AGT

Algebraic & Geometric Topology

Volume 24 (2024)

A strong Haken theorem

MARTIN SCHARLEMANN

msp

msp

Algebraic & Geometric Topology **24**:2 (2024) 717–753
DOI: 10.2140/agt.2024.24.717
Published: 12 April 2024

A strong Haken theorem

MARTIN SCHARLEMANN

Suppose $M = A \cup_T B$ is a Heegaard split compact orientable 3–manifold and $S \subset M$ is a reducing sphere for M. Haken (1968) showed that there is then also a reducing sphere S^* for the Heegaard splitting. Casson and Gordon (1987) extended the result to ∂–reducing disks in M and noted that in both cases S^* is obtained from S by a sequence of operations called 1–surgeries. Here we show that in fact one may take $S^* = S$.

57K35

It is a foundational theorem of Haken [4] that any Heegaard splitting $M = A \cup_T B$ of a closed orientable reducible 3–manifold M is reducible; that is, there is an essential sphere in the manifold that intersects T in a single circle. Casson and Gordon [1, Lemma 1.1] refined and generalized the theorem, showing that it applies also to essential disks, when M has boundary. More specifically, if S is a disjoint union of essential disks and 2–spheres in M then there is a similar family S^*, obtained from S by ambient 1–surgery and isotopy, such that each component of S^* intersects T in a single circle. In particular, if M is irreducible, so S consists entirely of disks, S^* is isotopic to S.

There is of course a more natural statement, in which S does not have to be replaced by S^*. I became interested in whether the natural statement is true because it would be the first step in a program to characterize generators of the Goeritz group of S^3; see Freedman and the author [3; 8]. Inquiring of experts, I learned that this more natural statement had been pursued by some, but not successfully. Here we present such a proof. A reader who would like to get the main idea in a short amount of time could start with the example in Section 11. Recently, Hensel and Schultens [6] have proposed an alternative proof that applies when M is closed and S consists entirely of spheres.

Here is an outline of the paper: Sections 1 and 2 are mostly a review of what is known; particularly the use of verticality in classical compression bodies, those which have no spheres in their boundary. We wish to allow sphere components in the boundary, and Section 3 explains how to recover the classical results in this context. Section 4 shows how to use these results to inductively reduce the proof of the main theorem to the case when S is connected. The proof when S is connected (the core of the proof) then occupies Sections 6 through 10.

1 Introduction and review

All manifolds considered will be orientable and, unless otherwise described, also compact. For M a 3–manifold, a closed surface $T \subset M$ is a *Heegaard surface* in M if the closed complementary components A and B are each compression bodies, defined below. This structure is called a *Heegaard splitting* and is typically written $M = A \cup_T B$. See, for example, [7] for an overview of the general theory of Heegaard surfaces. Among the foundational theorems of the subject is the following [1].

Suppose T is a Heegaard surface in a Heegaard split 3–manifold $M = A \cup_T B$ and D is a ∂ reducing disk for M, with $\partial D \subset \partial_- B \subset \partial M$.

Theorem 1.1 (Haken, Casson–Gordon) *There is a ∂–reducing disk E for M such that*

- $\partial E = \partial D$,
- *E intersects T in a single essential circle* (*ie E ∂–reduces T*).

Note that D and E are isotopic if M is irreducible; but if M is reducible then there is no claim that D and E are isotopic.

There is a similar foundational theorem, by Haken alone [4], that if M is reducible, there is a reducing sphere for M that intersects T in a single circle (ie it is a reducing sphere for T). But Haken made no claim that the reducing sphere for T is isotopic to a given reducing sphere for M.

The intention of this paper is to fill this gap in our understanding. We begin by retreating to a more general setting. For our purposes, a *compression body* C is a connected 3–manifold obtained from a (typically disconnected) closed surface $\partial_- C$ by attaching 1–handles to one end of a collar of $\partial_- C$. The closed connected surface $\partial C - \partial_- C$ is denoted $\partial_+ C$. This differs from what may be the standard notion in that we allow $\partial_- C$ to contain spheres, so C may be reducible. Put another way, we take the standard notion, but then allow the compression body to be punctured finitely many times. In particular, the compact 3–manifolds whose Heegaard splittings we study may have spheres as boundary components.

Suppose then that $M = A \cup_T B$ is a Heegaard splitting, with A and B compression bodies as above. A disk/sphere set $(S, \partial S) \subset (M, \partial M)$ is a properly embedded surface in M such that each component of S is either a disk or a sphere. A sphere in M is called *inessential* if it either bounds a ball or is parallel to a boundary component of M; a disk is inessential if it is parallel to a disk in ∂M. S may contain such inessential components, but these are easily dismissed, as we will see.

Definition 1.2 The Heegaard splitting T is *aligned* with S (or vice versa) if each component of S intersects T in at most one circle.

For example, a reducing sphere or ∂–reducing disk for T, typically defined as a sphere or disk that intersects T in a single essential circle, are each important examples of an aligned disk/sphere. This new

terminology is introduced in part because, in the mathematical context of this paper, the word "reduce" is used in multiple ways that can be confusing. More importantly, once we generalize compression bodies as above, so that some boundary components may be spheres, there are essential spheres and disks in M that may miss T entirely and others that may intersect T only in curves that are inessential in T. We need to take these disks and spheres into account.

Theorem 1.3 *Suppose that $(S,\partial S)\subset(M,\partial M)$ is a disk/sphere set in M. Then there is an isotopy of T such that afterwards T is aligned with S.*

Moreover, such an isotopy can be found so that, after the alignment, the annular components $S\cap A$, if any, form a vertical family of spanning annuli in the compression body A, and similarly for $S\cap B$.

The terminology "vertical family of spanning annuli" is defined in Section 2.

Note that a disk/sphere set S may contain inessential disks or spheres, or essential disks whose boundaries are inessential in ∂M. Each of these are examples in which the disk or sphere could lie entirely in one of the compression bodies and so be disjoint from T. In the classical setting, Theorem 1.3 has this immediate corollary:

Corollary 1.4 (strong Haken) *Suppose ∂M contains no sphere components. Suppose $S\subset M$ (resp. $(S,\partial S)\subset(M,\partial M)$) is a reducing sphere (resp. ∂–reducing disk) in M. Then S is isotopic to a reducing sphere (resp. ∂–reducing disk) for T.*

The assumption in Corollary 1.4 that there are no sphere components in ∂M puts us in the classical setting, where any reducing sphere S for M must intersect T.

2 Verticality in aspherical compression bodies

We first briefly review some classic facts and terminology for an aspherical compression body C, by which we mean that $\partial_- C$ contains no sphere components. Later, sphere components will add a small but interesting amount of complexity to this standard theory. See [7] for a fuller account of the classical theory. Unstated in that account (and others) is the following elementary observation, which further supports the use of the term "aspherical":

Proposition 2.1 *An aspherical compression body C is irreducible.*

Proof Let Δ be the cocores of the 1–handles used in the construction of C from the collar $\partial_- C\times I$. If C contained a reducing sphere S, that is a sphere that does not bound a ball, a standard innermost disk argument on $S\cap\Delta$ would show that there is a reducing sphere in the collar $\partial_- C\times I$. But since C is assumed to be aspherical, $\partial_- C$ contains no spheres, and it is classical that a collar of a closed orientable surface that is not a sphere is irreducible. (For example, its universal cover is a collar of R^2; the interior of this collar is R^3; and R^3 is known to be irreducible by the Schoenflies theorem [10].) □

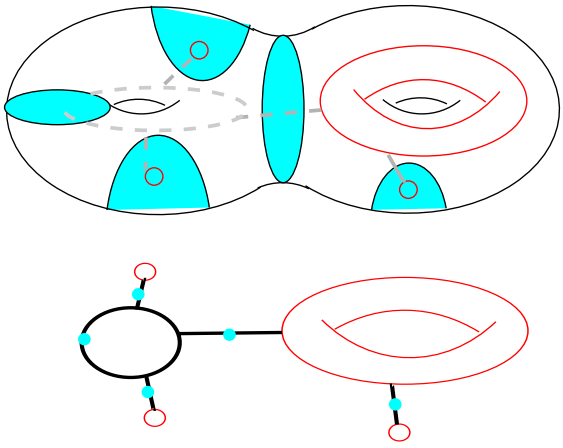

Figure 1: 2–handles and dual spine in a compression body.

Definition 2.2 A properly embedded family $(\Delta, \partial\Delta) \subset (C, \partial C)$ of disks is a *complete collection of meridian disks* for C if $C - \eta(\Delta)$ consists of a collar of $\partial_- C$ and some 3–balls.

That there is such a family of disks follows from the definition of a compression body: take Δ to be the cocores of the 1–handles used in the construction. Given two complete collections Δ and Δ' of meridian disks in an aspherical compression body, it is possible to make them disjoint by a sequence of 2–handle slides, viewing the disks as cocores of 2–handles. (The slides are often more easily viewed dually, as slides of 1–handles.) The argument in brief is this: If Δ and Δ' are two complete collections of meridians, an innermost disk argument (which relies on asphericity) can be used to remove all circles of intersection. A disk cut off from Δ' by an outermost arc γ of $\Delta' \cap \Delta$ in Δ' determines a way of sliding the 2–handle in Δ containing γ over some other members of Δ to eliminate γ without creating more intersection arcs. Continue until all arcs are gone. (A bit more detail is contained in Phase 2 of the proof of Proposition 3.4.)

Visually, one can think of the cores of the balls and 1–handles as a properly embedded graph in C, with some valence 1 vertices on $\partial_- C$, so that the union Σ of the graph and $\partial_- C$ has C as its regular neighborhood. Σ is called a *spine* of the compression body. As already noted, a spine for C is far from unique, but one can move from any spine to any other spine by sliding ends of edges in the graph over other edges, or over components of $\partial_- C$, dual to the 2–handle slides described above. (See [9] or [7].) For most arguments it is sufficient and also simplifying to disregard any valence-one vertex that is not on $\partial_- C$ and the "canceling" edge to which it is attached (but these do briefly appear in the proof of Corollary 5.5); to disregard all valence-two vertices by amalgamating the incident edges into a single edge; and, via a slight perturbation, to require all vertices not on $\partial_- C$ to be of valence three. We can, by edge slides, ensure that only a single edge of the spine is incident to each component of $\partial_- C$; this choice of spine is also sometimes useful.

The spine can be defined as above even when $\partial_- C$ contains spheres. Figure 1 shows a schematic picture of a (nonaspherical) compression body, viewed first with its (aqua) two-handle structure and then its dual

1–handle (spinal) structure. $\partial_- C$ is the union of a torus and 3 spheres; the genus two $\partial_+ C$ appears in the spinal diagram only as an imagined boundary of a regular neighborhood of the spine.

Definition 2.3 A properly embedded arc α in a compression body C is *spanning* if one end of α lies on each of $\partial_- C$ and $\partial_+ C$. Similarly, a properly embedded annulus in C is *spanning* if one end lies in each of $\partial_- C$ and $\partial_+ C$. (Hence, each spanning arc in a spanning annulus is also spanning in the compression body.)

A disjoint collection of spanning arcs α in a compression body is a *vertical family of arcs* if there is a complete collection Δ of meridian disks for C such that

- $\alpha \cap \Delta = \varnothing$ and
- for N, the components of $C - \Delta$ that are a collar of $\partial_- C$, there is a homeomorphism

$$h\colon \partial_- C \times (I, \{0\}) \to (N, \partial_- C)$$

such that $h(\mathfrak{p} \times I) = \alpha$, where $\mathfrak{p}$ is a collection of points in $\partial_- C$.

A word of caution: We will show in Proposition 2.8 that any two vertical arcs with endpoints on the same component $F \subset \partial_- C$ are properly isotopic in C. This is obvious if the two constitute a vertical family. If they are each vertical, but not as a vertical family, proof is required because the collection of meridian disks referred to in Definition 2.3 may differ for the two arcs.

There is a relatively simple but quite useful way of characterizing a vertical family of arcs. To that end, let α be a family of spanning arcs in C and $\hat{p} = \alpha \cap \partial_- C$ be their endpoints in $\partial_- C$. An embedded family c of simple closed curves in $\partial_- C$ is a *circle family associated to* α if $\hat{p} \subset c$.

Lemma 2.4 *Suppose α is a family of spanning arcs in an aspherical compression body C.*

- *Suppose α is vertical and c is an associated circle family. Then there is a family $\mathcal{A}$ of disjoint spanning annuli in C such that $\mathcal{A}$ contains α and $\mathcal{A} \cap \partial_- C = c$.*
- *Suppose, on the other hand, there is a collection $\mathcal{A}$ of disjoint spanning annuli in C that contains α. Suppose further that in the family of circles $\mathcal{A} \cap \partial_- C$ associated to α, each circle is essential in $\partial_- C$. Then α is a vertical family.*

Proof One direction is clear: suppose α is a vertical family and $h\colon \partial_- C \times (I, \{0\}) \to (N, \partial_- C)$ is the homeomorphism from Definition 2.3; then $h(c \times I)$ is the required family of spanning annuli (after the technical adjustment, from general position, of moving the circles $h(c \times \{1\})$ off the disks in $h(\partial_- C \times \{1\})$ coming from the family Δ of meridian disks for C).

For the second claim, let Δ be any complete collection of meridians for C and consider the collection of curves $\Delta \cap \mathcal{A}$. If $\Delta \cap \mathcal{A} = \varnothing$ then $\mathcal{A}$ is a family of incompressible spanning annuli in the collar

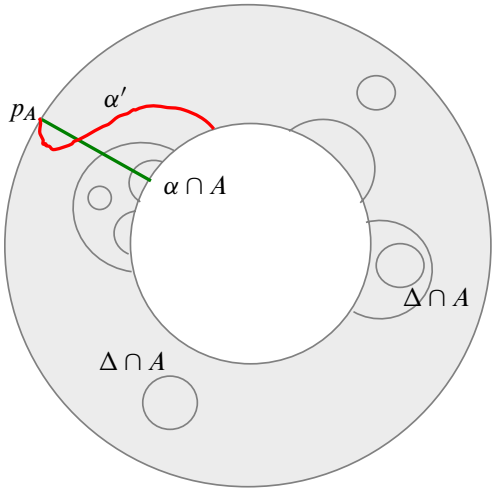

Figure 2: The spanning arc α' avoids $\Delta \cap A$.

$\partial_- C \times I$ and, by standard arguments, any family of incompressible spanning annuli in a collar is vertical. Furthermore, any family of spanning arcs in a vertical annulus can visibly be isotoped rel one end of the annulus to be a family of vertical arcs. So we are left with the case $\Delta \cap \mathcal{A} \neq \varnothing$.

Suppose $\Delta \cap \mathcal{A}$ contains a simple closed curve, necessarily inessential in Δ. If that curve were essential in a component $A \in \mathcal{A}$, then the end $A \cap \partial_- C \subset c$ would be nullhomotopic in C. Since the hypothesis is that each such circle is essential in $\partial_- C$, this would contradict the injectivity of $\pi_1(\partial_- C) \to \pi_1(C)$.

We conclude that each component of $\Delta \cap \mathcal{A}$ is either an inessential circle in $\mathcal{A}$ or an arc in $\mathcal{A}$ with both ends on $\partial_+ C$, since $\partial\Delta \subset \partial_+ C$. Such arcs are inessential in $\mathcal{A}$.

