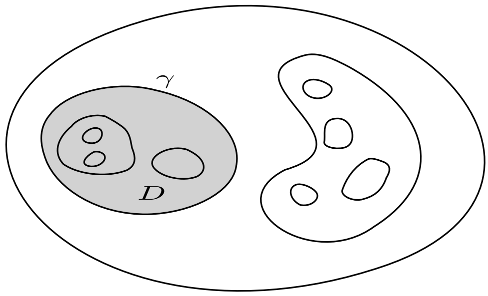

Figure 6: When c is not a disk.

Let B be some component of M cut open along $\Sigma_2 \cup \Delta_2$, and c some component of $\Sigma_1 \cap B$. We show that c is a disk. Assume for contradiction c is not a disk. Since the intersection is spinal, every curve on c is trivial in Σ_1. Hence c is a punctured disk. Let γ be one of the punctures, and $D \subset \Sigma_1$ the disk it bounds (see figure 6). By assumption $\partial D = \gamma \subset \partial B$, and $N_D(\partial(D)) \cap B = \gamma$ (that is, near its boundary D is outside B). Since the intersection of Σ_1 and Σ_2 is essential $\gamma \not\subset \Sigma_2$. Since $\Sigma_1 \cap \Delta_2$ consists of arcs, $\gamma \not\subset \Delta_2$. Hence γ must have parts on Σ_2 and parts on Δ_2 (say *above* Σ_2). Clearly part of D is below Σ_2. But the boundary of this part of D is a non-empty collection of simple closed curves in $\Sigma_1 \cap \Sigma_2$, all inessential in Σ_1, contradicting essentiality.

M cut open along $\Sigma_2 \cup \Delta_2$ consists of balls. Since the pieces of Σ_1 in each of these balls are disks, they further chop these balls up into balls, that is to say, M cut open along $\Sigma_1 \cup \Sigma_2 \cup \Delta_2$ consists of balls. As we saw, $\Sigma_1 \cap \Delta_2$ consists entirely of arcs and therefore Δ_2 cut open along Σ_1 consists of disks. Attaching the balls of M cut open along $\Sigma_1 \cup \Sigma_2 \cup \Delta_2$ to each other via these disks we get handlebodies of M cut open along $\Sigma_1 \cup \Sigma_2$. $\square$

9. The an-annular complex C

Using Σ found in Theorem 1.4 we define C to be $\Sigma \cup f(\Sigma)$. C is a complex, mostly a surface, but with some points that are not surface points. At these points C looks like the intersection of two surfaces. Denote this set by $\mathrm{sing}(\mathsf{C})$. However, in this section we will modify C and it will no longer be the union of two surface; we think of C as a collection of embedded surfaces with boundary, disjoint in their interiors, and with images of any two boundary components either disjoint or equal. Then $\mathrm{sing}(\mathsf{C})$ is the union of boundary components. Denote the genus of Σ by g. C has the following properties:

Properties 9.1.

(A) $\chi(\mathsf{C}) \geq 4 - 4g$.

(B) *All components of M cut open along C are handlebodies.*

(C′) *No piece of $\mathsf{C} \setminus \operatorname{sing}(\mathsf{C})$ is a disk.*

(D′) *Every curve of* $\operatorname{sing}(\mathsf{C})$ *is the union of an even number of boundary components.*

(E) *Every torus embedded in C bounds a solid torus.*

(F) C *is invariant under the involution.*

Properties (A), (D′) and (F) are obvious. Properties (B) and (C′) are Theorem 1.4. Property (E) was proved by Kobayashi and Rieck in [13, Corollary 1.3].

However, Property (C′) is insufficient as components of $\mathsf{C} \setminus \operatorname{sing}(\mathsf{C})$ may be annuli, preventing an Euler characteristic count. We need to replace Properties (C′) and (D′) with a stronger version, Properties (C) and (D) below. Achieving these properties is the context of this section and requires us to modify C. To see the relation between Properties (C′) and (D′) and Properties (C) and (D) we mention that in the process of modifying C, we remove from C a neighborhood of $\operatorname{sing}(\mathsf{C})$ and replace it by the boundary of that neighborhood, so all curves of $\operatorname{sing}(\mathsf{C})$ are arranged along tori and have valence three, where the *valence* of a curve of $\operatorname{sing}(\mathsf{C})$ is the number of surfaces adjacent to it locally. (This does not completely describes the modification we perform.)

The closure of a component of $\mathsf{C} \setminus \operatorname{sing}(\mathsf{C})$ is called a *sheet*. In Property (A) stated below we also consider a manifold admitting two strongly irreducible Heegaard splittings of genera g_1 and g_2.

Properties 9.2.

(A) $\chi(\mathsf{C}) \geq 4 - 4g$ *(or* $4 - 2(g_1 + g_2)$*).*

(B) *All components of M cut open along C are handlebodies.*

(C) *Every curve of* $\operatorname{sing}(\mathsf{C})$ *is the union of three boundary components, one of a sheet with negative Euler characteristic and two of annular sheets. These annuli close up, together with other annular sheets, to form tori bounding solid tori (denoted* $\{V_i\}_{i=1}^n$*). For each i,* $\operatorname{int} V_i \cap C = \emptyset$.

(D) *For each i, the number of annuli forming ∂V_i is even.*

(E) *Every torus embedded in* C *bounds a solid torus.*

(F) C *is invariant under the involution.*

Example 9.1. Whenever $\Sigma \cup f(\Sigma)$ contains no annuli, removing from C a neighborhood of $\mathrm{sing}(\mathsf{C})$ and replacing it by the boundary of that neighborhood is sufficient for achieving Properties 9.2. The following example shows that this requirement is sometimes impossible to impose: let M be a genus 2 manifold admitting a free involution and let Σ be a genus two Heegaard surface for M. Suppose Σ intersects $f(\Sigma)$ essentially and spinally and without annuli. It is easy to see that Σ cut open along $\Sigma \cap f(\Sigma)$ consists of two components, either both once punctured tori or both pairs of pants (similarly, $f(\Sigma)$ cut open along $\Sigma \cap f(\Sigma)$ consists two components homeomorphic to the components Σ cut open along $\Sigma \cap f(\Sigma)$). Denote by V one of the handlebodies obtained by cutting M open along Σ. We then see that $\partial(V \cap f(V))$ is a genus two surface and using Lemma 8.1 we deduce that $V \cap f(V)$ is a genus two handlebody. Therefore $f|_{V \cap f(V)}$ is a free involution, and the quotient of $V \cap f(V)$ by the involution has Euler characteristic $-\frac{1}{2}$, contradiction.