Consider what this means in a component $A \in \mathcal{A}$; let $c_A = A \cap \partial_- C \in c$ be the end of A in $\partial_- C$. It is easy to find spanning arcs α' in A with ends at the points $p_A = \hat{p} \cap c_A$, chosen so that α' avoids all components of $\Delta \cap A$. See Figure 2. But, as spanning arcs, $\alpha \cap A$ and α' are isotopic in A rel c_A (or, if one prefers, one can picture this as an isotopy near A that moves the curves $\Delta \cap A$ off of $\alpha \cap A$). After such an isotopy in each annulus, Δ and α are disjoint. Now apply classic innermost disk, outermost arc arguments to alter Δ until it becomes a complete collection of meridians disjoint from $\mathcal{A}$, the case we have already considered. More details of this classic argument appear in Phase 2 of the proof of Proposition 3.4. □

Lemma 2.4 suggests the following definition.

Definition 2.5 Suppose $\mathcal{A}$ is a family of disjoint spanning annuli in C and α is a collection of disjoint spanning arcs in $\mathcal{A}$, with at least one arc of α in each annulus of $\mathcal{A}$. $\mathcal{A}$ is a *vertical family of annuli* if and only if α is a vertical family of arcs.

Note that for $\mathcal{A}$ to be vertical we do not require that $\mathcal{A}$ be incompressible in C. This adds some complexity to our later arguments, particularly the proof of Proposition 3.8.

Proposition 2.6 *Suppose $\mathscr{A}$ is a vertical family of annuli in an aspherical compression body C. Then there is a complete collection of meridian disks for C that is disjoint from $\mathscr{A}$.*

Proof Let $\alpha \subset \mathscr{A}$ be a vertical family of spanning arcs as given in Definition 2.5. Since α is a vertical family of arcs, there is a complete collection Δ of meridian disks for C that is disjoint from α, so Δ intersects $\mathscr{A}$ only in inessential circles, and arcs with both ends incident to the end of ∂A at $\partial_+ C$. As noted in the proof of Lemma 2.4, a standard innermost disk, outermost arc argument can be used to alter Δ to be disjoint from $\mathscr{A}$. □

Corollary 2.7 *Suppose $(\mathscr{D}, \partial\mathscr{D}) \subset (C, \partial_+ C)$ is an embedded family of disks that is disjoint from an embedded family of vertical annuli $\mathscr{A}$ in an aspherical compression body C. Then there is a complete collection of meridian disks for C that is disjoint from $\mathscr{A} \cup \mathscr{D}$.*

Proof Proposition 2.6 shows that there is a complete collection disjoint from $\mathscr{A}$. But the same proof (which exploits asphericity through its use of Lemma 2.4) works here, if we also augment the curves $\Delta \cap \mathscr{A}$ with the circles $\Delta \cap \mathscr{D}$. □

Proposition 2.8 *Suppose F is a component of $\partial_- C$ and α and β are vertical arcs in C with endpoints $p, q \in F$. Then α and β are properly isotopic in C.*

Notice that the proposition does not claim that α and β are parallel, so in particular they do not necessarily constitute a vertical family. Indeed the isotopy from α to β that we will describe may involve crossings between α and β.

Proof Since C is aspherical, $\text{genus}(F) \geq 1$ and there are simple closed curves $c_\alpha, c_\beta \subset F$ such that

- $p \in c_\alpha$ and $q \in c_\beta$,
- c_α and c_β intersect in a single point.

Since α and β are each vertical, it follows from Lemma 2.4 that there are spanning annuli A_α and A_β in C that contain α and β, respectively, and whose ends on F are c_α and c_β, respectively. Since c_α and c_β intersect in a single point, this means that among the curves in $A_\alpha \cap A_\beta$ there is a single arc γ that spans each annulus, and no other arcs are incident to F. The annulus A_α then provides a proper isotopy from the spanning arc α to γ and the annulus A_β provides a proper isotopy from γ to β. Hence, α and β are properly isotopic in C. See Figure 3. □

We now embark on a technical lemma that uses these ideas, which we will need later. Begin with a closed connected surface F that is not a sphere, and say that circles α and β *essentially intersect* if they are not isotopic to disjoint circles and have been isotoped so that $|\alpha \cap \beta|$ is minimized. Suppose $\hat{a} \subset F$ is an embedded family of simple closed curves, not necessarily essential, and p_1 and p_2 are a pair of points disjoint from $\hat{a}$. (We only will need the case of two points; the argument below extends to any finite number, with some loss of clarity in statement and proof.)

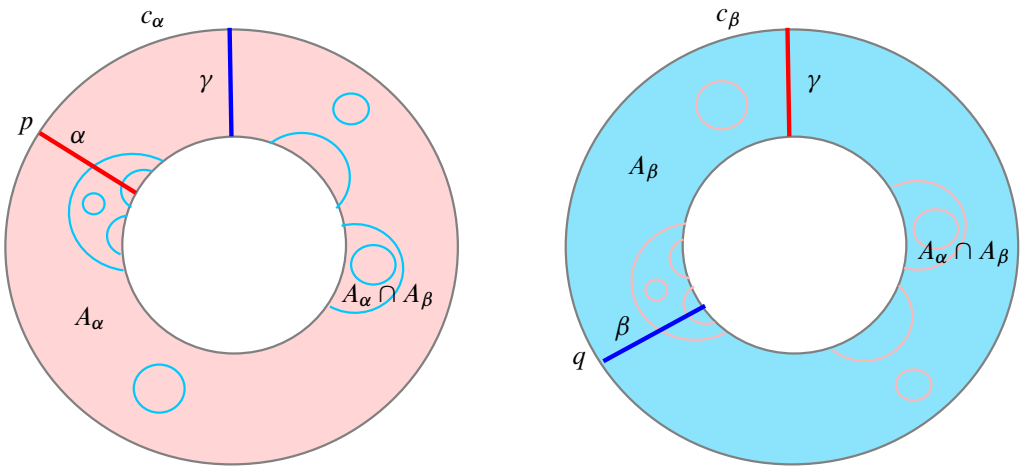

Figure 3: Arcs α and β both properly isotopic to γ.

Let $b' \subset F$ be a nonseparating simple closed curve in F that is not parallel to any $a \in \hat{a}$. For example, if all curves in $\hat{a}$ are separating, b' could be any nonseparating curve; if some curve $a \in \hat{a}$ is nonseparating, take b' to be a circle that intersects a once. Isotope b' in F so that it contains p_1 and p_2, and intersects $\hat{a}$ transversally if at all; call the result $b \subset F$. (Note that, following these requirements, $\hat{a}$ may not intersect b essentially, for example if an innermost disk in F cut off by an inessential $a \in \hat{a}$ contains p_i.) If b intersects $\hat{a}$, let q_i be points in $b \cap \hat{a}$ such that the subintervals $\sigma_i \subset b$ between p_i and q_i have interiors disjoint from $\hat{a}$ and are also disjoint from each other. Informally, we could say that q_i is the closest point in $\hat{a}$ to p_i along b, and σ_i is the path in b between p_i and q_i.

Since b is nonseparating there is a simple closed curve $x \subset F$ that intersects b exactly twice, with the same orientation (so the intersection is essential). Isotope x along b until the two points of intersection are exactly q_1 and q_2. See Figure 4.

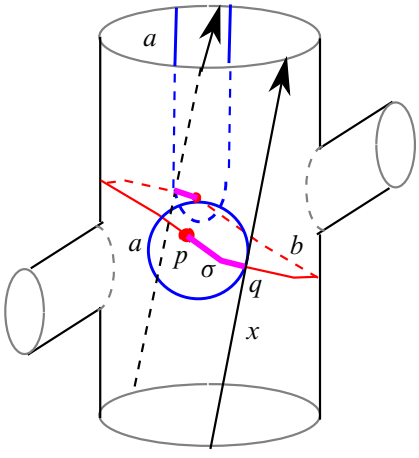

Figure 4: Preamble to Lemma 2.9.

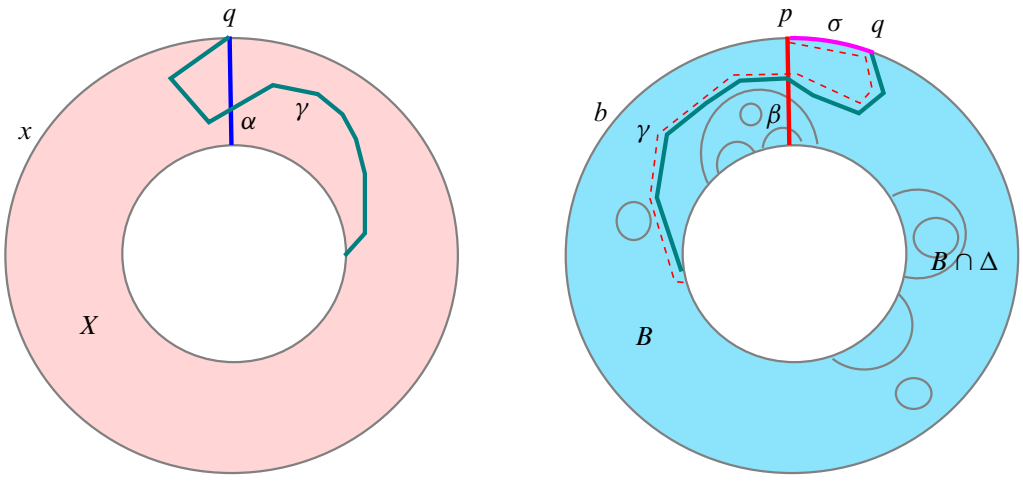

Figure 5: Concluding the proof of Lemma 2.9.

Lemma 2.9 *Let $(\mathcal{D}, \partial\mathcal{D}) \subset (C, \partial_+ C)$ and $\mathcal{A} \subset C$ be as in Corollary 2.7. Suppose $\hat{\beta} = \{\beta_i\}$ for $i = 1, 2$ is a vertical family of arcs in C whose endpoints $p_i \in \partial_- C$ are disjoint from the family of circles $\hat{a} = \mathcal{A} \cap \partial_- C$ in $\partial_- C$. Then $\hat{\beta}$ can be properly isotoped rel $\{p_i\}$ so that it is disjoint from $\mathcal{A} \cup \mathcal{D}$.*

Proof We suppose that both components of $\hat{\beta}$ are incident to the same component F of $\partial_- C$. The proof is essentially the same (indeed easier) if they are incident to different components of $\partial_- C$. Let Δ be a complete family of meridian disks as given in Corollary 2.7, so $\mathcal{A}$ lies entirely in a collar of $\partial_- C$. Per Lemma 2.4, let $B \subset C$ be a spanning annulus that contains the vertical pair $\hat{\beta}$ and has the curve b (from the preamble to this lemma) as its end $B \cap F$ on F.

Suppose first that b is disjoint from $\hat{a}$ and consider $B \cap (\Delta \cup \mathcal{D} \cup \mathcal{A})$. If there were a circle c of intersection that is essential in B, then it could not be in $\Delta \cup \mathcal{D}$, since b does not compress in C. The circle c could not be essential in $\mathcal{A}$, since b was chosen so that it is not isotopic to any element of $\hat{a}$, and it can't be inessential there either again since b does not compress in C. We deduce that there can be no essential circle of intersection, so any circles in $B \cap (\Delta \cup \mathcal{D} \cup \mathcal{A})$ are inessential in B. Also, any arc of intersection must have both ends on $\partial_+ C$ since b is disjoint from $\hat{a}$. It follows that the spanning arcs $\hat{\beta}$ of B can be properly isotoped in B to arcs that avoid $\Delta \cup \mathcal{D} \cup \mathcal{A}$. So, note, they are in the collar of $\partial_- C$ as well as being disjoint from $\mathcal{A} \cup \mathcal{D}$ as required.

Now suppose that b is not disjoint from $\hat{a}$ and let the points q_i, the subarcs σ_i of b and the simple closed curve $x \subset F$ be as described in the preamble to this lemma. By construction, each q_i is in the end of an annulus $A_i \subset \mathcal{A}$; let $\alpha_i \subset A_i$ be a spanning arc of A_i with an end on q_i. Since $\mathcal{A}$ is a vertical family of annuli, α_1 and α_2 are a vertical pair of spanning arcs. Per Lemma 2.4, there is a spanning annulus X that contains the α_i and has the curve x as its end $X \cap F$ on F. Since x essentially intersects b in these two points, $B \cap X$ contains exactly two spanning arcs γ_i, for $i = 1, 2$, each with one endpoint on the respective q_i.

In B the spanning arcs β_i can be properly isotoped rel p_i so that they are each very near the concatenation of σ_i and γ_i; in X the arcs γ_i can be properly isotoped rel q_i to α_i. See Figure 5. (One could also think of this as giving an ambient isotopy of the annulus B so that afterwards $\gamma_i = \alpha_i$.) The combination of these isotopies then leaves β_i parallel to the arc $\sigma_i \cup \alpha_i$. A slight push-off away from A_i leaves β_i disjoint from $\mathcal{A} \cup \mathcal{D}$, as required. □

3 Verticality in compression bodies

We no longer will assume that compression bodies are aspherical. That is, $\partial_- C$ may contain spheres. We will denote by $\widehat{C}$ the aspherical compression body obtained by attaching a 3–ball to each such sphere.

Figure 1 shows a particularly useful type of meridian disk to consider when $\partial_- C$ contains spheres.

Definition 3.1 A complete collection Δ of meridian disks in a compression body C is a *snug collection* if, for each sphere $F \subset \partial_- C$, the associated collar of F in $C - \Delta$ is incident to exactly one disk $D_F \in \Delta$.

The use of the word "snug" is motivated by a simple construction. Suppose Δ is a snug collection of meridian disks for C and $F \subset \partial_- C$ is a sphere. Then the associated disk $D_F \subset \Delta$ is completely determined by a spanning arc α_F in the collar of F in $C - \Delta$, and vice versa: the arc α_F is uniquely determined by D_F, by the light-bulb trick, and once α_F is given, D_F is recovered simply by taking a regular neighborhood of $\alpha_F \cup F$; this regular neighborhood is a collar of F, and the end of the collar away from F itself is the boundary union of a disk in $\partial_+ C$ and a copy of D_F. With that description, we picture D_F as sitting "snugly" around $\alpha_F \cup F$. See Figure 6.

Following immediately from Definition 3.1 is:

Lemma 3.2 *Suppose C is a compression body and $\widehat{\Delta}$ is a collection of meridian disks for C that is a complete collection for the aspherical compression body $\widehat{C}$. Then $\widehat{\Delta}$ is contained in a snug collection for C.*

Proof For each sphere component F of $\partial_- C$, let α_F be a properly embedded arc in $C - \widehat{\Delta}$ from F to $\partial_- C$ and construct a corresponding meridian disk D_F as just described. Then the union of $\widehat{\Delta}$ with all these new meridian disks is a snug collection for C. □

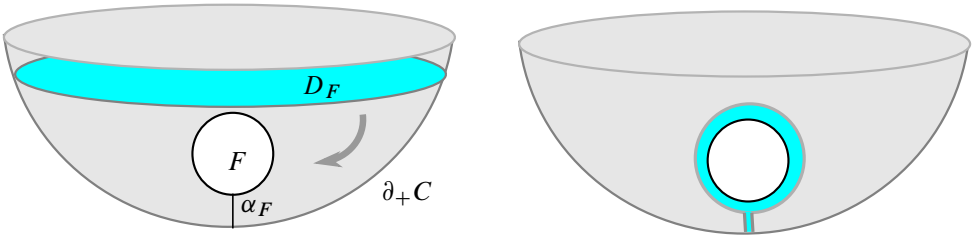

Figure 6: D_F snuggles down around $\alpha_F \cup F$.