We now state the main theorem of this section. In this theorem, Σ (resp. Σ_1 and Σ_2) are the surfaces found in Theorem 7.1 for a manifold with involution (resp. a manifold admitting two strongly irreducible Heegaard splittings).

Recall (Remark 1.1) that if M is non-Haken then any minimal genus Heegaard surface is strongly irreducible; hence the theorem below is not vacuous:

Theorem 9.1. *Let M be an irreducible, orientable, a-toroidal, non-Seifert fibered manifold of Heegaard genus at least two admitting an orientation preserving involution $f : M \to M$ and a strongly irreducible Heegaard surface Σ of genus g (resp. two strongly irreducible Heegaard surfaces or genera g_1 and g_2).*

Then there exists a complex $\mathsf{C} \subset M$ fulfilling Properties 9.2 (resp. Properties 9.2(A)–(E)). Moreover, if Σ (resp. Σ_1 and Σ_2) is the surface found in Theorem 7.1, we may assume that $(\Sigma \cup f(\Sigma)) \setminus (\cup_{i=1}^n V_i) = \mathsf{C} \cap (\cup_{i=1}^n V_i)$ (resp. $(\Sigma_1 \cup \Sigma_2) \setminus (\cup_{i=1}^n V_i) = \mathsf{C} \cap (\cup_{i=1}^n V_i)$).

Remark 9.1. The proof is constructive, giving an algorithm that takes the surface Σ (resp. Σ_1 and Σ_2) found in Theorem 1.4 as input and starting with $\mathsf{C} = \Sigma \cup f(\Sigma)$ (resp. $\Sigma_1 \cup \Sigma_2$) modifies C in finitely many steps until arriving at a complex (still denoted C) fulfilling Properties 9.2. (The algorithm given here is an equivariant version of the algorithm given in [17].)

Question 9.1. We view the cores of the solid tori described in Property 9.2(C) as a link in M. Not every link can arise this way (for example, one can see that Properties 9.2 imply an upper bound on the Heegaard genus of the link exterior and the number of its components). We ask what (and how many) links arise in this way and what other properties do they have.

Proof of Theorem 9.1. The proofs for manifold with involution and manifolds containing two strongly irreducible Heegaard surfaces are identical except for the invariance requirement (Property (F)), which makes the latter strictly easier. We therefore concentrate on the former only.

Starting with $\mathsf{C} = \Sigma \cup f(\Sigma)$, we modify C to fulfill Properties 9.2. As noted above Properties 9.1 are already satisfied, and (unless replaced with stronger properties) they must be preserved throughout the work; that is to say they are *invariants* of the algorithm.

During the modifications of C we construct the solid tori V_i and enlarge them, step by step. However, we never modify C outside these solid tori. This guarantees that for the final complex $(\Sigma \cup f(\Sigma)) \cap (M \setminus (\cup_{i=1}^n V_n)) = \mathsf{C} \cap (M \setminus (\cup_{i=1}^n V_n))$ as required. We will not refer to this again.

The next invariant counts the number of sheets attached to a solid torus $V \subset M$ with $\partial V \subset \mathsf{C}$. We count with multiplicity, that is, a sheet with n boundary components on a V is counted as n sheets attached to V. (Sheets inside V are not counted.) The invariant is:

Invariant 1. *Let $V \subset M$ be a solid torus, $\partial V \subset \mathsf{C}$. Then the number of sheets attached to V is even.*

Proof. Let $\gamma \subset \partial V$ be a curve of $\mathrm{sing}(\mathsf{C})$. Since the valence of γ is four and exactly two sheets attached to γ are part of ∂V, at γ there are either zero, one, or two sheets attached to V. We need to show that the number of the curves with one sheet attached to V is even. By Property (C$'$) γ is essential in ∂V.

Assume first that the slope defined by $\mathrm{sing}(\mathsf{C})$ is the meridian of V. We can then remove ∂V from C, obtaining an immersed surface $\mathscr{S}$. If at γ two sheets of $\mathscr{S}$ are attached to ∂V from outside (resp. inside) V, push $\mathscr{S}$ near γ out of (resp. into) V, removing γ from ∂V. If some component (say F) of $\mathscr{S} \cap \mathrm{int}(V)$ is not a meridian disk then F is either boundary parallel, compressible or boundary compressible. In the first case, F is a boundary parallel annulus (since ∂F is essential in ∂V) and we push F out of V without changing the parity of the number of sheets attached to V.

In the second case, we compress F. If F is boundary compressible but not boundary parallel then F is compressible, so we may ignore the third case (see, for example, [13, Lemma 2.7]). Finally, we see that every component of $\mathscr{S}$ in $\mathrm{int}(V)$ is a meridian disk; by construction the number of meridian disks of $\mathscr{S} \cap V$ has the same parity as the number of curves on ∂V where a single sheet was attached to V. We constructed $\mathscr{S}$ by removing ∂V from C, isotopy and compression. Hence $\mathscr{S}$ is homologous to the null-homologous complex C and the number of times S intersects the core of V is even. This number is exactly the number of meridian disks of $\mathscr{S} \cap V$, proving Invariant 1 in this case.

Next assume that the slope defined by $\mathrm{sing}(\mathsf{C})$ is not meridional. If at γ zero (resp. one, two) sheets are attached to ∂V from outside V, then two (resp. one, zero) sheets are attached to ∂V from *inside* V. Hence the number of sheets attached to V (from outside) is even if and only if the number of sheets attached to ∂V from inside is even. Let F be the closure of a component of $\Sigma \cap \mathrm{int}(V)$ (F may intersect $f(\Sigma)$ in its interior and so may not be a sheet). In [15, Section 2; 19, Theorem 3.3] it was shown that if a strongly irreducible Heegaard surface Σ intersects a solid torus V so that each curve of $\Sigma \cap \partial V$ is a non-meridional essential curve of ∂V, then a component F of $\Sigma \cap V$ is either an annulus, or a twice punctured torus, or a four times punctured sphere; in particular $|\partial F| = 2$ or $|\partial F| = 4$. The same holds for every component of $f(\Sigma) \cap V$. Summing up these numbers gives the number of sheets attached to ∂V from inside; hence this number is even as required. □

The main tool used in this section is:

Definition 9.1. A solid torus V embedded in M is called a *maximal solid torus* if $\partial V \subset \mathsf{C}$ and V is maximal with respect to inclusion among all such solid tori.