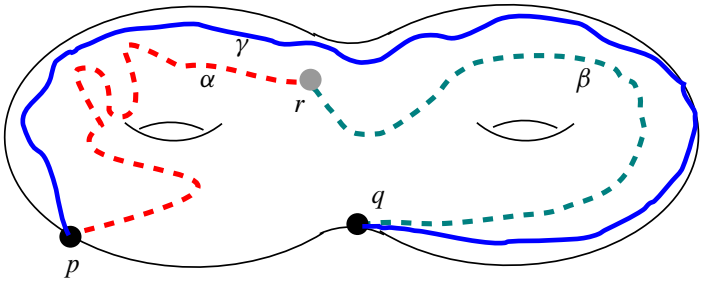

Figure 7

Following Definition 2.2 we noted that for an aspherical compression body, two complete collections of meridian disks can be handle-slid and isotoped to be disjoint. As a useful warm-up we will show that this is also true for snug collections, in case $\partial_- C$ contains spheres. This is the key lemma:

Lemma 3.3 *Suppose C is a compression-body with $p, q \in \partial_+ C$ and $r \in \operatorname{int}(C)$. Suppose α and β are arcs from p and q, respectively, to r in C. Then there is a proper isotopy of β to α in C, fixing r.*

Proof Let Σ be a spine for the compression-body C. By general position, we may take Σ to be disjoint from the path $\alpha \cup \beta$. Since $\pi_1(\partial_+ C) \to \pi_1(C)$ is surjective there is a path γ in ∂C such that the closed curve $\alpha \cup \beta \cup \gamma$ is nullhomotopic in C. See Figure 7. Slide the end of β at q along γ to p so that β becomes an arc β' (parallel to the concatenation of γ and β) also from p to r, one that is homotopic to α rel endpoints. A sophisticated version of the light-bulb trick [5, Proposition 4] then shows that α and β' are isotopic rel endpoints. (Early versions of this paper appealed to the far more complex [2, Theorem 0] to provide such an isotopy.) □

Proposition 3.4 *Suppose Δ and Δ' are snug collections of meridian disks for C. Then Δ can be made disjoint from Δ' by a sequence of handle slides and proper isotopies.*

Proof Let $\mathcal{F} = \{F_i\}$ for $1 \le i \le n$ be the collection of spherical boundary components of C. Since Δ (resp. Δ') is snug, to each F_i there corresponds a properly embedded arc α_i (resp. α_i') in C from F_i to $\partial_+ C$ and this arc determines the meridian disk in $D_i \subset \Delta$ (resp. $D_i' \subset \Delta'$) associated to F_i as described after Definition 3.1. The proof in the aspherical case (as outlined following Definition 2.2; see also [7]) was achieved by isotopies and slides reducing $|\Delta \cap \Delta'|$. In the general case the proof proceeds in two phases.

Phase 1 We will properly isotope the arcs $\{\alpha_i\}$ to $\{\alpha_i'\}$ for $1 \le i \le n$. The associated ambient isotopy of Δ in C may increase $|\Delta \cap \Delta'|$ but in this first phase we don't care. Once each $\alpha_i = \alpha_i'$, each snug disk D_i can be made parallel to D_i' by construction.

Pick a sphere component F_i with associated arcs α_i and α_i'. Isotope the end of α_i on F_i to the end r of α_i' at F_i. Temporarily attach a ball B to F_i and apply Lemma 3.3 to the arcs α and α', after which α and α' coincide. By general position, we can assume the isotopy misses the center b of B and by the

light-bulb trick that it never passes through the radius of B between b and r. Now use radial projection from b to push the isotopy entirely out of B and thus back into C.

Having established how to do the isotopy for a single α_i, observe that we can perform such an isotopy simultaneously on all α_i for $1 \le i \le n$. Indeed, anytime the isotopy of α_i is to cross α_j with $i \ne j$ we can avoid the crossing by pushing it along α_j, over the sphere F_j, and then back along α_j; in short, use the light-bulb trick.

Phase 2 We eliminate $\Delta \cap \Delta'$ by reducing $|\Delta \cap \Delta'|$, as in the aspherical case. After Phase 1, the disks $\{D_i\}$ for $1 \le i \le n$ are parallel to the disks $\{D_i'\}$ for $1 \le i \le n$; until the end of this phase we take them to coincide and also to be fixed, neither isotoped nor slid. Denote the complement in Δ (resp. Δ') of this collection of disks $\{D_i\}$ by $\widehat{\Delta}$ (resp. $\widehat{\Delta}'$), since they constitute a complete collection of meridians in $\widehat{C}$. Moreover, the component of $C - \{D_i\}$ containing $\widehat{\Delta}$ and $\widehat{\Delta}'$ is homeomorphic to $\widehat{C}$, so that is how we will designate that component.

Motivated by that last observation, we now complete the proof by isotoping and sliding $\widehat{\Delta}$, much as in the aspherical case, to reduce $|\widehat{\Delta} \cap \widehat{\Delta}'|$. Suppose first there are circles of intersection and let $E' \subset \widehat{\Delta}'$ be a disk with interior disjoint from $\widehat{\Delta}$ cut off by an innermost such circle of intersection in $\widehat{\Delta}'$. Then $\partial E'$ also bounds a disk $E \subset \widehat{\Delta}$ (which may further intersect $\widehat{\Delta}'$). Although C is no longer aspherical, the sphere $E \cup E'$ lies entirely in $\widehat{C}$, which is aspherical, so $E \cup E'$ bounds a ball in $\widehat{C}$, through which we can isotope E past E', reducing $|\widehat{\Delta} \cap \widehat{\Delta}'|$ by at least one.

Once all the circles of intersection are eliminated as described, we consider arcs in $\widehat{\Delta} \cap \widehat{\Delta}'$. An outermost such arc in $\widehat{\Delta}'$ cuts off a disk E' from $\widehat{\Delta}'$ that is disjoint from $\widehat{\Delta}$; the same arc cuts off a disk E from $\widehat{\Delta}$ (which may further intersect $\widehat{\Delta}'$). The properly embedded disk $E \cup E' \subset \widehat{C}$ has boundary on $\partial_+ \widehat{C}$ and its interior is disjoint from Δ. The latter fact means that its boundary lies on one end of the collar $\widehat{C} - \eta(\Delta)$ of a nonspherical component F of $\partial_- C$. But in a collar of F any properly embedded disk is ∂–parallel. Use the disk in the end of the collar (the other end from F itself) to which $E \cup E'$ is parallel to slide E past E' (possibly sliding it over other disks in Δ, including those in $\{D_i\}$), thereby reducing $|\widehat{\Delta} \cap \widehat{\Delta}'|$ by at least one.

Once $\widehat{\Delta}$ and $\widehat{\Delta}'$ are disjoint, slightly push the disks $\{D_i\}$ off the presently coinciding disks $\{D_i'\}$ so that Δ and Δ' are disjoint. □

Energized by these observations we will now show that all the results of Section 2 remain true (in an appropriate form) in compression bodies that are not aspherical; that is, even when there are sphere components of $\partial_- C$. Here are the analogous results, with edits on statement in boldface, and proofs annotated as appropriate:

Lemma 3.5 (cf Lemma 2.4) *Suppose $\hat{\alpha}$ is a family of spanning arcs in compression body C.*

- *Suppose $\hat{\alpha}$ is vertical and c is an associated circle family. Then there is a family $\mathcal{A}$ of disjoint spanning annuli in C such that $\mathcal{A}$ contains $\hat{\alpha}$ and $\mathcal{A} \cap \partial_- C = c$.*

- *Suppose, on the other hand, there is a collection $\mathscr{A}$ of disjoint spanning annuli in C that contains $\hat{\alpha}$. Suppose further that*
  - ***at most one arc in $\hat{\alpha}$ is incident to each sphere component of $\partial_- C$**, and*
  - *in the family of circles $\mathscr{A} \cap \partial_- C$ associated to $\hat{\alpha}$, each circle **lying in a nonspherical component of $\partial_- C$** is essential.*

  Then α is a vertical family.

Proof The proof of the first statement is unchanged.

For the second, observe that by Lemma 2.4 there is a collection $\widehat{\Delta}$ of meridian disks in $\widehat{C}$ such that $\widehat{\Delta}$ is disjoint from each arc $\alpha \in \hat{\alpha}$ that is incident to a nonspherical component of $\partial_- C$. By general position, $\widehat{\Delta}$ can be taken to be disjoint from the balls $C - \widehat{C}$ and so lie in C.

Now consider an arc $\alpha' \in \hat{\alpha}$ that is incident to a sphere F in $\partial_- C$. It may be that $\widehat{\Delta}$ intersects α'. In this case, push a neighborhood of each point of intersection along α' and then over F. Note that this last operation is not an isotopy of $\widehat{\Delta}$ in C, since it pops across F, but that's unimportant — afterwards the (new) $\widehat{\Delta}$ is completely disjoint from α'. Repeat the operation for every component of $\hat{\alpha}$ that is incident to a sphere in $\partial_- C$, so that $\widehat{\Delta}$ is disjoint from all of $\hat{\alpha}$. Now apply the proof of Lemma 3.2, expanding $\widehat{\Delta}$ by adding a snug meridian disk for each sphere in $\partial_- C$, using the corresponding arc in $\hat{\alpha}$ to define the snug meridian disk for spheres that are incident to $\hat{\alpha}$. □

Proposition 3.6 (cf Corollary 2.7) *Suppose $(\mathscr{D}, \partial\mathscr{D}) \subset (C, \partial_+ C)$ is an embedded family of disks that is disjoint from an embedded family of vertical annuli $\mathscr{A}$ in C. Then there is a complete collection of meridian disks for C that is disjoint from $\mathscr{A} \cup \mathscr{D}$.*

Proof Let $\alpha \subset \mathscr{A}$ be a vertical family of spanning arcs as given in Definition 2.5. This means there is a complete collection Δ of meridian disks for C that is disjoint from α, so Δ intersects $\mathscr{A}$ only in inessential circles, and in arcs with both ends incident to the end of $\mathscr{A}$ at $\partial_+ C$.

Let C' be the compression body obtained by attaching a ball to each sphere component of $\partial_- C$ that is *not incident to* $\mathscr{A}$. Because Δ is a complete collection in C, it is also a complete collection in C', since attaching a ball to a collar of a sphere just creates a ball. Consider the curves $\Delta \cap (\mathscr{A} \cup \mathscr{D})$, and proceed as usual, much as in Phase 2 of the proof of Proposition 3.4:

If there are circles of intersection, an innermost one in Δ cuts off a disk $E \subset \Delta$ and a disk $E' \subset (\mathscr{A} \cup \mathscr{D})$ which together form a sphere whose interior is disjoint from $\mathscr{A}$ and so bounds a ball in C'. In C', E' can be isotoped across E, reducing $|\Delta \cap (\mathscr{A} \cup \mathscr{D})|$. On the other hand, if there are no circles of intersection, then an arc of intersection γ outermost in $\mathscr{A} \cup \mathscr{D}$ cuts off a disk $E' \subset (\mathscr{A} \cup \mathscr{D})$ and a disk $E \subset \Delta$ which together form a properly embedded disk E'' in $C' - \Delta$ whose boundary lies on $\partial_+ C$. Since E'' lies in $C' - \Delta$, it lies in a collar of $\partial_- C'$ and so is parallel to a disk in the other end of the collar. (If the relevant component of $\partial_- C'$ is a sphere, we may have to reset E to be the other half of the disk in Δ in which γ lies to accomplish this.) The disk allows us to slide E past E' and so reduce $|\Delta \cap (\mathscr{A} \cup \mathscr{D})|$.

The upshot is that eventually, with slides and isotopies, Δ can be made disjoint from $\Delta \cap (\mathscr{A} \cup \mathscr{D})$ *in* C'. The isotopies themselves can't be done in C, since sphere boundary components *disjoint from* $\mathscr{A}$ may get in the way, but the result of the isotopy shows how to alter Δ (not necessarily by isotopy) to a family of disks Δ' disjoint from $\mathscr{A} \cup \mathscr{D}$ that is complete in C'. Now apply the argument of Lemma 3.2, adding a snug disk to Δ' for each sphere component of $\partial_- C$ that was not incident to $\mathscr{A}$ and so bounded a ball in C'. These additional snug disks, when added to Δ', create a complete collection of meridian disks for C that is disjoint from $\mathscr{A} \cup \mathscr{D}$, as required. □

Proposition 3.7 (cf Proposition 2.8) *Suppose α and β are vertical arcs in C with endpoints p and q in a component $F \subset \partial_- C$. Then α and β are properly isotopic in C.*

Proof If F is not a sphere, apply the argument of Proposition 2.8. If F is a sphere, apply Lemma 3.3. □

Proposition 3.8 (cf Lemma 2.9) *Suppose $(\mathscr{D}, \partial\mathscr{D}) \subset (C, \partial_+ C)$ is an embedded family of disks that is disjoint from an embedded family of vertical annuli $\mathscr{A}$ in C. Suppose $\hat{\beta} = \{\beta_i\}$ for $i = 1, 2$ is a vertical family of arcs in C whose endpoints $p_i \in \partial_- C$ are disjoint from the family of circles $\hat{a} = \mathscr{A} \cap \partial_- C$ in $\partial_- C$. Then β can be properly isotoped rel $\{p_i\}$ so that it is disjoint from $\mathscr{A} \cup \mathscr{D}$.*

Proof The proof, like the statement, is essentially identical to that of Lemma 2.9, with this alteration when $F \subset \partial C_-$ is a sphere: Use Lemma 3.3 to isotope the vertical (hence parallel) pair $\hat{\beta}$ rel p_i until the arcs are parallel to the vertical family of spanning arcs of $\mathscr{A}$ that are incident to F. In particular, we can then take $\hat{\beta}$ to lie in the same collar $F \times I$ as $\mathscr{A}$ does, and to be parallel to $\mathscr{A}$ in that collar. It is then a simple matter, as in the proof of Lemma 2.9, to isotope each arc in $\hat{\beta}$ rel p_i very near to the concatenation of arcs σ_i disjoint from $\mathscr{A}$ and arcs α_i in $\mathscr{A}$ and, once so positioned, to push $\hat{\beta}$ off of $\mathscr{A} \cup \mathscr{D}$. □

Let us now return to the world and language of Heegaard splittings with a lemma on verticality, closely related to ∂–reduction of Heegaard splittings.

Suppose $M = A \cup_T B$ is a Heegaard splitting of a compact orientable 3–manifold M and $(E, \partial E) \subset (M, \partial_- B)$ is a properly embedded disk, intersecting T in a single circle, so that the annulus $E \cap B$ is vertical in B and the disk $E \cap A$ is essential in A. Since $E \cap B$ is vertical, there is a complete collection of meridian disks Δ in the compression body B such that a component N of $B - \Delta$ is a collar of $\partial_- B$ in which $E \cap B$ is a vertical annulus. Parametrize E as a unit disk with center $b \in E \cap A$ and $E \cap B$ the set of points in E with radius $\frac{1}{2} \le r \le 1$. Let ρ be a vertical radius of E, with ρ_A the half in the disk $E \cap A$ and ρ_B the half in the annulus $E \cap B$.