By definition a maximal solid torus is an embedded solid torus; in particular, a solid torus embedded in its interior but not in its boundary cannot be a maximal solid torus. Let $\{V_i\}_{i=1}^n$ be the set of all maximal solid tori in M, which is finite since the complex C is.

We would like maximal solid tori to be disjoint; this is not quite the case. For future reference we state this lemma for any complex C fulfilling Property E; in particular, Property F (invariance) is not used in the proof.

Lemma 9.1. *Let C be a complex fulfilling Property E. Then any two distinct maximal solid tori are either disjoint or intersect in a single simple closed curve that is essential in the boundary of both and longitudinal in* (*at least*) *one.*

Proof. Let V_1 and V_2 be distinct maximal solid tori so that $V_1 \cap V_2 \neq \emptyset$. We first show that $V_1 \cap V_2$ is a single simple closed curve. Let $W = V_1 \cup V_2$. Let $\{N_i\}_{i=1}^k$ be the closures of the components of $M \setminus W$. If, for some i, ∂N_i contains an embedded surface (say S) then S is a torus (it has zero Euler characteristic since it is made up of annuli, and is orientable since it locally separates W from N_i in the orientable manifold M). By Property E, S bounds a solid torus in M, and by maximality this solid torus cannot contain V_1 or V_2. Therefore it must contain N_i and we conclude that (since N_i is connected) the solid torus is N_i itself. If N_i is a solid torus for all i then M is the union of $N(W)$ with solid tori. This gives a decomposition of M into solid tori that intersect in annuli. If some slope is meridional, M is reducible or a lens space; else, M is a Seifert fibered space; all conclusions contradict our assumptions.

Therefore, we may assume that some component (say N_1) is not a solid torus and hence no component of ∂N_1 is an embedded surface. Thus there is some curve on ∂N_1 (say γ) where V_1 is tangent to V_2. A neighborhood of γ in C separates a neighborhood of γ in M into four regions, two non-adjacent (say east and west) from V_1 and V_2, and the other two (north and south) from N_1. If $V_1 \cap V_2 = \gamma$ we are done. Thus we may assume $\partial V_1 \cap \partial V_2$ contains at least one more component. Note that ∂V_1 is a torus, formed by gluing an annulus connecting (say) the southeast corner of γ to the northeast corner to itself along γ. Denote this annulus by A_{V_1}, and similarly denote A_{V_2} the annulus connecting the southwest corner to the northwest corner, so that gluing A_{V_2} to itself at γ gives ∂V_2. By assumption A_{V_1} is not disjoint from A_{V_2} in its interior. Let T' be an embedded torus obtained from cut and pasting annuli of A_{V_1} and A_{V_2} cut open along $A_{V_1} \cap A_{V_2}$. Then T' is a toral component of ∂N_1 and by the previous paragraph N_1 is a torus, contradicting out assumption. This shows that $V_1 \cap V_2$ is a single curve.

Next, we show that the slope of $V_1 \cap V_2$ is longitudinal in V_1 or V_2. For contradiction assume that the slope of the intersection is not longitudinal in either solid torus. If it is meridional in both then M contains a non-separating sphere and if it is meridional in one and cabled in the other (i.e., neither meridional nor longitudinal) then M contains a lens space summand, both contradicting our assumptions. So we may assume the slope is cabled in both. Consider W be $N(V_1 \cup V_2)$ which is a Seifert fibered space over

D^2 with exactly two exceptional fibers. Denote ∂W by T. If T bounds a solid torus then either M reduces or M is a Seifert fibered space. Thus T is a torus not bounding a solid torus. By assumption M is irreducible and a-toroidal and therefore T bounds a knot exterior contained in a ball, say X (for details see, for example, [13]). If X were $\mathrm{cl}(M \setminus W)$ then T would be essential, contradicting our assumptions. Hence $X = W$. In [13, Theorem 1.1] Kobayashi and Rieck proved that if a strongly irreducible Heegaard surface intersects a torus bounding a knot exterior contained in a ball in curves that are all essential in the torus, then the slope defined by these curves is meridional. In our case $T \cap \Sigma$ is the slope of a regular fiber in the Seifert fibration by construction, which is not meridional (note that X is a torus knot exterior), contradiction. $\square$

We now modify C in four steps (we do not rename C after each step):

Step 1: Amalgamating maximal solid tori.

Definition 9.2. Let $V_1, V_2 \subset M$ be solid tori such that $\partial V_1, \partial V_2 \subset \mathsf{C}$ and $V_1 \cap V_2$ is a simple closed curve γ, so that γ is essential in ∂V_1 and ∂V_2 and longitudinal in at least one of V_1, V_2. Let $N(\gamma)$ be a small neighborhood of γ, invariant if γ is. Replacing C by $(\mathsf{C} \setminus \mathsf{C} \cap N(\gamma)) \cup (\mathrm{cl}(\partial N(\gamma) \setminus (V_1 \cup V_2)))$ is called *amalgamating* V_1 and V_2 along γ (or simply amalgamating along γ, or amalgamating V_1 and V_2). The two annuli $\mathrm{cl}(\partial N(\gamma) \setminus (V_1 \cup V_2))$ are denoted A_1 and A_2, the solid torus obtained by amalgamating V_1 and V_2 is denoted V, and its boundary is denoted T. Note that $V_1, V_2 \subset V$ and exactly one curve was removed from $\mathrm{sing}(\mathsf{C})$; no other curve of $\mathrm{sing}(\mathsf{C})$ has changed.

Suppose there exist maximal solid tori (say V_1 and V_2) so that $V_1 \cap V_2 \neq \emptyset$. Amalgamate V_1 and V_2 (which can be done by Lemma 9.1). We show that the resulting solid torus V is a maximal solid torus: let U be a maximal solid torus containing V. If U is embedded prior to the amalgamation then $V_1, V_2 \subset U$, contradicting their maximality. Else, prior to the amalgamation U is pinched at γ and broken up to two solid tori, one containing V_1 and the other containing V_2. By maximality, these solid tori are V_1 and V_2 themselves and $U = V$. Therefore V is a maximal solid torus as desired. We verify Property E:

Lemma 9.2. *Let V_1 and V_2 be maximal solid tori in a complex fulfilling Property E and assume V_1 can be amalgamated with V_2. Then C fulfills Property E after amalgamation.*

Proof. Let $T \subset \mathsf{C}$ be a torus after the amalgamation. Then one of the following holds:

(1) $A_1 \not\subset T$ or $A_2 \not\subset T$.

(2) $A_1 \subset T$ and $A_2 \subset T$.

In case (1) T is embedded in C prior to amalgamation. Since C fulfills Property E before the amalgamation T bounds a solid torus. In case (2), prior to the amalgamation there are two tori (say T' and T'') so that $T' \cap T'' = \gamma$ and T is obtained from T' and T'' via surgery. By Property E, T' and T'' bound solid tori (say V' and V'' respectively; note that if $V' \subset V''$ we cannot amalgamate the two). Let U', U'' be the maximal solid tori containing V', V'' respectively. Then $\gamma \subset U'$ and hence so are at least two of the four sheets adjacent to γ. Thus $U' \cap V_1$ or $U' \cap V_2$ contains a sheet, and by Lemma 9.1 either $U' = V_1$ or $U' = V_2$, say the former. Similarly either $U'' = V_1$ or $U'' = V_2$. Since $T' \cap T'' = \gamma$ we see that $U'' = V_2$. Therefore $V' \cap V'' \subset U' \cap U'' = V_1 \cap V_2 = \gamma$, and V' can be amalgamated to V'' along γ. Clearly, T bounds the amalgamation of V' and V''. $\square$

If γ is an invariant curve, we perform the amalgamation invariantly. Else, we amalgamate along $f(\gamma)$; we verify that this can be done: Let V_3 be a maximal solid torus distinct from V_1, V_2 above. If V (the result of amalgamating V_1 and V_2) intersects V_3, by Lemma 9.1 the intersection is a single essential curve that is longitudinal in at least one of the two solid tori. Thus we can amalgamate along $f(\gamma)$ (either amalgamating V and V_3 or amalgamating two maximal solid tori, both distinct from V). After this, Property F is recovered.

We continue amalgamating as long as possible, always performing the amalgamation invariantly. This process reduces $|\mathrm{sing}(\mathsf{C})|$ and hence terminates. When it does, any two maximal solid tori are disjoint and C is invariant. We may now replace Lemma 9.1 with the stronger property below, which is our next invariant:

Invariant 2. *Any two maximal solid tori are disjoint.*

We check invariants:

Property A. $\chi(\mathsf{C})$ has not changed.

Property B. The new components of M cut open along C are solid tori.

Property C$'$. The new sheets are annuli.

Property D′. Some curves are removed from sing(C) and the number of sheets attached to all other curves is unchanged.

Property E. See Lemma 9.2

Property F. By construction.

Invariant 1. In the proof of Lemma 9.2 we saw that any new solid torus (after amalgamating V_1 and V_2) is the amalgamation of two solid tori V', V'' at γ. Prior to the amalgamation, the number of sheets attached to V', V'' is even, and the number of sheets attached to the amalgamation of V' and V'' is the sum of these numbers minus four.

Step 2: Cleaning maximal solid tori. We remove from C every sheet that is in the interior of a maximal solid torus.[7] As a result, the valence of curves of sing(C) on the boundary of each maximal solid torus is either three or four. Let γ be a curve on the boundary of a maximal solid torus V with valence four. We equivariantly deform C by adding a small neighborhood of γ to V, splitting γ into two curves of valence three. This completely describes the modification of C in Step 2.

We show that the tori embedded in C after Step 2 are exactly the boundaries of maximal solid tori before Step 2. In one direction, if V is a maximal solid torus prior to Step 2 then clearly ∂V is a torus embedded in C after Step 2. For the other direction, let $T \subset \mathsf{C}$ be an embedded torus after Step 2. Let T' be the image of the embedding prior to Step 2. If T' is not embedded then T' has a double curve (say γ') on a valence four curves of sing(C) on the boundary of a maximal solid torus. Locally near γ', T' has four annuli, two on ∂V and two attached to ∂V. Let A' be one of the annuli of T' attached to ∂V. It is easy to use the annuli of T' cut open along double curves to cut and paste an embedded torus (say T'') containing A'. By Property E, T'' bounds a solid torus and this solid torus is contained in a maximal solid torus, say V''. By construction $A' \subset V''$ and therefore $V'' \neq V$ and $V'' \cap V \neq \emptyset$, contradicting Invariant 2.

So we may assume that T' is embedded. Then by Property E T' bounds a solid torus, say V. V is contained in some maximal solid torus, say U.

[7]Note that $\Sigma \cup f(\Sigma_t)$ may contains many components inside a maximal solid torus. In that case C will be modified very drastically in Step 2. For example, if V is a maximal solid torus and $\mathsf{C} \cap V$ looks like a grid cross S^1 then in Step 2 many annuli are removed, which is the reason this step is important for the algorithm constructed here.

If $V \neq U$ then parts of ∂V are in the interior of U and are thrown out in Step 2, contradicting choice of T. Hence $V = U$ and $T = \partial U$ as required.

This proves the following invariant, which is stronger than Property E and therefore replaces it:

Invariant 3. *Every torus embedded in C bounds a maximal solid torus that does not intersect C in its interior.*

We call a curve $\gamma \in \text{sing}(\mathsf{C})$ that is on the boundary of a maximal solid torus a *boundary curve* and a sheet on the boundary of a maximal solid torus a *boundary sheet*. If $\gamma \in \text{sing}(\mathsf{C})$ is a boundary curve then by Invariant 2 it is on the boundary of exactly one maximal solid torus and hence of the three sheets attached to γ exactly two are boundary sheets. We replace Property D$'$ by Property D$''$ to accommodate boundary curves:

Property D$''$. If $\gamma \subset \text{sing}(\mathsf{C})$ is not a boundary curve then γ is the union of four boundary components. Every boundary curve has valence three. The boundary of a maximal solid torus consists of an even number of boundary sheets.

Note that Property D$''$ implies Invariant 1 and hence replaces it. We now check our invariants, proving Property D$''$.

Property A. Since no sheet is a disk the Euler characteristic is no more negative than it was.

Property B. The new components of M cut open along C are solid tori.

Property C$'$. The only new sheets are boundary sheets, and they are all annuli.

Property D$''$. For non-boundary curves there is nothing new to prove. For $\gamma \in \mathsf{C}$ a boundary curve, this follows immediately from Invariant 1.