Let $E \times [-1, 1]$ be a collar of the disk E in M and consider the manifold $M_0 = M - (E \times (-\epsilon, \epsilon))$, the complement of a thinner collar of E. It has a natural Heegaard splitting, obtained by moving the solid cylinders $(E \cap A) \times (-1, -\epsilon]$ and $(E \cap A) \times [\epsilon, -1)$ from A to B. Classically, this operation (when E is essential) is called ∂–reducing T along E [7, Definition 3.5]. We denote this splitting by $M_0 = A_0 \cup_{T_0} B_0$,

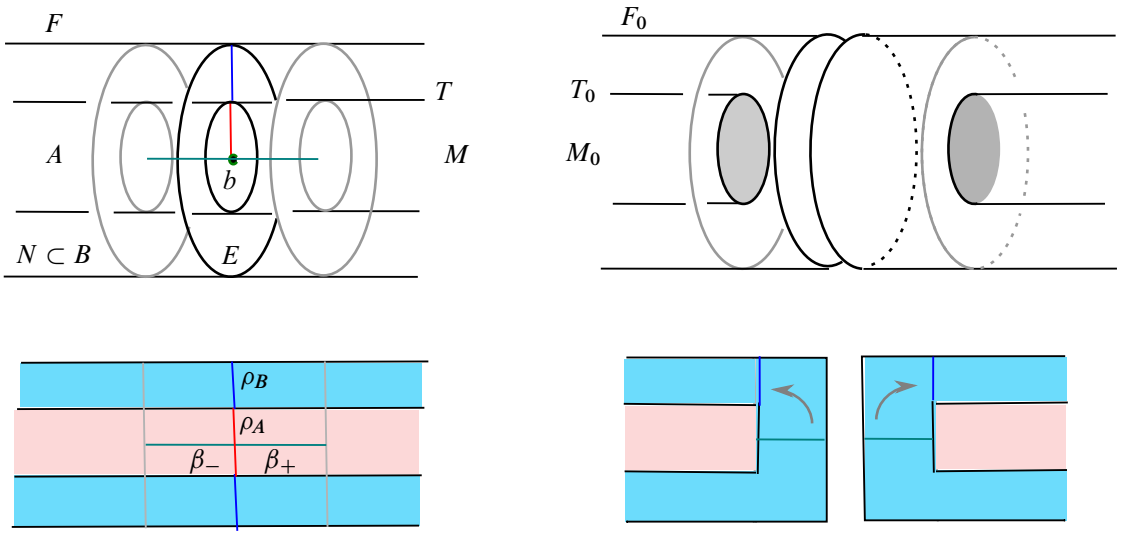

Figure 8

recognizing that if E is separating, it describes a Heegaard splitting of each component. Denote the spanning arcs $b \times [-1,-\epsilon]$ and $b \times [\epsilon,1]$ in B_0 by β_- and β_+, respectively. See the top two panes of Figure 8, with a schematic rendering below.

Lemma 3.9 *The spanning arcs $\beta_\pm$ are a vertical family of arcs in B_0.*

Proof The complete collection of meridian disks Δ for B is disjoint from the annulus $E \cap B$, so remains in B_0. Viewed in the collar component $N \cong (F \times I)$ in the complement of Δ to which $E \cap B$ belongs, the operation described cuts the component $F \subset \partial_- B$ by $\partial E \subset F$, then caps off the boundary circles by disks to get a new surface F_0 and extends the collar structure to $F \times I$. The rectangles $\rho \times [\epsilon,1]$ and $\rho \times [-1,-\epsilon]$ provide isotopies in M_0 from $\beta_\pm$ to the vertical arcs $\rho_B \times \{\pm 1\}$, illustrating that $\beta_\pm$ is a vertical family. □

4 Reducing Theorem 1.3 to the case S is connected

To begin the proof of Theorem 1.3 note that (unsurprisingly) we may as well assume each component of S is essential; that is no sphere in S bounds a ball and no sphere or disk in S is ∂–parallel. This can be accomplished simply by isotoping all inessential components well away from T. So henceforth we will assume all components of S are essential, including perhaps disks whose boundaries are inessential in ∂M but which are not ∂–parallel in M.

Assign a simple notion of complexity (g,s) to the pair (M,T), with g the genus of T and s the number of spherical boundary components of M. We will induct on this pair, noting that there is nothing to prove if $g = 0$ and $s \leq 2$.

Suppose then that we are given a disk/sphere set $(S, \partial S) \subset (M, \partial M)$ in which all components are essential. We begin with:

Assumption 4.1 (inductive assumption) Theorem 1.3 is true for Heegaard splittings of manifolds that have lower complexity than that of (M, T).

With this inductive assumption we have:

Proposition 4.2 *It suffices to prove Theorem 1.3 for a single component S_0 of S.*

Proof Let $M = A \cup_T B$ be a Heegaard splitting, $S \subset M$ be a disk/sphere set, in which each component is essential in M, and let S_0 be a component of S that is aligned with T. The goal is to isotope the other components of S so that they are also aligned, using the inductive Assumption 4.1.

Case 1 S_0 is a sphere and $S_0 \cap T = \varnothing$ or an inessential curve in T.

If S_0 is disjoint from T, say $S_0 \subset B$, then it cuts off from M a punctured ball. This follows from Proposition 2.1, which shows that S_0 bounds a ball in the aspherical compression body $\widehat{B}$ and so a punctured 3–ball in B itself. Any component of $S - S_0$ lying in the punctured 3–ball is automatically aligned, since it is disjoint from T. Removing the punctured 3–ball from B leaves a compression body B_0 with still at least one spherical boundary component, namely S_0. The Heegaard split $M_0 = A \cup_T B_0$ is unchanged, except there are fewer boundary spheres in B_0 than in B because S_0 is essential. Now align all remaining components of $S - S_0$ using the inductive assumption, completing the construction.

Suppose next that S_0 intersects T in a single circle that bounds a disk D_T in T, and S_0 can't be isotoped off of T. Then S_0 again bounds a punctured ball in M with $m \geq 1$ spheres of ∂M lying in A and $n \geq 1$ spheres of ∂M lying in B. S_0 itself is cut by T into hemispheres $D_A = S_0 \cap A$ and $D_B = S_0 \cap B$. A useful picture can be obtained by regarding D_A (say) as the cocore of a thin 1–handle in A connecting a copy A_+ of A with m fewer punctures to a boundary component $T_- = D_T \cup D_A$ of an m–punctured ball in A. In this picture, S_0 and T_- are parallel in $\widehat{B}$; the interior of the collar between them has n punctures in B itself. See Figure 9.

Let β be the core of the 1–handle, divided by S_0 into a subarc β_+ incident to $T_+ = \partial A_+$ and β_- incident to the sphere T_-. Now cut M along S_0, dividing it into two pieces. One is a copy $M_+ = A_+ \cup_{T_+} B_+$ of M, but with m fewer punctures in A_+ and $n-1$ fewer in B_+ (a copy of S_0 is now a spherical boundary component of B_+). The other is an $m+n+1$ punctured 3–sphere M_-, Heegaard split by the sphere T_-. (Neither of the spanning arcs β_+ nor β_- play a role in these splittings yet.)

Now apply the inductive assumption to align T_+ and T_- with the disk/sphere set $S - S_0$ (not shown in Figure 9). Afterwards, reattach M_+ to M_- along the copies of S_0 in each. The result is again M, and S is aligned with the two parts T_- and T_+ in T. But to recover T itself, while ensuring that S remains aligned, we need to ensure that β can be properly isotoped rel S_0 so that it is disjoint from $S - S_0$. Such

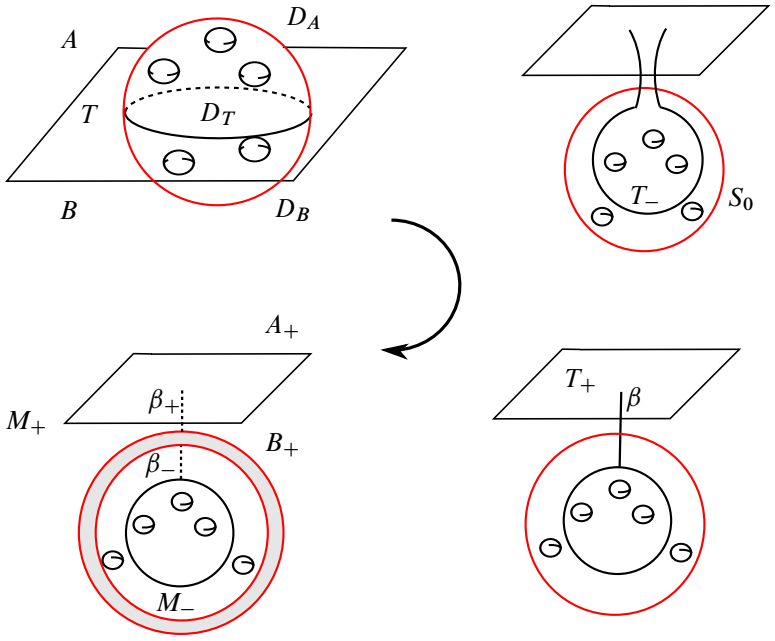

Figure 9: Clockwise through the inductive step in Case 1.

a proper isotopy of β will determine an isotopy of T by viewing β as the core of a tube (the remaining part of T) connecting T_+ to T_-. But once $S - S_0$ is aligned, the proper isotopy of β can be found by first applying Proposition 3.8 to β_+ and the family $S \cap B_0$ of disks and annuli in the compression body B_+ and then proceeding similarly with the arc β_- in M_-.

Case 2 S_0 is a sphere that intersects T in an essential curve.

As in Case 1, S_0 is cut by T into hemispheres $D_A = S_0 \cap A$ and $D_B = S_0 \cap B$ and we can consider D_A (say) as the cocore of a thin 1–handle in A. Continuing as in Case 1, denote the arc core of the 1–handle by β; S_0 again divides the arc β into two arcs which we label $\beta_\pm$.

If S_0 separates, then it divides M into two manifolds, say $M_\pm$ containing, respectively, $\beta_\pm$. Apply the same argument in each that was applied in Case 1 to the manifold M_+.

If S_0 is a nonseparating sphere then we can regard $S - S_0$ as a disk sphere set in the manifold $M_0 = M - \eta(S_0)$. Since S_0 is two-sided, two copies $S_\pm$ of S_0 appear as spheres in ∂M_0. Choose the labeling such that each arc $\beta_\pm$ has one end in the corresponding $S_\pm$. M_0 has lower complexity (the genus is lower) so the inductive assumption applies, and the spheres in $S - S_0$ can be aligned with T_0. Apply Proposition 3.8 to the arcs $\beta_\pm$ and then reconstruct (M, T), now with T aligned with S, as in Case 1.

Case 3 S_0 is a separating disk.

Suppose, with no loss of generality, that $\partial S_0 \subset \partial_- B$, so S_0 intersects A in a separating disk D_A and B in a separating vertical spanning annulus. As in the previous cases, let $M_\pm$ be the manifolds obtained

from M by cutting along S_0, β the core of the 1–handle in A whose cocore is D_A, and $\beta_\pm$ its two subarcs in $M_\pm$, respectively.

The compression body $A-\eta(D_A)$ consists of two compression bodies, $A_\pm$ in $M_\pm$, respectively. As described in the preamble to Lemma 3.9, the complement $B_\pm$ of $A_\pm$ in $M_\pm$ is a compression body, in which $\beta_\pm$ is a vertical spanning arc. So the surfaces $T_\pm$ obtained from T by compressing along D_A are Heegaard splitting surfaces for $M_\pm$, and the pairs $(M_\pm, T_\pm)$ have lower complexity than (M,T).

Now apply the inductive hypothesis: Isotope each of $T_\pm$ in $M_\pm$ so that they align with the components of $S-S_0$ lying in $M_\pm$. As in Case 1, apply Proposition 3.8 to each of $\beta_\pm$ and then reattach M_+ to M_- along disks in $\partial M_\pm$ centered on the points $\beta_\pm\cap\partial M_\pm$ and simultaneously reattach β_+ to β_- at those points. The result is an arc isotopic to β which is disjoint from $S-S_0$. Moreover, the original Heegaard surface T can be recovered from $T_\pm$ by tubing them together along β and, since β is now disjoint from $S-S_0$, all of T is aligned with S.

Case 4 S_0 is a nonseparating disk.

Near S_0 the argument is the same as in Case 3. Now, however, the manifold M_0 obtained by cutting along S_0 is connected. The construction of its Heegaard splitting $M_0=A_0\cup_{T_0}B_0$ and vertical spanning arcs $\beta_\pm$ proceeds as in Case 3, and, since $\text{genus}(T_0)=\text{genus}(T)-1$, we can again apply the inductive hypothesis to align $S-S_0$ with T_0.

If ∂S_0 separates the component F of $\partial_- B\subset\partial M$ in which it lies, say into surfaces $F_\pm$, the argument concludes just as in Case 3. If ∂S_0 is nonseparating in F, then we encounter the technical point that Proposition 3.8 requires that β be a vertical family of arcs. But this follows from Lemma 3.9. □

5 Breaking symmetry: stem swaps

Applications of Lemma 3.3 extend beyond Propositions 2.8 and 3.8. But the arguments will require *breaking symmetry*: given a Heegaard splitting $M=A\cup_T B$ of a compact orientable 3–manifold M and Σ a spine for B, we can, and typically will, regard B as a thin regular neighborhood of Σ, with T as the boundary of that thin regular neighborhood. This allows general position to be invoked as if B were a graph embedded in M. Edge slides of Σ can be viewed as isotopies of T in M and therefore typically are of little consequence. We have encountered this idea in the previous section: the boundary of a tubular neighborhood of an arc β there represented an annulus in T; a proper isotopy of β was there interpreted as an isotopy of T. We can then regard A as the closure of $M-\eta(\Sigma)$; a properly embedded arc in A then appears as an arc whose interior lies in $M-\Sigma$ and whose endpoints may be incident to Σ. We describe such an arc as a properly embedded arc in A whose endpoints lie on Σ. This point of view is crucial to what follows; without it many of the statements might appear to be nonsense.

Let R be a sphere component of $\partial_- B$. Let Σ be a spine for B for which a single edge σ is incident to R.

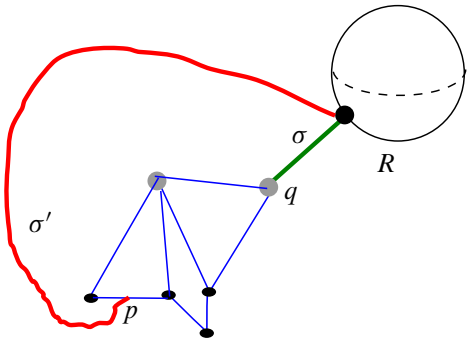

Figure 10: A stem swap for the case $p, q \notin \partial_- B \subset \Sigma$.

Definition 5.1 The complex $\sigma \cup R$ is called a *flower*, with σ the *stem* and R the *blossom*. The point $\sigma \cap R$ is the *base* of the blossom, and the other end of σ is the *base* of both the stem and the flower.

Now suppose σ' is a properly embedded arc in A from the base of the blossom R to a point p in $\Sigma - \sigma$. See Figure 10 for an example when p and q lie on edges of the spine.

Proposition 5.2 (stem swapping) *The complex Σ' obtained from Σ by replacing the arc σ with the arc σ' is, up to isotopy, also a spine for B. That is, T is isotopic in M to the boundary of a regular neighborhood of Σ'.*

Proof Given the spine Σ as described, there is a natural alternative Heegaard splitting for M in which R is regarded as lying in $\partial_- A$ instead of $\partial_- B$. It is obtained by deleting the flower $\sigma \cup R$ from Σ, leaving R as an additional component of $\partial_- A$. Call the resulting spine Σ_- and let A_+ be the complementary compression body (so $M = A_+ \cup_{T'} \eta(\Sigma_-)$). Apply the argument of Lemma 3.3 to A_+, with $\beta = \sigma$, $\alpha = \sigma'$ and $r = R$. (See Phase 1 of the proof of Proposition 3.4 for how we can regard the sphere R as the point r.) Let γ be the path in $\partial_+ A_+ = \partial(\eta(\Sigma_-))$ given by Lemma 3.3. Note that in Figure 10 some edges in the spine Σ_- are shown, but we do not claim that the path γ from Lemma 3.3 is a subgraph of Σ_-. Rather, the path is on the boundary of a *regular neighborhood* of Σ_- and does not necessarily project to an embedded path in Σ_- itself. Note further that after the stem swap the edge in Σ that contains p in its interior (if p is on an edge and not on $\partial_- B$) becomes two edges in Σ' and, dually, when q is not on $\partial_- B \subset \Sigma$, it is natural to concatenate the two edges of Σ that are incident to q into a single edge of Σ'.