Property F. Since the image of a maximal solid torus is a maximal solid torus, C is invariant.

Invariant 2. The set of maximal solid tori was not changed in Step 2.

Step 3: Curves of sing(C) *not on maximal solid tori.* Let γ be a curve of sing(C) not on the boundary of a maximal solid torus. Note that such a curve was not changed from the original complex.

Definition 9.3. A map from a torus into C is called *admissible* if it is a homeomorphism on the torus except at a finite set of double curves. On double curves the map is 2-to-1 into curves of sing(C).

Thus each double curve either double covers its image or is identified with another double curve. Since $\mathsf{C} \setminus \operatorname{sing}(\mathsf{C})$ contains no disks, the image of the torus cut open along the double curves consists of annuli. Note that we do not require annuli adjacent to double curves to cross each other, that is to say, an annulus coming from the south may be connected to an annulus from the east, while an annulus from the north is connected to an annulus from the west (so an admissible map need not be self-transverse as a map into M).

Lemma 9.3. *The only admissible maps are boundaries of maximal solid tori.*

Proof. For contradiction assume that there exists an admissible map $g : T \to \mathsf{C}$ that is not the boundary of a maximal solid torus. By Invariant 3 every torus embedded in C is the boundary of a maximal solid torus; therefore, the map considered is not an embedding and has a double curve in $\operatorname{sing}(\mathsf{C})$, say γ. Since boundary curves have valence three, γ is not a boundary curve. Therefore $\mathsf{C} \cap N(\gamma)$ was not changed in Steps 1 and 2, and $\mathsf{C} \cap N(\gamma)$ is the intersection of two annuli. Thus γ is the image of two distinct curves on T and these curves cut T into two annuli, say A and A'. Since both boundary components of A map to γ, A defines an admissible map with γ on its boundary and fewer double curves than g. Continuing in this way, we construct an embedding of the torus into C that intersects some curve of $\operatorname{sing}(\mathsf{C})$ that is not on the boundary of a maximal solid torus, contradicting Invariant 3. Thus every admissible map is an embedding and hence the boundary of a maximal solid torus. $\square$

Let γ be a curve of $\operatorname{sing}(\mathsf{C})$ not on the boundary of a maximal solid torus. Replace C by $(\mathsf{C} \setminus N(\gamma)) \cup (\partial N(\gamma))$, introducing a new solid torus. This construction can be done equivariantly by either considering pairs of involute curves, or using the Invariant Neighborhood Theorem on invariant curves.

Let T be a torus embedded in C after Step 3. It is easy to see that T has one of the following two forms: either prior to Step 3 there is some non-boundary curve γ and T is $\partial N(\gamma)$, or prior to Step 3 T is an admissible map. Hence by Lemma 9.3, either T bounds a solid torus V given by $N(\gamma)$ (for some non-boundary curve γ) or T bounds a solid torus V that was a maximal solid tori prior to Step 3. Thus we see that after Step 3 the set of maximal solid tori consists of neighborhoods of non-boundary curves and maximal solid tori prior to Step 3; clearly, distinct maximal solid tori are disjoint. Every curve of $\operatorname{sing}(\mathsf{C})$ is a boundary curve and every non-boundary sheet has its boundary on maximal solid tori. We emphasize that since maximal solid tori are disjoint, every curve of $\operatorname{sing}(\mathsf{C})$ is the boundary of exactly two boundary sheets and one non-boundary sheet (although all three may be annuli). We replace Properties C′ and D″ by Properties C″ and D, which are very close to the required Properties C and D. (In fact, if we could replace "non-positive" in Property C″ by "negative" we'd be done.)

Property C″. Every curve of sing(C) is the union of three boundary components, one of a non-boundary sheet with non-positive Euler characteristic and two annular boundary sheets. These boundary sheets close up, together with other boundary sheets, to form tori bounding solid tori. These solid tori do not intersect C in their interior.

Property D. The number of annuli forming each torus described in Property C is even.

We now check invariance of the properties achieved so far.

Property A. The Euler characteristic was not changed in Step 3.

Property B. All new components of M cut open along C are solid tori.

Property C″. By construction.

Property D. Every new torus has four annuli on its boundary.

Property F. By construction.

Invariant 2. By construction.

Invariant 3. By Lemma 9.3 and the construction.

Remark. We pause for a moment to review what we achieved so far. Recall from Example 9.1 that at the onset our only concern were annular sheets (of course now these sheets are best described as annular non-boundary sheets). Many annular sheets were removed in Step 2. The crucial property we achieved by using maximal solid tori is that chains of (boundary and non-boundary) annular sheets do not close up to form tori, except for boundary of maximal solid tori. This allows us to remove annular non-boundary sheets in Step 4.

Step 4: Getting rid of annular non-boundary sheets. Let A be an annular non-boundary sheet. Assume (for contradiction) that A connects a maximal solid torus (say V) to itself. We use A and an annulus of ∂V cut open along ∂A to form a torus, say T. Let U be the maximal solid torus that T bounds (which exists since every torus bounds a maximal solid torus). Clearly, U and V are distinct maximal solid tori and $U \cap V \neq \emptyset$, contradiction. Thus A connects two distinct maximal solid tori, say V_1 and V_2. Assume (for contradiction) that the slopes defined by ∂A on ∂V_1 and ∂V_2 are both not longitudinal. If both slopes are meridional then M contains a non-separating sphere and if one slope is meridional and the other cabled then M contains a lens space summand, both conclusions contradicting our assumptions. Finally, if the slope is cabled in both V_1 and V_2 then $N(V_1 \cup A \cup V_2)$ is a Seifert fibered space over the disk with exactly two exceptional fibers. As in the proof of Lemma 9.1 it is easy to argue that $\partial N(V_1 \cup A \cup V_2)$ is a torus not bounding a solid torus. Hence $\partial N(V_1 \cup A \cup V_2)$ bounds a knot exterior X, and (since M is a-toroidal) $X = N(V_1 \cup A \cup V_2)$. Then we have:

Claim 1. *$\Sigma \cap \partial X$ or $f(\Sigma) \cap \partial X$ is non-empty and consists of fibers in the Seifert fibration of X.*

Proof. Since M cut open along C consists of handlebodies (and not compression bodies) C is connected. Since $\mathrm{cl}(M \setminus X)$ is not a solid torus, C is not contained in X. Hence $\mathsf{C} \cap \partial X \neq \emptyset$. A fiber in the Seifert fibration is given by a curve on ∂X parallel to ∂A; it is now easy to see that all curves of $\mathsf{C} \cap \partial X$ are parallel (in ∂X) to such a curve, and hence are fibers.

Denote the set of maximal solid tori $V_1, \ldots, V_n$; by Remark 9.1 $(\Sigma \cup f(\Sigma)) \cap (M \setminus (\cup_{i=1}^n V_n)) = \mathsf{C} \cap (M \setminus (\cup_{i=1}^n V_n))$. Since $\partial X \subset (M \setminus (\cup_{i=1}^n V_n))$, we have that $\Sigma \cap \partial X$ or $f(\Sigma) \cap \partial X$ is non-empty and consists of fibers, proving the claim. $\square$

However, by Kobayashi and Rieck [13, Theorem 1.1] if a strongly irreducible Heegaard surface Σ (or $f(\Sigma)$) intersects a knot exterior X contained in a ball and $\Sigma \cap \partial X$ consists of a non-empty collection of curves that are all essential in ∂X then these curves are meridional. The meridian of a torus knot exterior is not a fiber, contradiction. We conclude that A connects two distinct maximal solid tori and is longitudinal in at least one of the two. Similar to Step 1 we amalgamate V_1 with V_2 along A by replacing C with $(\mathsf{C} \setminus N(A)) \cup \mathrm{cl}(\partial N(A) \setminus (V_1 \cup V_2))$, see figure 7. Denote the new component of M cut open along C by V. V is a solid torus.

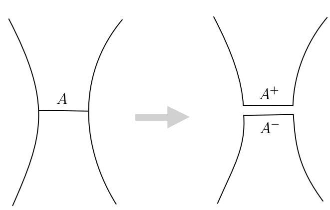

Figure 7: Amalgamation along an annulus.

Let T be a torus embedded in C after the amalgamation. Denote the two parallel copies of A in $\partial N(A)$ by A^+ and A^-. First (cf. Lemma 9.2(1)) suppose $A^+ \not\subset T$ or $A^- \not\subset T$. Then T is embedded prior to the amalgamation and hence T is the boundary of a maximal solid torus (say U) prior to amalgamation. It is straight-forward to see that $U \neq V_1$ and $U \neq V_2$, hence $U \cap V_1 = \emptyset$ and $U \cap V_2 = \emptyset$, that is, U existed as a maximal solid torus prior to the amalgamation. Next, (cf. Lemma 9.2(2)) suppose $A^+ \subset T$ and $A^- \subset T$. Denote the boundary components of A^+ and A^- by S_1^+, S_2^+, S_1^- and S_2^- where $S_i^\epsilon = A^\epsilon \cap V_i$ ($\epsilon = \pm$, $i = 1, 2$). We follow T along, starting at S_1^+, moving across A^+ to S_2^+, continuing until we get to A^-. Assume (for contradiction) that the boundary component we get to is S_1^-. Then the annulus we have traversed from S_1^+ to S_1^- forms an embedded torus (say T') prior to amalgamation. Then T' bounds a maximal solid torus (say U') and it is easy to see that $U' \neq V_1$, and $U' \cap V_1 \neq \emptyset$; contradiction. Thus we conclude that the boundary component we get to is S_2^-. The annulus we traversed from S_2^+ to S_2^- forms an embedded torus (say T') prior to the amalgamation, and the maximal solid torus it bounds intersects V_2. Hence, this maximal solid torus is V_2. Similarly around V_1, and we conclude that T is simply the boundary of the new solid torus. Thus, the set of maximal solid tori after amalgamation is exactly the set of maximal solid tori prior to amalgamation (except V_1 and V_2), with the new solid torus V replacing V_1 and V_2. Note that distinct maximal solid tori are disjoint.

Exactly as in Step 1 we notice that any two maximal solid tori can be amalgamated: let A' be an annulus connecting maximal solid tori. The argument in the beginning of Step 4 shows that A connects distinct maximal solid tori and is longitudinal in at least one. If A is not invariant we retrieve invariance of C by amalgamating along $f(A)$. Iterating Step 4 as long as we can, the process reduces $|\mathrm{sing}(\mathsf{C})|$ and terminates when no non-boundary sheet is an annulus. Note that after Step 4, every torus embedded in C bounds a maximal solid torus and distinct maximal solid tori are disjoint.

We now verify Properties 9.2:

(A) The Euler characteristic of C was not changed in Step 4.

(B) All new components of M cut open along C are solid tori.

(C) By construction, every curve of $\mathrm{sing}(\mathsf{C})$ is on three sheets, two boundary sheets and one non-boundary sheet. Boundary sheets close up to form tori bounding solid tori. In Step 4 we removed all non-boundary annular sheets.

(D) This property is preserved since after amalgamating (say) V_1 and V_2 to obtain V, the number of sheets attached to V is the sum of the sheets attached to V_1 and V_2 minus four.

(E) By construction, every torus embedded in C bounds a (maximal) solid torus.

(F) C is invariant by construction.

This completes the proof of Theorem 9.1. $\square$

10. Constructing the invariant Heegaard surface S

Proof of Theorem 1.1. By Theorem 9.1 M admits a complex C fulfilling Properties 9.2(A)–(F). Recall that the solid tori components of M cut open along C are denoted $\{V_i\}_{i=1}^n$. Let K^* be the complex $\mathsf{C} \cup (\cup_{i=1}^n V_i)$. Note that $\chi(\mathsf{K}^*) = \chi(\mathsf{C})$ and therefore by Property 9.2(A) $\chi(\mathsf{K}^*) \geq 4 - 4g$. By Property F, K^* is invariant under f. We call the components of $\mathsf{K}^* \setminus (\cup_{i=1}^n V_i)$ *sheets* (recall that V_i are close solid tori and therefore the annuli forming ∂V_i are not sheets).

Let K be the complex obtained from K^* by puncturing every sheet once or twice (if necessary for invariance). Note that the only sheets that are punctured twice are sheets that are invariant but admit no fixed point, and every such sheet has Euler characteristic divisible by 2. By Property C every sheet has negative Euler characteristic, we see that every sheet that is punctured once has Euler characteristic at least as negative as -1 and every sheet that is punctured twice has Euler characteristic at least as negative as -2. Every puncture reduces the Euler characteristic by exactly one, and we see that the Euler characteristic is doubled at worst, that is to say, $\chi(\mathsf{K}) \geq 8 - 8g(\Sigma)$.