Returning to the original splitting, sliding an end of σ along γ does not change the fact that Σ is a spine for B and, viewing T as the boundary of a regular neighborhood of Σ, the slide defines an isotopy of T in M. After the slide, according to Lemma 3.3, σ and σ' have the same endpoints at R and p; then σ can be isotoped to σ' rel its endpoints, completing the proof. (Note that passing σ through σ', as must be allowed to invoke Lemma 3.3, has no significance in this context.) □

Definition 5.3 The operation of Proposition 5.2 in which we replace the stem σ with σ' is called a *stem swap*. If the base of the stem σ' is the same as that of σ, it is called a *local stem swap*.

Definition 5.4 Suppose $M = A \cup_T B$, and Σ is a spine for B. A sphere R_e that intersects Σ in a single point in the interior of an edge e is an *edge-reducing sphere* for Σ and the associated edge e is called a *reducing edge* in Σ.

There is a broader context in which we will consider stem swaps: Let $\mathfrak{R}$ be an embedded collection of edge-reducing spheres for Σ, chosen so that no edge of Σ intersects more than one sphere in $\mathfrak{R}$. (The latter condition, that each edge of Σ intersect at most one sphere in $\mathfrak{R}$, is discussed at the beginning of Section 8.) Let $M_{\mathfrak{R}}$ be a component of $M - \mathfrak{R}$ and $\mathfrak{R}_0 \subset \mathfrak{R}$ be the collection incident to $M_{\mathfrak{R}}$. (Note that a nonseparating sphere in $\mathfrak{R}$ may be incident to $M_{\mathfrak{R}}$ on both its sides. We will be working with each side independently, so this makes very little difference in the argument.)

For a sphere $R_e \in \mathfrak{R}_0$, and $e \in \Sigma$ the corresponding edge, the segment (or segments) $e \cap M_{\mathfrak{R}}$ can each be regarded as a stem in $M_{\mathfrak{R}}$, with blossom (one side of) R_e. A stem swap on this flower can be defined for an arc $\sigma' \subset M_{\mathfrak{R}}$ with interior disjoint from Σ that runs from the point $e \cap R_e$ to a point in $\Sigma \cap M_{\mathfrak{R}}$. Such a swap can be viewed in M as a way of replacing e with another reducing edge e' for R_e that differs from e inside of $M_{\mathfrak{R}}$, leaving the other segment (if any) of e inside $M_{\mathfrak{R}}$ alone.

Corollary 5.5 *If σ and σ' both lie in $M_{\mathfrak{R}}$, then the isotopy of T described in Proposition 5.2 can be assumed to take place entirely in $M_{\mathfrak{R}}$.*

Proof The manifold $M_{\mathfrak{R}}$ has a natural Heegaard splitting $M_{\mathfrak{R}} = A_{\mathfrak{R}} \cup_{T_0} B_{\mathfrak{R}}$ induced by that of M, in which each boundary sphere $R \in \mathfrak{R}_0$ is assigned to $\partial_- B_{\mathfrak{R}}$. We describe this construction:

Recall the setting: Σ is a spine for B and B itself is *a thin regular neighborhood* of Σ. Thus an edge-reducing sphere $R \in \mathfrak{R}$ intersects B in a tiny disk, centered at the point $R \cap \Sigma$. This disk is a meridian of the tubular neighborhood of the reducing edge that contains the point $R \cap \Sigma$. The rest of R, all but this tiny disk, is a disk lying in A. So R is a reducing sphere for the Heegaard splitting of M.

In the classical theory of Heegaard splittings — see eg [7] — such a reducing sphere naturally induces a Heegaard splitting for the manifold $\overline{M}$ obtained by reducing M along R; that is, $\overline{M}$ is obtained by removing an open collar $\eta(R)$ of the sphere R and *attaching 3–balls* to the two copies $R_\pm$ of R at the ends of the collar. The classical argument then gives a natural Heegaard splitting on each component of $\overline{M}$: replace the annulus $T \cap \eta(R)$ by equatorial disks in the two balls attached to $R_\pm$. Translated to our setting, the original spine Σ thereby induces a natural spine on each component of $\overline{M}$: the reducing edge is broken in two when $\eta(R)$ is removed, and at each side of the break, a valence-one vertex is attached, corresponding to the attached ball.

For understanding $M_{\mathfrak{R}}$, we don't care about $\overline{M}$ and the unconventional (because of the valence one vertex) spine just described. We care about the manifold $M - \eta(R)$, in which there are two new sphere boundary components created, but no balls are attached. But the classical construction suggests how to construct a natural Heegaard splitting for the manifold $M - \eta(R)$ and a natural spine for it: simply regard both spheres $R_\pm$ as new components of $\partial_- B$ and attach them at the breaks in the reducing edge where, above,

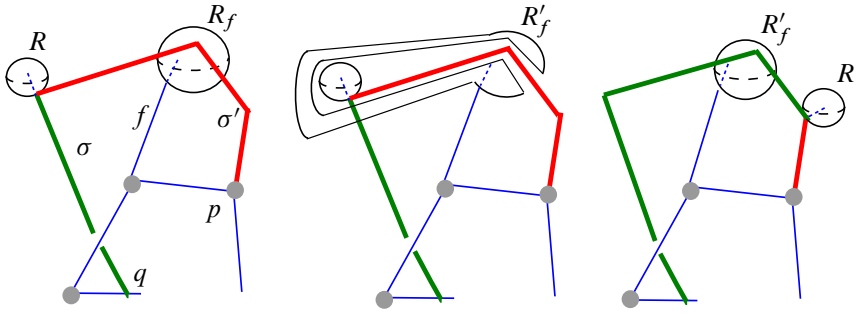

Figure 11: Blossoms R_f and R'_f.

we had added a valence 1 vertex. This Heegaard splitting for $M - \eta(R)$ is topologically equivalent to taking the classical construction of the splitting on $\overline{M}$ and removing two balls from the compression body $\overline{B}$.

When applied to all spheres in $\mathfrak{R}$ simultaneously, the result of this construction is a natural Heegaard splitting on each component of $M - \eta(\mathfrak{R})$. On $M_{\mathfrak{R}}$ it gives the splitting $A_{\mathfrak{R}} \cup_{T_0} B_{\mathfrak{R}}$ which was promised above, and also a natural spine $\Sigma_{\mathfrak{R}}$ for $B_{\mathfrak{R}}$. The required isotopy then follows, by applying Proposition 5.2 to the Heegaard splitting $M_{\mathfrak{R}} = A_{\mathfrak{R}} \cup_{T_0} B_{\mathfrak{R}}$, with $B_{\mathfrak{R}}$ a thin regular neighborhood of the spine $\Sigma_{\mathfrak{R}}$. □

Suppose, in a stem swap, that σ' intersects an edge-reducing sphere R_f, with associated edge $f \neq \sigma$. See the first panel of Figure 11. (Note that f is an edge in Σ but if $p \in f$ then f becomes two edges in Σ'.) Although R_f is no longer an edge-reducing sphere for Σ', there is a natural way to construct a corresponding edge-reducing sphere R'_f for Σ', one that intersects f in the same point, but now intersects σ instead of σ'. At the closest point in which σ' intersects R_f, tube a tiny neighborhood in R_f of the intersection point to its end at R and then around R. Repeat until the resulting sphere is disjoint from σ', as shown in the second panel of Figure 11. One way to visualize the process is to imagine ambiently isotoping R'_f, in a neighborhood of σ', to the position of R_f, as shown in the third panel of Figure 11. The effect of the ambient isotopy is as if R is a bead sitting on the embedded arc $\sigma \cup \sigma'$ and the ambient isotopy moves the bead along this arc and through R_f. We will call R'_f the *swap-mate* of R_f (and vice versa).

Here is an application.

Suppose R_0 is a reducing sphere for a reducing edge $e_0 \in \Sigma$ and $\sigma \subset e_0$ is one of the two segments into which R_0 divides e_0. Let $\sigma' \subset A - R_0$ be an arc whose ends are the same as those of σ but is otherwise disjoint from σ. Let e'_0 be the arc obtained from e by replacing σ with σ'. Let $\eta(R_0)$ be the interior of a collar neighborhood of R_0 on the side away from σ.

Viewing $\sigma \cup R_0$ as a flower in the manifold $M - \eta(R_0)$, and the substitution of σ' for σ as a local stem swap, it follows from the proof of Proposition 5.2 that the 1–complex Σ' obtained from Σ by replacing e_0 with e'_0 is also a spine for B. That is, T is isotopic in M to the boundary of a regular neighborhood of Σ'. Moreover, e'_0 remains a reducing edge in Σ' with edge-reducing sphere R_0.

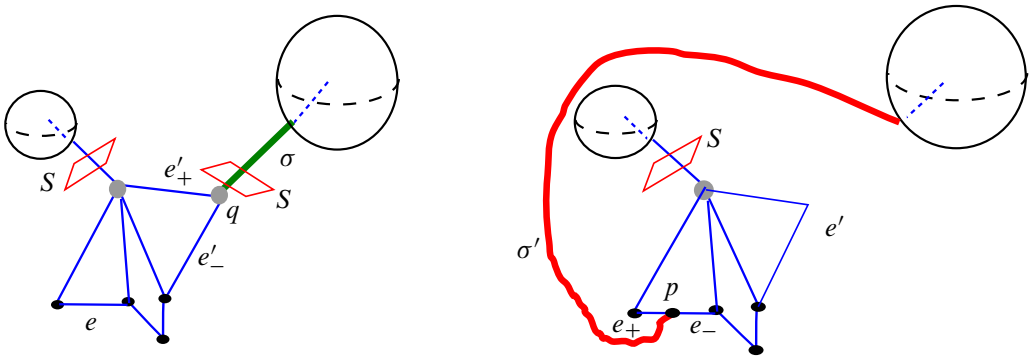

Figure 12: Spines Σ and Σ'.

With this as context, we have:

Lemma 5.6 *Suppose $\mathscr{E}$ is a collection of edges in Σ, with $e_0 \in \mathscr{E}$, and let $\mathscr{E}_r \subset \mathscr{E}$ be the set of reducing edges for Σ that lie in $\mathscr{E}$. Similarly, suppose $\mathscr{E}'$ is a collection of edges in Σ' containing the edge e'_0 constructed above, and $\mathscr{E}'_r \subset \mathscr{E}'$ is the set of reducing edges for Σ' that lie in $\mathscr{E}'$. If $\mathscr{E}' - e'_0 \subset \mathscr{E} - e_0$ then $\mathscr{E}'_r - e'_0 \subset \mathscr{E}_r - e_0$.*

Proof Let f be an edge in $\mathscr{E}'_r$ other than e'_0, and R'_f be a corresponding edge-reducing sphere for Σ'. Then R'_f is disjoint from e'_0, so, although it may intersect e_0, any intersection points lie in $\sigma \subset e_0$. The swap-mate R_f of R'_f then may intersect σ' but by construction it will not intersect σ. Hence, R_f is disjoint from e_0 (as well as all edges of Σ other than f). Hence, R_f is an edge-reducing sphere for Σ and $f \in \mathscr{E}_r$. □

Consider as usual a Heegaard splitting $M = A \cup_T B$, where B is viewed as a thin regular neighborhood of a spine Σ. Suppose $\mathscr{E}$ is a collection of edges in Σ and $\mathscr{E}_r \subset \mathscr{E}$ is the set of reducing edges for Σ that lie in $\mathscr{E}$. (For example, $\mathscr{E}$ might be the set of edges that intersects a specific essential sphere S in M, as in the discussion that will follow Corollary 5.5. This motivates the appearance of the red parallelograms in Figure 12.) Suppose $\mathfrak{R}$ is an embedded collection of edge-reducing spheres for Σ, one associated to each edge in $\mathscr{E}_r$. Let $M_{\mathfrak{R}}$ be a component of $M - \mathfrak{R}$ and consider a sphere $R_0 \in \mathfrak{R}_0 \subset \partial M_{\mathfrak{R}}$. Then, as just described before Lemma 5.6, a segment of the associated reducing edge e_0 that lies in $M_{\mathfrak{R}}$ can be regarded in $M_{\mathfrak{R}}$ as a stem σ with blossom R_0. (The rest of e_0 is shown as a dotted extension in Figure 12.) Let σ' be another arc properly embedded in $M_{\mathfrak{R}}$ which has the same ends as σ but is otherwise disjoint from Σ, and let Σ' be the spine for B constructed as above for the local stem swap of σ to σ'. Notice that because $\operatorname{int}(M_{\mathfrak{R}})$ is disjoint from the spheres $\mathfrak{R}$, $\operatorname{int}(\sigma') \subset \operatorname{int}(M_{\mathfrak{R}})$ is also disjoint from $\mathfrak{R}$.

Proposition 5.7 *Suppose $\mathscr{E}'$ is a subcollection of the edges $\mathscr{E} - e_0$, together possibly with the edge e'_0, and denote by $\mathscr{E}'_r \subset \mathscr{E}'$ the set of reducing edges for Σ' in $\mathscr{E}'$. There is a collection of edge-reducing spheres $\mathfrak{R}'$ for Σ', one associated to each edge in $\mathscr{E}'_r$, such that $\mathfrak{R}' \subset \mathfrak{R}$.*

Proof From Lemma 5.6 we know that $\mathscr{E}'_r - e'_0 \subset \mathscr{E}_r - e_0$. Since σ' is in $M_{\mathfrak{R}}$, it is disjoint from $\mathfrak{R}$, so for each edge f in $\mathscr{E}'_r - e'_0$ we can just use the corresponding edge reducing sphere for f in Σ. In the same vein, since R_0 is disjoint from σ', R_0 is an edge-reducing sphere for e'_0 in Σ'. □

There is an analogous result for more general stem swaps, but it is more difficult to formulate and prove. To that end, suppose $\sigma' \subset M_{\mathfrak{R}}$ has one end at the base of R_0 and the other at a point $p \in \Sigma$. Here p is not a vertex of Σ, nor a point in $\mathfrak{R}$, and $\operatorname{int}(\sigma')$ is disjoint from Σ. If p lies on an edge of Σ, the edge is not one that is also incident to the base point q of σ.

Consider the stem swap as described in Proposition 5.2. After the stem swap, one difference between the two spines Σ and Σ' (other than the obvious switch from σ to σ') is that if p lies on an edge $e \subset \Sigma$ then e becomes two edges $e_\pm$ in Σ' and if the base point q of σ lies on an edge $e' \subset \Sigma'$ then e' began as two edges $e'_\pm$ in Σ. See Figure 12.

Definition 5.8 A collection of edges $\mathscr{E}'$ in Σ' is *consistent with the swap* of σ to σ' (or *swap-consistent*) if, when p and/or q lie on edges as just described, $\mathscr{E}'$ has these properties:

- $\mathscr{E}' - \{e_\pm, e', \sigma'\} \subset \mathscr{E}$.
- If either $e_\pm$ is in $\mathscr{E}'$ then $e \in \mathscr{E}$.
- If both $e'_\pm \notin \mathscr{E}'$ then $e' \notin \mathscr{E}'$. Or, equivalently, if $e' \in \mathscr{E}'$ then at least one of $e'_\pm \in \mathscr{E}$.
- Suppose e is a reducing edge in $\mathscr{E}$ with R_e the corresponding edge-reducing sphere in $\mathfrak{R}$. Then the segment e_+ or e_- not incident to R_e is not in $\mathscr{E}'$. There must be such a segment since by hypothesis $p \notin \mathfrak{R}$.

(In the case that p and/or q lie on $\partial_- B \subset \Sigma$, so the edges e and/or e' are not defined, statements about these edges are deleted.)