Let $S = \partial N\mathsf{K}$. Since K is invariant, so is S. On one side (away from K) S bounds components of M cut open along C (that are all handlebodies by Property B) glued to each other along disks that correspond to the punctures of K. Thus S bounds a handlebody on that side. On the side containing K, S bounds the solid tori V_i glued along pieces of the form punctured sheet cross interval. Since a punctured sheet deformation retracts to a spine that contains the boundary of the sheet, a punctured sheet cross interval deformation retracts to a neighborhood of that spine. It is now easy to see that this component of M cut open along S is obtained from $\cup_{i=1}^n V_i$ by attaching 1-handles; hence it too is a handlebody. Thus S is an invariant Heegaard surface and the complementary handlebodies are invariant.

We calculate its genus: $\chi(S) = \chi(\partial N\mathsf{K}) = 2\chi(N\mathsf{K}) = 2\chi(\mathsf{K}) \geq 16 - 16g(\Sigma)$. Thus $2 - 2g(S) \geq 16 - 16g(\Sigma)$. Solving for $g(S)$ we get $g(S) \leq 8g(\Sigma) - 7$. This completes the proof of Theorem 1.1. $\square$

11. Constructing a Heegaard surface for N

Proof of Theorem 1.2. The surface S found in Theorem 1.1 is an invariant Heegaard surface for M and the involution preserves the sides of S. Pick a handlebody of $M \setminus S$, say H.

Claim 2. *The quotient of H by the involution is a handlebody.*

We prove the claim by induction on the genus of H, denoted $g(H)$. For balls, this is a result of Waldhausen [23]. Assume $g(H) > 0$. By the Equivariant Disk Theorem H admits equivariant essential disks, either an invariant disk D or two disjoint disks D, D' that are involutes of each other. The image of the equivariant disks is a single disk $f(D)$. We cut H along the equivariant disks, obtaining H_D. H_D consists of one, two or three handlebodies, all of genus lower than $g(H)$. We cut $f(H)$ along $f(D)$ obtaining $f(H)_{f(D)}$. The projection f induces a cover $f|H_D : H_D \to f(H)_{f(D)}$. By induction, $f(H)_{f(D)}$ consists of handlebodies. Gluing these handlebodies to each other along $f(D)$ we see that the image of H is a handlebody. This proves the claim.

We see that N cut open along the image of S (denoted $S/(f)$) consists of two handlebodies, and therefore $S/(f)$ is a Heegaard surface for $M/(f)$. In Section 10 we saw that $\chi(S) \geq 16 - 16g(\Sigma)$. If $f|_S$ has no fixed points then f induces an unbranched cover $F|_S : S \to S/(f)$. In that case the Euler characteristic is multiplicative and we get: $2 - 2g(S/(f)) = \chi(S/(f)) \geq 8 - 8g(\Sigma)$; solving for $g(S/(f))$ we see that $g(S/(f)) \leq 4g(\Sigma) - 3$. It is easy to see that if the cover $f|_S : S \to S/(f)$ is branched the genus of $S/(f)$ is even lower.

This completes the proof of Theorem 1.2. $\square$

12. Bounding the bridge number of the branch set

In this section we prove Theorem 1.3, bounding the complexity of the branch set of the double cover $f : M \to N$. The branch set is a link in N, denoted L. To measure the complexity of $L \subset N$ we fix a Heegaard surface F for N and isotope the link to intersect each of the handlebodies of N cut open along F in boundary parallel arcs. (To see that this is possible, pick any Heegaard function corresponding to F and pull the maxima of L above zero and the minima below.) We define:

Definition 12.1. Let N be a manifold, $L \subset N$ a link, $F \subset N$ a Heegaard surface, and denote the complementary handlebodies by H_1, H_2. The *bridge number of* L *with respect to* F is the minimal number of arcs in $L' \cap H_1$ for any link L' isotopic to L, subject to the constraint that $L \cap H_1$ and $L \cap H_2$ consists of boundary parallel arcs.

Proof of Theorem 1.3. Note that we need to show that the $b(k)$ is bounded above by $g(S) + 1$, where S is an invariant Heegaard surface for M found in Theorem 1.1. For a double cover $f : M \to N$ the *singular set* is the set of fixed points. Similar to Claim 2 we have:

Claim 3. *Let H be a handlebody of genus $g(H)$ and f be an orientation preserving involution on H. Then the singular set of h consists of at most $g(H) + 1$ arcs, and these arcs are boundary parallel.*

Remark. It is easy to construct involutions that realize the bound above.

We prove the claim by induction on $g(H)$. For balls, this is a result of Waldhausen [23]. Assume $g(H) > 0$. By the Equivariant Disk Theorem H admits equivariant disks, either an invariant disk D (Case 1 and 2 below) or two disjoint disks D, D' that are involutes of each other (Case 3). By the classification of involutions on a disk we know that in the first case the intersection of the singular set with D is either a properly embedded arc (Case 1) or in a single point (Case 2). We prove the claim in each case:

Case 1: A single invariant disk D that intersects the singular set in an arc. Cutting H along D we obtain a handlebody H_D. If D does not separate H we are left with a genus $g(H) - 1$ handlebody and are done by induction. If D does separate H, the two complementary pieces are exchanged by h (note that $h|D$ is a reflection and so orientation reversing) and the singular set of h consists of a single arc.

Case 2: A single invariant disk D that intersects the singular set in a point. If D does not separate H, cutting H open along D we obtain a handlebody of genus $g(H) - 1$. By induction the singular set consists of at most $g(H)$ boundary parallel arcs. If D separates, cutting H open along D we obtain two handlebodies (say H_1 and H_2 of genera g_1 and g_2) with $g_1, g_2 < g(H)$

and $g_1 + g_2 = g(H)$. Since $f|_D$ is orientation preserving, H_1 and H_2 are invariant under f. By induction, the singular set in $f|H_i$ consists of at most $g_i + 1$ boundary parallel arcs $(1 = 1, 2)$. Since $g_1 + g_2 = g(H)$, adding these numbers we get $g_1 + 1 + g_2 + 1 = g(H) + 2$. Luckily, gluing along D, two arcs are identified, becoming a single boundary parallel arc and reducing the number of singular arcs by one.