Lemma 5.9 *Suppose $\mathscr{E}'$ is consistent with the swap described above. Then there is collection of edge-reducing spheres $\mathfrak{R}'$ for Σ', one associated to each reducing edge in $\mathscr{E}'$, such that $\mathfrak{R}' \subset \mathfrak{R}$.*

Proof Consider any reducing edge $f \in \mathscr{E}'$. If $f = \sigma'$ use R_0 for the corresponding sphere in $\mathfrak{R}'$. In any other case, since f is a reducing edge for an edge in Σ', a corresponding edge-reducing sphere R'_f is automatically disjoint from $\operatorname{int}(\sigma')$ since R'_f only intersects Σ' in a single point. Its swap-mate R_f is then an edge-reducing sphere for Σ, because it is disjoint from $\operatorname{int}(\sigma)$. We do not know that $R_f \in \mathfrak{R}$ and in fact it can't be if $\operatorname{int}(\sigma')$ intersects R_f, since σ' was chosen, following Proposition 5.7, to be in $M_{\mathfrak{R}}$. With this in mind, consider the possibilities:

If $f \notin \{e_\pm, e', \sigma'\}$ then $f \in \mathscr{E}$, since $\mathscr{E}'$ is consistent with the swap. Then R_f is an edge-reducing sphere for f in Σ, so f is a reducing edge in $\mathscr{E}$. As originally defined prior to Proposition 5.7, $\mathscr{E}_r$ is the set of reducing edges in $\mathscr{E}$, so $f \in \mathscr{E}_r$. Since $\mathfrak{R}$ contains an edge-reducing sphere for each edge in $\mathscr{E}_r$,

$\mathfrak{R}$ contains an edge-reducing sphere for f. By construction this sphere is disjoint from both $\mathrm{int}(\sigma)$ and $\mathrm{int}(\sigma')$, the latter by choice of σ'. Include this as the sphere in $\mathfrak{R}'$ that corresponds to f.

As noted at the start, if $f = \sigma'$, use R_0.

If $f = e'$ then one of $e'_\pm$, say e'_+, is in $\mathscr{E}$, since $\mathscr{E}'$ is consistent with the swap. R'_f may as well be taken to pass through $e'_+ \subset e'$. Then R_f is an edge-reducing sphere for Σ that passes through e'_+. Hence, e'_+ is a reducing edge in $\mathscr{E}$. The edge-reducing sphere in $\mathfrak{R}$ corresponding to e'_+ is again disjoint from both $\mathrm{int}(\sigma)$ and $\mathrm{int}(\sigma')$. Include this as the sphere in $\mathfrak{R}'$ that corresponds to f.

If f is one of the edges $e_\pm$, say e_+, then $e \in \mathscr{E}$, since $\mathscr{E}'$ is consistent with the swap. As before, the sphere R_f shows that e is a reducing edge for Σ and so has a corresponding edge-reducing sphere R in $\mathfrak{R}$. Include it in $\mathfrak{R}'$ to correspond to $f = e_+$. The last condition in Definition 5.8 ensures that $e_- \notin \mathscr{E}'$, so no corresponding edge-reducing sphere is included in $\mathfrak{R}'$. In simple terms, R appears only once in $\mathfrak{R}'$. The condition also ensures that f is the subedge of e in Σ' that is incident to R. □

6 When $\partial S \subset \partial_- B \subset \partial M$: early considerations

We will begin the proof of Theorem 1.3 in the case that S is connected. In conjunction with Proposition 4.2, this will complete the proof of Theorem 1.3.

6.1 Preliminary remarks

What will be most important for our purposes is not that S is connected, but that S is entirely disjoint either from all of $\partial_- A$ or all of $\partial_- B$, as is naturally the case when S is connected. So we henceforth assume with no loss of generality that $\partial S \subset \partial_- B$. Following that assumption, the compression bodies A and B play very different roles in the proof. We will be studying spines of B and will take for A the complement in M of a regular neighborhood $\eta(\Sigma)$ of such a spine Σ. In particular, each sphere component R of $\partial_- B$ is part of Σ. As noted in the discussion of spines following Definition 2.2, we can choose Σ so that each sphere component R is incident to exactly one edge of Σ; in that case we are in a position to apply the key idea of stem swapping to alter Σ, as in Proposition 5.2.

In contrast, the sphere components of $\partial_- A$ play almost no role in the proof, other than requiring a small change in language. Since in Theorem 1.3 the isotopy class of S remains fixed (indeed, that is the point of the theorem), we must be careful not to pass any part of S through a sphere component of $\partial_- A$, but the constructions we make use of will avoid this. For example, underlying a stem swap in Σ is the slide and isotopy of an edge of Σ. (See Proposition 5.2.) But these can be made to avoid sphere components of $\partial_- A$, essentially by general position. More explicitly, let $\widehat{M}$ be the 3–manifold obtained from M by attaching a ball to each sphere component of $\partial_- A$. A slide or isotopy of an edge of Σ can avoid the centers of these balls by general position, and then be radially moved outside the entire balls and back into A.

A more subtle problem arises when, for example, we want to use a classical innermost disk (or outermost arc) argument to move a surface F in A so that it is disjoint from S. In the classical setting we find a circle c in $F \cap S$ that bounds a disk $E_S \subset S - F$ and a disk $E_F \subset F$ and argue that one can isotope E_F past E_S, reducing the number of intersections, via a ball whose boundary is the sphere $E_F \cup E_S$. But the existence of such a ball requires A to be irreducible, an assumption that fails when $\partial_- A$ contains spheres. It will turn out that this fraught situation can always be avoided here by *redefining* F to be the surface obtained by a simple disk-exchange, replacing $E_F \subset F$ with a push-off of $E_S \subset S$.

A useful way to visualize and describe this process of redefining F is to imagine, both in the argument and in the figures, a host of bubbles floating around in A, corresponding to sphere components of $\partial_- A$. These bubbles cannot pass through S (or Σ), but typically each bubble can pass "through" other surfaces we construct, in the sense that, when needed, the constructed surface F can be redefined to pass on the other side of the bubble. As shorthand for this process (which we have already seen in Phase 2 of the proof of Proposition 3.4) we will describe the process as a *porous isotopy* of F (equivalent to an actual isotopy in $\widehat{M}$), since the bubbles appear to pass through F.

6.2 The argument begins

Let Σ denote a spine of B and, as usual, take B to be a thin regular neighborhood of Σ.

Let $(\Delta, \partial\Delta) \subset (A, T)$ be a collection of meridian disks for A that constitute a complete collection of meridian disks for $\widehat{A}$, the compression body obtained from A by capping off all spherical boundary components by balls. Let $B_+ = B \cup \eta(\Delta)$; since Δ is complete for $\widehat{A}$, the complement of B_+ is the union of punctured balls and a punctured collar of $\partial_- A \subset \partial M$. The deformation retraction of B to Σ will carry Δ to disks in $M - \Sigma$; continue to denote these by Δ.

Suppose an edge e of Σ is disjoint from Δ. A point on e corresponds to a meridian of B whose boundary lies on ∂B_+. If it is inessential in ∂B_+ then it bounds a disk in A, so such a meridian can be completed to a sphere intersecting e in a single point. In other words, e is a reducing edge of Σ.

The other possibility is that the boundary of the meridian disk for e is essential on ∂B_+, so it, together with an essential curve in $\partial_- A$, bounds an essential spanning annulus $a_e \subset A$. Together, the meridian disk of e and the annulus a_e comprise a boundary reducing disk for M, in fact one that also ∂–reduces the splitting surface T. (In particular, the disk is aligned with T.) We will eliminate from consideration this possibility by a straightforward trick, which we now describe.

Lemma 6.1 *There is a collection $\mathcal{C} \subset \partial_- A$ of disjoint essential simple closed curves with the property that $\mathcal{C}$ intersects any essential simple closed curve in $\partial_- A$ that bounds a disk in M.*

Proof Suppose A_0 is a genus $g \geq 1$ component of $\partial_- A$. By standard duality arguments, the collection $K \subset A_0$ of simple closed curves that compress in M can generate at most a g–dimensional subspace of $H_1(A_0, \mathbb{R}) \cong \mathbb{R}^{2g}$. More specifically, one can find a nonseparating collection $c_1, \ldots, c_g$ of disjoint simple closed curves in A_0 such that $\mathcal{C}_- = \bigcup_{i=1}^{g} c_i$ generates a complementary g–dimensional subspace

of $H_1(A_0,\mathbb{R})$, and therefore essentially intersects any *nonseparating* curve in K. It is easy to add to $\mathscr{C}_-$ a further disjoint collection of $2g-3$ simple closed curves, each nonseparating, so that the result $\mathscr{C}_0 \subset A_0$ has complement a collection of $2g-2$ pairs of pants. Any curve in A_0 that is disjoint from $\mathscr{C}_0$ is parallel to a curve in $\mathscr{C}_0$ and so must be nonseparating. Since it is disjoint from $\mathscr{C}_- \subset \mathscr{C}_0$ it cannot be in K.

Do the same in each component of $\partial_- A$; the result is the required collection C. □

Following Lemma 6.1 add to the collection of disks Δ the disjoint collection of annuli

$$\mathscr{C} \times I \subset \partial_- A \times I \subset M - B_+,$$

and continue to call the complete collection of meridional disks and these spanning annuli Δ. Then a meridian of an edge e of Σ that is disjoint from the (newly augmented) Δ cannot be part of a ∂–reducing disk for T and so must be part of a reducing sphere. Since the collection S of reducing spheres and ∂–reducing disks we are considering have no contact with $\partial_- A$, arcs of $S \cap \Delta$ are nowhere incident to $\partial_- A$. Additionally, no circle in $S \cap \Delta$ can be essential in an annulus in $\mathscr{C} \times I$, since no circle in $\mathscr{C}$ bounds a disk in M. Hence, the annuli which we have added to Δ intersect S much as a disk would: each circle of intersection bounds a disk in the annulus and each arc of intersection cuts off a disk from the same end of the annulus. As a result, the arguments cited below, usually applied to disk components of Δ, apply also to the newly added annuli components $\mathscr{C} \times I$.

7 Reducing edges and S

Lemma 7.1 *Suppose a spine Σ for B and a collection Δ of meridians and annuli, as just described, have been chosen to minimize the pair $(|\Sigma \cap S|, |\partial\Delta \cap S|)$ (lexicographically ordered, with Σ, S and Δ all in general position). Then Σ intersects $\operatorname{int}(S)$ only in reducing edges.*

Notes:

- We do not care about the number of circles in $\Delta \cap S$.
- If S is a disk and intersects Σ transversally only in $\partial S \subset \partial_- B$, then S is aligned with $T = \partial(\eta(\Sigma))$ and intersects B in a vertical annulus, completing the proof of Theorem 1.3 in this case. In addition, S is a ∂–reducing disk for T if ∂S is essential in $\partial_- B$.
- If S is a sphere and intersects Σ transversally only in a single point, then S is aligned with T, completing the proof of Theorem 1.3 in this case. Moreover, if the circle $S \cap T$ is essential in T, S is a reducing sphere for T.

Proof Recall from a standard proof of Haken's theorem — see eg [7; 9, Proposition 2.2] — that $(\Sigma \cup \Delta) \cap S$ (ignoring circles of intersection) can be viewed as a graph Γ in S in which points of $\Sigma \cap S$ are the vertices and $\Delta \cap S$ are the edges. As discussed in [9] in the preamble to Proposition 2.2 there, this is accomplished

by extending the disks and annuli Δ via a retraction $B \to \Sigma$ so that it becomes a collection of disks and annuli whose embedded interior is disjoint from Σ and whose singular boundary lies on Σ. When S is a disk we will, with slight abuse of notation, also regard ∂S as a vertex in the graph, since it lies in $\partial_- B \subset \Sigma$. (This can be made sensible by imagining capping off ∂S by an imaginary disk outside of M.)

Borrowing further from the preamble to [9, Proposition 2.2], an edge in Γ is a loop if both ends lie on the same vertex, called the base vertex for the loop. A loop is *inessential* if it bounds a disk in S whose interior is disjoint from Σ, otherwise it is *essential*. A vertex in Γ is *isolated* if it is incident to no edge in Γ.

It is shown in [9] that if Σ and Δ are chosen to minimize the pair $(|\Sigma \cap S|, |\partial\Delta \cap S|)$ then

- there are no inessential loops,
- any innermost loop in the graph Γ bounds a disk in S that contains only isolated vertices, and
- if there are no loops in Γ then every vertex is isolated.

It follows that either S is disjoint from Σ (so it is aligned and we are done) or there is at least one isolated vertex. An isolated vertex represents a point p in an edge e of Σ which is incident to no element of Δ. The point p defines a meridional disk D_B of $B = \eta(\Sigma)$, and the fact that the curve $\partial D_B \subset \partial_+ A$ is disjoint from Δ ensures that ∂D_B is parallel to a curve in $\partial_- A$ that is inessential. Thus ∂D_B also bounds a disk D_A in A. Then $D_A \cup D_B$ is a reducing sphere, so e is a reducing edge in Σ. This establishes the original Haken theorem and, if there are no loops at all, also Lemma 7.1. That there are no loops is what we now show.

Consider an innermost loop, consisting of a vertex $p \in \Sigma \cap S$ and an edge lying in a component D of Δ. Together, they define a circle c in S that bounds a disk $E \subset S$ whose interior, by the argument of [9, Proposition 2.2], contains only isolated vertices and so intersects Σ only in reducing edges. Remembering that we are taking $A = M - \eta(\Sigma)$, the 3–manifold $A_- = A - \eta(D)$ can be viewed as $M - \eta(D \cup \Sigma)$, so c is parallel in E to a circle c' in ∂A_- bounding a subdisk E_- of E. E_- is the complement in E of the collar in E between c and c'. Since E_- intersects Σ only in reducing edges, it follows immediately that c' is nullhomotopic in A_- and then by Dehn's lemma that it bounds an embedded disk E' entirely in A_-.

By standard innermost disk arguments we can find an E' such that its interior is disjoint from Δ. Now split D in two by compressing the loop to the vertex along E' and replace D in Δ by these two pieces, creating a new complete (for $\widehat{A}$) collection of disks and annuli Δ', with $|\partial\Delta' \cap S| \le |\partial\Delta \cap S| - 2$. Since we have introduced no new vertices, this contradicts our assumption that $(|\Sigma \cap S|, |\partial\Delta \cap S|)$ is minimal. □

Note that the new Δ' may intersect S in many more circles than Δ did, but we don't care.

8 Edge-reducing spheres for Σ

Recall from Section 5 that, given a reducing edge e in Σ, an associated edge-reducing sphere R_e is a sphere in M that passes once through e. Any other edge-reducing sphere R'_e passing once through e is

porously isotopic to R_e in M (ie isotopic in $\widehat{M}$) via edge-reducing spheres. Indeed, the segment of e between the points of intersection with Σ provides an isotopy from the meridian disk $R_e \cap B$ to $R'_e \cap B$; this can be extended to a porous isotopy of $R_e \cap A$ to $R'_e \cap A$ since $\widehat{A}$ is irreducible. So R_e is well-defined up to porous isotopy.

Let Σ be a spine for B in general position with respect to the disk/sphere S, and suppose $\mathscr{E}$ is a collection of edges in Σ. Let $\mathfrak{R}$ be a corresponding embedded collection of edge-reducing spheres transverse to S, one for each reducing edge in $\mathscr{E}$. Let $|\mathfrak{R} \cap S|$ denote the number of components of intersection.

Definition 8.1 The *weight* $w(\mathfrak{R})$ of $\mathfrak{R}$ is $|\mathfrak{R} \cap S|$. Porously isotope $\mathfrak{R}$ via edge-reducing spheres so that its weight is minimized, and call the result $\mathfrak{R}(\mathscr{E})$. Then the *weight* $w(\mathscr{E})$ of $\mathscr{E}$ is $w(\mathfrak{R}(\mathscr{E}))$.

Consider the stem swap as defined in Proposition 5.2 and Corollary 5.5 and suppose $\mathscr{E}'$ is a collection of edges in Σ that is swap-consistent with $\mathscr{E}$.

Lemma 8.2 *There is a collection $\mathfrak{R}'$ of edge-reducing spheres for Σ', one for each reducing edge in $\mathscr{E}'$ such that $w(\mathfrak{R}') \le w(\mathfrak{R})$.*

Proof This is immediate from Lemma 5.9. □

Corollary 8.3 *Suppose in Lemma 8.2 that $\mathfrak{R}$ is $\mathfrak{R}(\mathscr{E})$. Then $w(\mathscr{E}') \le w(\mathscr{E})$.*

Proof Let $\mathfrak{R}'$ be the collection of spheres given in Lemma 8.2. By definition $w(\mathscr{E}') \le w(\mathfrak{R}')$ so, by Lemma 8.2,

$$w(\mathscr{E}') \le w(\mathfrak{R}') \le w(\mathfrak{R}) = w(\mathfrak{R}(\mathscr{E})) = w(\mathscr{E}).$$ □

Here is a motivating example: For Σ a spine of B in general position with respect to S, let $\mathscr{E}$ be the set of edges that intersect S, with the set of edge-reducing spheres $\mathfrak{R} = \mathfrak{R}(\mathscr{E})$ corresponding to the reducing edges of $\mathscr{E}$. As usual, let $M_{\mathfrak{R}}$ be a component of $M - \mathfrak{R}$ and $\mathfrak{R}_0$ be the collection of spheres in $\partial M_{\mathfrak{R}}$ that comes from $\mathfrak{R}$. Suppose R_0 is a sphere in $\mathfrak{R}_0$ with stem σ, and suppose σ' is an arc in $M_{\mathfrak{R}}$ from the base of R_0 to a point p in an edge e of Σ, very near an end vertex of e, so that the subinterval of e between p and the end vertex does not intersect S.

Perform an edge swap and choose $\mathscr{E}'$ to be the set of edges in Σ' that intersect S.

Proposition 8.4 *$\mathscr{E}'$ is swap-consistent with $\mathscr{E}$.*

Proof All but the last property of Definition 5.8 is immediate, because S will intersect an edge if and only if it intersects some subedge. The last property of Definition 5.8 follows from our construction: since σ' lies in a component $M_{\mathfrak{R}}$ of $M - \mathfrak{R}$, the point p lies between the sphere in $\mathfrak{R}$ corresponding to e and an end vertex v of e, and the segment of e between p and v is disjoint from S by construction and therefore not in $\mathscr{E}'$. □

Define the weight $w(\Sigma)$ of Σ to be $w(\mathscr{E})$, and similarly $w(\Sigma') = w(\mathscr{E}')$.

Corollary 8.5 *Given a stem swap as described in Propositions 5.2 or 5.7 for $\mathfrak{R}(\mathscr{E})$, $w(\Sigma') \le w(\Sigma)$.*

Proof This follows immediately from Proposition 8.4 and Corollary 8.3. □

We will need a modest variant of Corollary 8.5 that is similar in proof but a bit more complicated. As before, let $\mathscr{E}$ be the set of edges in a spine Σ that intersect S, with the set of edge-reducing spheres $\mathfrak{R} = \mathfrak{R}(\mathscr{E})$ corresponding to the reducing edges of $\mathscr{E}$. Suppose $e_0 \in \mathscr{E}$ with corresponding edge-reducing sphere $R_0 \in \mathfrak{R}$. Then, by definition,

$$w(\Sigma) = w(\mathscr{E}) = w(\mathfrak{R}) = w(\mathfrak{R} - R_0) + w(R_0) = w(\mathfrak{R} - R_0) + |R_0 \cap S|.$$

Let $\mathfrak{R}_- = \mathfrak{R} - R_0$, $\mathscr{E}_- = \mathscr{E} - e_0$ and $M_{\mathfrak{R}_-}$ be the component of $M - \mathfrak{R}_-$ that contains R_0. Perform an edge swap *in* $M_{\mathfrak{R}_-}$ as in the motivating example: replace the stem σ of a sphere $\mathfrak{a}$ in $\mathfrak{R}_-$ with σ', an arc in $M_{\mathfrak{R}_-}$ from the base of $\mathfrak{a}$ to a point p in an edge e of Σ, very near an end vertex of e, so that the subinterval of e between p and the end vertex does not intersect S. Notice that, in this set-up, R_0 is essentially invisible: the new stem σ' is allowed to pass through R_0. The swap-mate R_0' of R_0 is an edge-reducing sphere for e_0 in Σ' that is disjoint from $\mathfrak{R}_- = \mathfrak{R} - R_0$

As in the motivating example, let $\mathscr{E}'$ be the set of edges in Σ' that intersects S and further define $\mathscr{E}'_- = \mathscr{E}' - e_0$.

Proposition 8.6 $$w(\Sigma') \le w(\Sigma) - |R_0 \cap S| + |R_0' \cap S|.$$

Proof As in the motivating example, $\mathscr{E}'_-$ is consistent with the swap, so by Lemma 5.9 there is a collection $\mathfrak{R}'_- \subset \mathfrak{R}_- = \mathfrak{R} - R_0$ of edge-reducing spheres associated to the edge-reducing spheres of $\mathscr{E}'_-$. Then $\mathfrak{R}'_- \cup R_0'$ is a collection of edge-reducing spheres for $\mathscr{E}'$. Thus,

$$\begin{aligned} w(\Sigma') = w(\mathscr{E}') \le w(\mathfrak{R}'_-) + w(R_0') \le w(\mathfrak{R}_-) + w(R_0') &= w(\mathfrak{R}) - w(R_0) + w(R_0') \\ &= w(\Sigma) - w(R_0) + w(R_0'). \end{aligned}$$ □

9 Minimizing $w(\mathfrak{R}) = |\mathfrak{R} \cap S|$

Following Lemma 7.1, consider all spines that intersect S only in reducing edges, and define $\mathscr{E}$ for each such spine to be as in the motivating example from Section 8: the collection of edges that intersect S. Let Σ be a spine for which $w(\Sigma) = w(\mathscr{E})$ is minimized and let $\mathfrak{R}(\Sigma)$ denote the corresponding collection of edge-reducing spheres for Σ. In other words, among all such spines and collections of edge-reducing spheres, choose that which minimizes the number $|\mathfrak{R} \cap S|$ of (circle) components of intersection.

Proposition 9.1 $\mathfrak{R}(\Sigma)$ *is disjoint from S.*

Note that for this proposition we don't care about how often the reducing edges of the spine Σ intersects S. We revert to the notation $\mathfrak{R}$ for $\mathfrak{R}(\Sigma)$.

Proof Suppose, contrary to the conclusion, $\mathfrak{R} \cap S \neq \varnothing$. Among the components of $\mathfrak{R} \cap S$, pick c to be one that is innermost in S. Let $E \subset S$ be the disk that c bounds in S and let $M_{\mathfrak{R}}$ be the component of $M - \mathfrak{R}$ in which E lies. Let $R_0 \in \mathfrak{R}$ be the edge-reducing sphere on which c lies, $e_0 \subset \Sigma$ the corresponding edge, p be the base $e_0 \cap R_0$ of R_0, and $D \subset R_0$ be the disk c bounds in $R_0 - p$. Finally, as in Proposition 8.6 let $\mathfrak{R}_- = \mathfrak{R} - R_0$ and $M_{\mathfrak{R}_-} \supset M_{\mathfrak{R}}$ be the component of $M - \mathfrak{R}_-$ that contains R_0.

Claim 1 *After local stem swaps as in Proposition 5.7 we can take e_0 to be disjoint from E.*

Let $v_\pm$ be the vertices at the ends of e_0, with $e_\pm$ the incident components of $e_0 - p$. In a bicollar neighborhood of R_0, denote the side of R_0 incident to $e_\pm$ by, respectively, $R_\pm$, with the convention that a neighborhood of ∂E is incident to R_+. It is straightforward to find a point $p' \in R_0$ and arcs $e'_\pm$ in $M_{\mathfrak{R}} - E$, each with one end at the respective vertex $v_\pm$ and other end incident to p' via the respective side $R_\pm$.

It is not quite correct that replacing each of $e_\pm$ with $e'_\pm$ is a local stem swap, since the arcs are incident to R_0 at different points. But this can be easily fixed: Let γ be an arc from p' to p in R_0 and $\gamma_\pm$ be slight push-offs into $R_\pm$. Then replacing each $e_\pm$ with, respectively, $e'_\pm \cup \gamma_\pm$ is a local stem swap. Attach the two arcs at $p \in R_0$ to get a new reducing edge e'_0 for R_0, and then use the arc γ to isotope e'_0 back to the reducing edge $e'_+ \cup e'_-$, which is disjoint from E, as required. See Figure 13. Revert to e_0, p, etc as notation for $e'_+ \cup e'_-$, now disjoint from E.

Claim 2 *After local stem swaps we can assume that each stem that intersects E, intersects it always with the same orientation.*

Figure 14 shows how to use a local stem swap to cancel adjacent intersections with opposite orientations, proving the claim.

Notice that if E is nonseparating in $M_{\mathfrak{R}}$ we could do a local stem swap so that each stem intersects E algebraically zero times. Following Claim 2, this implies that we could make all stems disjoint from E. Once E intersects no stems, replace the subdisk D of R_0 that does not contain p with a copy of E. The result R'_0 is still an edge-reducing sphere for e_0, but the circle c (and perhaps more circles) of intersection with S has been removed. That is,

$$w(R'_0) = |R'_0 \cap S| \leq |R_0 \cap S| - 1 = w(R_0) - 1.$$

Hence, $w(\Sigma') < w(\mathfrak{R}) = w(\Sigma)$, contradicting our hypothesis that $w(\Sigma)$ is minimal.

So we henceforth proceed under the assumption that E is separating, but hoping for the same conclusion: that we can arrange for all stems to be disjoint from E, so that R'_0 as defined above leads to the same contradiction. Since E is separating, a stem that always passes through E with the same orientation can pass through at most once. So we henceforth assume that each stem that intersects E intersects it exactly once.

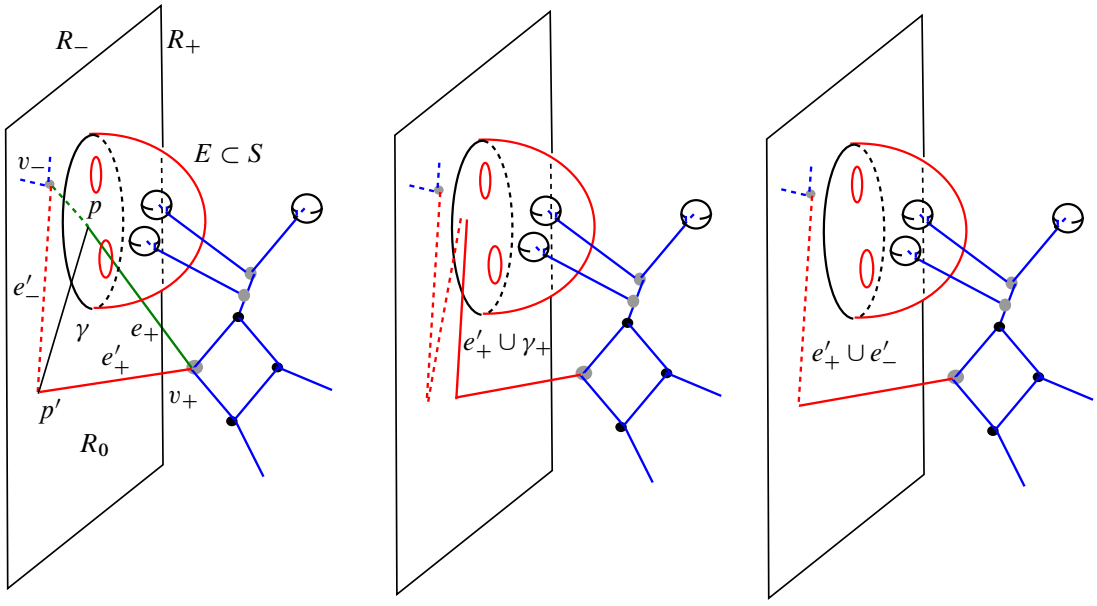

Figure 13: Making e_0 disjoint from E by local stem swaps.

In a bicollar neighborhood of the disk E, let E_+ be the side of E on which v_+ lies, and E_- be the other side of E. Consider a stem σ of a boundary sphere $\mathfrak{a}$ of $M_{\mathfrak{R}-}$. If σ intersects E, the subsegment of $\sigma - E$ that is incident to the blossom $\mathfrak{a}$ passes through one of $E_\pm$. Let $\hat{\sigma}_\pm$ be the collection of those stems intersecting E for which this subsegment passes through, respectively, $E_\pm$. If $\sigma \in \hat{\sigma}_+$, it is straightforward to find an alternative stem σ' from $\mathfrak{a}$ to a point very near v_+ so that σ' misses E. A stem swap to σ' is as in Proposition 5.2, and so by Corollary 8.5 does not increase weight. Hence, we have proven:

Claim 3 *After stem swaps, we may assume that each stem that intersects E is in $\hat{\sigma}_-$.*

Following Claim 3, we move to swap those stems in $\hat{\sigma}_-$ for ones that are disjoint from E. Let σ be the stem of a boundary sphere $\mathfrak{a}$ of $M_{\mathfrak{R}-}$, and assume that $\sigma \in \hat{\sigma}_-$. Then it is straightforward to find an alternative stem σ' for $\mathfrak{a}$ that is disjoint from E and ends in a point very near v_-, for example by concatenating an arc in E_- with an arc in R_- and an arc parallel to e_-. See Figure 15. A problem is, that such an arc intersects the disk $D \subset R_0$, so, after such a swap, R_0 is no longer an edge-reducing sphere

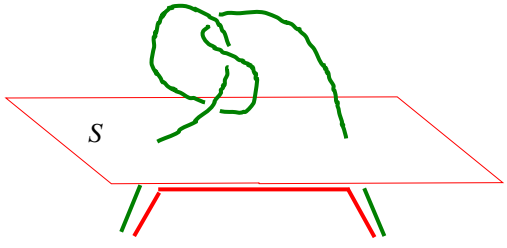

Figure 14: A local stem swap.