Case 3: Two disjoint disks D_1, D_2 are exchanged by f. Then $f|_{D_2}, f|_{D_2}$ do not admit a fixed point and therefore the singular set does not intersect D_1 or D_2. Cutting H along D_1 and D_2, we get at most three components, all handlebodies. If there are one or two components, the sum of their genera is strictly less than $g(H)$ and by induction the branch set consists of at most $g(H)$ boundary parallel arcs that remain boundary parallel after gluing. If there are three components, then two components (say H_1 and H_2) are exchanged by the involution and the last component (say $H_{1,2}$) is invariant. Since D_1 and D_2 are not boundary parallel H_1 and H_2 have positive genus, and therefore the genus of $H_{1,2}$ is strictly less than $g(H)$. Since f has no fixed points in H_1 or H_2 the singular set of h is the same as the singular set of $h|_{H_{1,2}}$ and the result follows from the inductive hypothesis. This completes the proof the Claim 3.

Checking the same three cases, one easily proves the following claim. To avoid repetition the details are omitted:

Claim 4. *Let H be a handlebody and $f : H \to H$ an orientation preserving involution. Then $H/(f)$ is a handlebody and the branch set of H consists entirely of boundary parallel arcs.*

Since an involution is injective on the singular set, the branch set of $f : H \to H/(f)$ has the same number of arcs as its singular set. Theorem 1.3 follows from Claims 3 and 4. $\square$

Acknowledgments

Y.R. was supported in part by JSPS grant P00024 and The 21st Century COE Program "Constitution of wide-angle mathematical basis focused on knots" (Project LeaderAkio Kawauchi).

We are very grateful to Tsuyoshi Kobayashi and Marc Lackenby for many helpful conversations and the anonymous referee for her/his comments. We thank Sean Bowman for the illustrations.

Parts of Y.R.'s work was carried out when he was a JSPS post doctoral fellow of Tsuyoshi Kobayashi in Nara Women's University, and when he was visiting Akio Kawauchi in Osaka City University as a part of his 21st Century COE Program "Constitution of wide-angle mathematical basis focused on knots". He is grateful to both, and the math departments of Nara Women's University and Osaka City University for their warm hospitality.

References

[1] M. Boileau and J.-P. Otal, *Scindements de Heegaard et groupe des homéotopies des petites variétés de Seifert*, Invent. Math. **106**(1) (1991), 85–107.

[2] G. Burde and H. Zieschang, Knots, de Gruyter Studies in Mathematics, **5**, Walter de Gruyter & Co., Berlin, 2nd edition, 2003.

[3] A. J. Casson and C. McA. Gordon, *Reducing Heegaard splittings*, Topol. Appl. **27**(3) (1987), 275–283.

[4] J. Cerf, Les diffeomorphismes de la sphere de dimension trois ($\Gamma(4) = 0$), Lucture Notes Math., 53, Springer, Berlin, 1968.

[5] D. Cooper and M. Scharlemann, *The structure of a solvmanifold's Heegaard splittings*, Proceedings of 6th Gökova Geometry-Topology Conference, **23** (1999), 1–18.

[6] W. Haken, *Some results on surfaces in* 3*-manifolds*, in 'Studies in Modern Topology', ed. P. J. Hilton, Math. Assoc. Amer. (distributed by Prentice-Hall, Englewood Cliffs, N.J.), 1968, 39–98.

[7] J. Hempel, *3-manifolds*, AMS Chelsea Publishing, Providence, RI, 2004, Reprint of the 1976 original.

[8] C. Hodgson and J. H. Rubinstein, *Involutions and isotopies of lens spaces*, Knot theory and manifolds (Vancouver, B.C., 1983), Springer, Berlin, 1985, 60–96.

[9] W. Jaco, *Lectures on three-manifold topology*, CBMS Regional Conference Series in Mathematics, **43**, American Mathematical Society, Providence, RI, 1980.

[10] T. Kobayashi and Y. Rieck, *Heegaard genus of the connected sum of m-small knots*, Comm. Anal. Geom. **14**(5) (2006), 1037–1077.

[11] T. Kobayashi and Y. Rieck, *Manifolds admitting both strongly irreducible and weakly reducible minimal genus Heegaard splittings*, Comm. Anal. Geom. **17**(4) (2009), 637–649.

[12] T. Kobayashi and Y. Rieck, *On the growth rate of tunnel number of knots*, J. Reine Angew. Math. **592** (2006), 63–78.

[13] T. Kobayashi and Y. Rieck, *Local detection of strongly irreducible Heegaard splittings via knot exteriors*, Topology Appl. **138**(1–3) (2004), 239–251.

[14] J. W. Morgan and H. Bass (eds), *The Smith conjecture*, Pure and Applied Mathematics, **112**, Academic Press Inc., Orlando, FL, 1984; Papers presented at the symposium held at Columbia University, New York, 1979.

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

[16] Y. Rieck, *Heegaard structures of manifolds in the Dehn filing space*, Topology, 39(3) (2000), 619–641.

[17] Y. Rieck, *An-annular complexes in 3-manifolds*, Kyungpook Math. J. **45**(4) (2005), 549–559.

[18] H. Rubinstein and M. Scharlemann, *Comparing Heegaard splittings of non-Haken* 3-*manifolds*, Topology, **35**(4) (1996), 1005–1026.

[19] M. Scharlemann, *Local detection of strongly irreducible Heegaard splittings*, Topology Appl. **90**(1–3) (1998), 135–147.

[20] M. Scharlemann, *Heegaard splittings of 3-manifolds*, in 'Low Dimensional Topology', eds. B. Li, S. Wang and X. Zhao, New Stud. Adv. Math., **3**, Int. Press, Somerville, MA, 2003, 25–39.

[21] P. Scott, *The geometries of* 3-*manifolds*, Bull. London Math. Soc. **15**(5) (1983), 401–487.

[22] P. Scott, *Homotopy implies isotopy for some Seifert fibre spaces*, Topology, **24**(3) (1985), 341–351.

[23] F. Waldhausen, *Über Involutionen der* 3-*Sphäre*, Topology, 8 (1969), 81–91.

Department of mathematical Sciences
University of Arkansas
Fayetteville, AR 72701
USA
E-mail address: yoav@uark.edu

Department of Mathematics and Statistics
University of Melbourne
Melbourne, Parkville 3010
Australia
E-mail address: rubin@ms.unimelb.edu.au

Received November 13, 2009