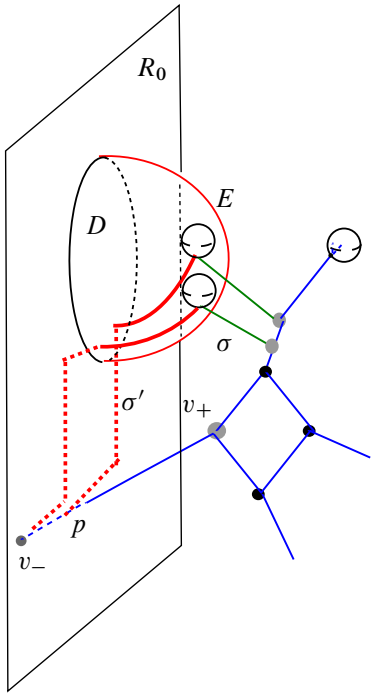

Figure 15

for the new spine. However, if such swaps are performed simultaneously on all stems in $\hat{\sigma}_-$, we have seen that the swap-mate of R_0 is an edge-reducing sphere for the new spine Σ', as required. But observe in Figure 15 that the swap-mate is exactly R_0'! So we can now appeal to Proposition 8.6:

$$w(\Sigma') \le w(\Sigma) - |R_0 \cap S| + |R_0' \cap S| \le w(\Sigma) - 1.$$

The contradiction proves Proposition 9.1. □

10 Conclusion

Proposition 10.1 *Suppose Σ intersects S only in reducing edges, and the associated set $\mathfrak{R}$ of edge-reducing spheres is disjoint from S. Then T can be isotoped (via edge slides of Σ) so that S is aligned with T.*

Proof We will proceed by stem swaps, chosen so that they do not affect the hypothesis that $\mathfrak{R} \cap S = \varnothing$. Let $M_{\mathfrak{R}}$ be the component of $M - \mathfrak{R}$ that contains S, and $\mathfrak{R}_0 \subset \partial M_{\mathfrak{R}}$ the collection of sphere components that come from $\mathfrak{R}$. In $M_{\mathfrak{R}}$ each $\mathfrak{a} \in \mathfrak{R}_0$ is the blossom of a flower whose stem typically intersects S. (A nonseparating sphere in $\mathfrak{R}$ may appear twice in $\mathfrak{R}_0$, with one or both stems intersecting S.) Denote by $\hat{\sigma}$ the collection of all stems of $\mathfrak{R}_0$ that intersect S. The proof will be by induction on $|\hat{\sigma} \cap S|$. If $|\hat{\sigma} \cap S| = 0$ then either S is a sphere disjoint from Σ and therefore aligned, or S is a disk. In the latter case our convention of which compression body to call B has $\partial S \subset \partial_- B \subset \Sigma$, so $T \cap S$ is a single circle

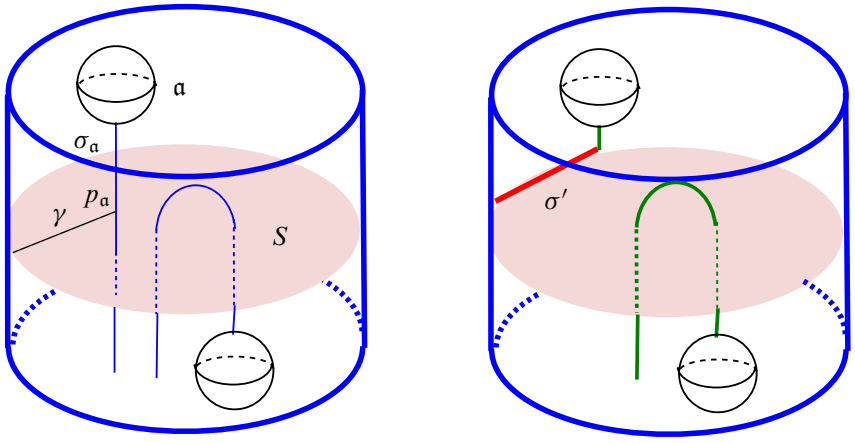

Figure 16: Swap lowering $|\hat\sigma \cap S|$, for S a disk.

parallel to ∂S in S. Again this means that S is aligned. Suppose then that $|\hat\sigma \cap S| > 0$ and inductively assume that the proposition is known to be true for lower values of $|\hat\sigma \cap S|$. Consider the possibilities:

Case 1 S is a disk.

Since $|\hat\sigma \cap S| > 0$ there is a blossom $𝔞 \in \mathfrak{R}_0$ with stem $\sigma \in \hat\sigma$. Let $\sigma_𝔞 \subset \sigma$ be the segment of $\sigma - S$ whose interior is disjoint from S and whose endpoints are the blossom $𝔞$ and a point $p_𝔞$ in S. Let γ be an arc in S that runs from $p_𝔞$ to ∂S that avoids all other points of $\hat\sigma \cap S$. Push the arc $\gamma \cup \sigma_𝔞$ off of S in the direction of $\sigma_𝔞$ so that it becomes a stem σ' for $𝔞$. Do a stem swap from σ to σ', and let Σ' be the result. See Figure 16. Since σ' is disjoint from S, σ is thereby removed from $\hat\sigma$, lowering $|\hat\sigma \cap S|$ by at least one. The stem swap does not affect other reducing edges or their edge-reducing spheres, so the latter remain disjoint from S. By Proposition 5.2 Σ' is still a spine of B, so T is isotopic in M to a regular neighborhood of Σ'. The inductive hypothesis implies that then T can be isotoped so that S is aligned with T, as required.

Case 2 S is a sphere.

Although S could be nonseparating in M, it cannot be nonseparating in $M_{\mathfrak{R}}$. Here is the argument: Suppose $S \subset M_{\mathfrak{R}}$ is nonseparating. If $\hat\sigma$ were disjoint from S then S would have no intersections with the Heegaard surface T at all and so $S \subset A$. But in a compression body such as A, all spheres separate, a contradiction. We will inductively reach the same contradiction by showing that if $\hat\sigma$ does intersect S there is a local stem swap that lowers $|\hat\sigma \cap S|$: Since S is nonseparating there is a circle c in $M_{\mathfrak{R}} - \Sigma$ that intersects S in a single point p. Let γ be a path in S from p to a point in $\sigma \cap S$, where $\sigma \in \hat\sigma$ and γ is chosen so that its interior is disjoint from $\hat\sigma$. Band sum σ to γ along a band perpendicular to S, with γ as its core. The result is an edge σ' that is obtained from σ by a local stem swap and intersects S in one fewer point than σ does, as required. See Figure 17.

So S is separating in $M_{\mathfrak{R}}$. This implies that no stem can intersect S more than once algebraically and so, following local stem swaps as in Claim 2 of Proposition 9.1 (see Figure 14), no more than once geometrically. If no stem intersects S at all, then $S \subset A$ and so S is aligned, finishing the proof.

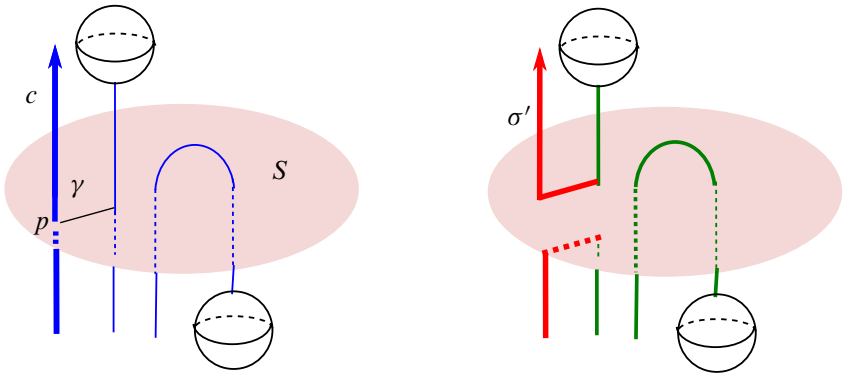

Figure 17: Swap lowering $|\hat{\sigma} \cap S|$, for S a nonseparating sphere.

Suppose, on the other hand, there is at least one stem $\sigma_!$ that intersects S exactly once. Repeat the argument of Case 1 for all stems other than $\sigma_!$, using the point $p_! = \sigma_i \cap S$ in place of ∂S in the argument. The result is that, after a sequence of stem swaps, all stems other than $\sigma_!$ are disjoint from S. This means that $S \cap \Sigma$ consists of the single point $p_!$. In other words, T intersects S in a single circle, and so S is aligned. □

The sequence of Proposition 4.2, Lemma 7.1, and Propositions 9.1 and 10.1 establishes Theorem 1.3. □

11 The Zupan example

Some time ago, Alex Zupan proposed a simple example for which the strong Haken theorem seemed unlikely (personal communication, 2019). The initial setting is of a Heegaard split 3–manifold $M = A \cup_T B$ that is the connected sum of compact manifolds M_1, M_2 and M_3, as shown in Figure 18. The blue indicates the spine Σ of B, say and, following our convention throughout the proof, B is to be thought of as a thin regular neighborhood of Σ. The spine is not shown inside of the punctured summands M_1 and M_2 because those parts are irrelevant to the argument; psychologically it's best to think of these as spherical boundary components of M lying in $\partial_- B$, so M_1 and M_2 are balls.

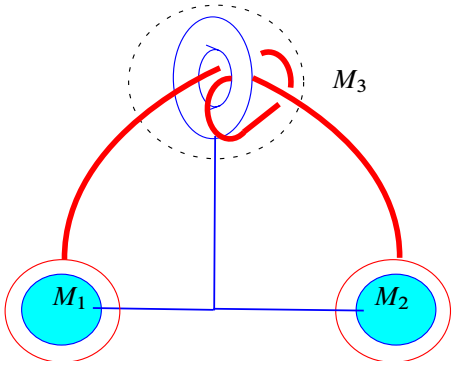

Figure 18: The initial setting.

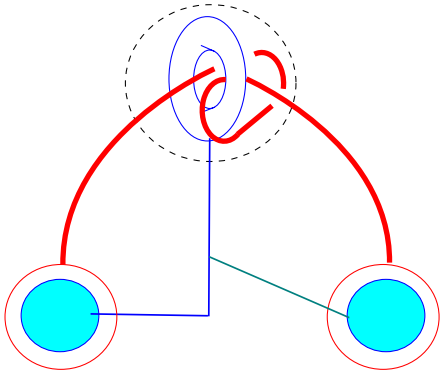

Figure 19: One blue edge now teal.

In the figure, M_3 is a solid torus and what we see is the punctured M_3, lying in M as a summand. We will continue the argument for this special case, in which M_3 is a solid torus and M_1 and M_2 are balls, but the argument works in general. An important role is played by the complement A of Σ outside M_1 and M_2. This is a solid torus: indeed, the region in the figure between the torus and the cyan balls is a twice punctured solid torus; A is obtained by removing both a collar of the torus boundary component and the blue arcs, all part of Σ. Removing the collar does not change the topology, but removing the blue arcs changes the twice-punctured solid torus into an unpunctured solid torus A.

Zupan proposed the following sort of reducing sphere S for M: the tube sum of the reducing spheres for M_1 and M_2 along a tube in M_3 which can be arbitrarily complicated. The outside of the tube is shown in red in Figure 18. The reducing sphere S is not aligned with T because it intersects Σ in two points, one near each of M_1 and M_2. The goal is then to isotope T through M so that it will be aligned with S. This is done by modifying Σ by what is ultimately a stem swap, and we will describe how the stem swap is obtained by an edge-slide of Σ. The edge-slide induces an isotopy of T in M because T is the boundary of a regular neighborhood of Σ. Note that in such an edge slide, passing one of the blue arcs through the red tube is perfectly legitimate.

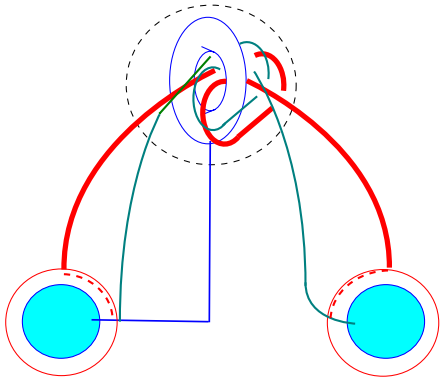

Figure 20: Teal edge now homotopic to red tube.

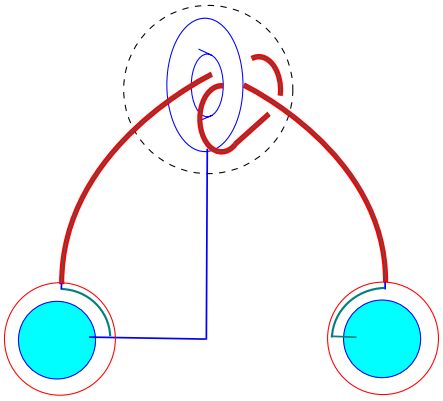

Figure 21: Teal edge isotoped into red tube.

Figure 19 is the same, but we have distinguished part of Σ (the rightmost edge) by turning it teal and beginning to slide it on the rest of the spine.

Now we invoke the viewpoint and notation of Proposition 5.2: There is a related Heegaard splitting of M available to us, in which the sphere boundary component at M_2 is not viewed as part of $\partial_- B$ but as part of $\partial_- A$, and the teal arc is also added to A. This changes A into a punctured solid torus A_+ and the spine of its complement into Σ_-, obtained by deleting from Σ both the teal edge and the sphere boundary component at M_2.

And so we apply Lemma 3.3, with A_+ playing the role of compression-body C; the boundary sphere at M_2 playing the role of the point r; the other end of the teal arc playing the role of q; the teal arc playing the role of β; and the union of the core of the red tube and the two dotted arcs in Figure 20 playing the role of α. Specifically, as the proof of Lemma 3.3 describes, because $\pi_1(\partial A_+) \to \pi_1(A_+)$ is surjective, and the slides take place in ∂A_+, one can slide the end of the teal arc around on the rest of Σ_- (technically on the boundary of a thin regular neighborhood of Σ_-) until it is *homotopic* rel endpoints to the path that is the union of the core of the tube of S and the two dotted red arcs shown in Figure 20. Hass and Thompson [5, Proposition 4] then shows that α and β are isotopic rel endpoints.

The result of the isotopy is shown in Figure 21; the teal edge now goes right through the tube, never intersecting S. Thus S now intersects Σ in only a single point, near the boundary sphere at M_1. In other words, S is aligned with T.

References

[1] **A J Casson**, **C M Gordon**, *Reducing Heegaard splittings*, Topology Appl. 27 (1987) 275–283 MR Zbl

[2] **M Freedman**, **M Scharlemann**, *Dehn's lemma for immersed loops*, Math. Res. Lett. 25 (2018) 1827–1836 MR Zbl

[3] **M Freedman**, **M Scharlemann**, *Uniqueness in Haken's theorem*, Michigan Math. J. 74 (2024) 119–142 MR

[4] **W Haken**, *Some results on surfaces in* 3*–manifolds*, from "Studies in modern topology" (P J Hilton, editor), Stud. Math. 5, Math. Assoc. Amer., Englewood Cliffs, NJ (1968) 39–98 MR Zbl

[5] **J Hass**, **A Thompson**, *Neon bulbs and the unknotting of arcs in manifolds*, J. Knot Theory Ramifications 6 (1997) 235–242 MR Zbl

[6] **S Hensel**, **J Schultens**, *Strong Haken via sphere complexes*, preprint (2021) arXiv 2102.09831

[7] **M Scharlemann**, *Heegaard splittings of compact* 3*–manifolds*, from "Handbook of geometric topology" (R J Daverman, R B Sher, editors), North-Holland, Amsterdam (2002) 921–953 MR Zbl

[8] **M Scharlemann**, *Generating the Goeritz group of* S^3, preprint (2020) arXiv 2011.10613

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

[10] **A Schoenflies**, *Beiträge zur Theorie der Punktmengen, III*, Math. Ann. 62 (1906) 286–328 MR Zbl

Department of Mathematics, University of California, Santa Barbara
Santa Barbara, CA, United States

mgscharl@math.ucsb.edu

Received: 8 April 2020 Revised: 1 August 2022

ALGEBRAIC & GEOMETRIC TOPOLOGY

msp.org/agt

See inside back cover or msp.org/agt for submission instructions.

The subscription price for 2024 is US $705/year for the electronic version, and $1040/year (+$70, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP. Algebraic & Geometric Topology is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

Algebraic & Geometric Topology (ISSN 1472-2747 printed, 1472-2739 electronic) is published 9 times per year and continuously online, by Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840. Periodical rate postage paid at Oakland, CA 94615-9651, and additional mailing offices. POSTMASTER: send address changes to Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840.

AGT peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

https://msp.org/

ALGEBRAIC & GEOMETRIC TOPOLOGY

Volume 24 Issue 2 (pages 595–1223) 2024