Algebraic & Geometric Topology 10 (2010) 2251–2275

Heegaard splittings with large subsurface distances

JESSE JOHNSON
YAIR MINSKY
YOAV MORIAH

We show that subsurfaces of a Heegaard surface for which the relative Hempel distance of the splitting is sufficiently high have to appear in any Heegaard surface of genus bounded by half that distance.

57M99; 57R99

1 Introduction

It was shown by Scharlemann and Tomova [20] that if a 3–manifold has a Heegaard surface Σ so that the Hempel distance $d(\Sigma)$ of the splitting [7] is greater than twice its genus $g(\Sigma)$, then any other Heegaard surface Λ of genus $g(\Lambda) < d(\Sigma)/2$ is a stabilization of Σ, ie, the result of attaching trivial handles to Σ. Hartshorn [5] had proved a similar result in which the surface Λ is incompressible.

In this paper, we generalize this theorem to the case where the Heegaard splitting Σ is not necessarily of high distance but has a proper essential subsurface $F \subset \Sigma$ so that the "subsurface distance" measured in F is large:

Theorem 1.1 *Let Σ be a Heegaard surface in a closed 3–manifold M of genus $g(\Sigma) \geq 2$, and let $F \subset \Sigma$ be a compact essential subsurface. Let Λ be another Heegaard surface for M of genus $g(\Lambda)$. If*

$$d_F(\Sigma) > 2g(\Lambda) + c(F),$$

then, up to ambient isotopy, the intersection $\Lambda \cap \Sigma$ contains F.

Here $c(F) = 0$ unless F is an annulus, 4–holed sphere, or 1– or 2–holed torus, in which case $c(F) = 2$.

Here, $d_F(\Sigma)$ is the distance between the subsurface projections of the disk sets of the handlebodies on the two sides of Σ to the "arc and curve complex" of F. For precise definitions see Section 2. This result can be paraphrased as follows: If the two disk sets

Published: 23 November 2010 DOI: 10.2140/agt.2010.10.2251

of a Heegaard splitting intersect on a subsurface of the Heegaard surface in a relatively "complicated" way, then any other Heegaard surface whose genus is not too large must contain that subsurface.

One can view this in the context of a number of results in 3–manifold topology and geometry where "local" complication, as measured in a subsurface, yields some persistent topological or geometric feature. An example with a combinatorial flavor is in Masur–Minsky [15; 16], where one studies certain quasi-geodesic paths in the space of markings on a surface. If the endpoints of such a path have projections to the arc and curve complex of a subsurface F which are sufficiently far apart, then the path is forced to have a long subinterval in which ∂F is a part of the markings.

In [18; 3], Brock, Canary and Minsky consider the geometry of hyperbolic structures on $\Sigma \times \mathbb{R}$, as controlled by their "ending invariants," which can be thought of as markings on Σ. A subsurface F for which the projections of the ending invariants are far apart yields a "wide" region in the manifold isotopic to $F \times [0,1]$, where the length of ∂F is very short.

A similar, but incompletely understood, set of phenomena can occur for Heegaard splittings of hyperbolic 3–manifolds, where the disk sets play the role of the ending invariants. This is the subject of work in progress by Brock, Namazi, Souto and others.

In the purely topological setting, Lackenby [14] studies Heegaard splittings of manifolds obtained by sufficiently complicated gluings along incompressible, acylindrical boundaries of genus at least 2. Bachman and Schleimer [1] obtain lower bounds on Heegaard genus in a surface bundle whose monodromy map is sufficiently complicated. Bachman, Schleimer and Sedgwick [2] study Heegaard genus for manifolds glued along tori by complicated maps.

The methods of our proof are extensions, via subsurface projections, of the work in Johnson [10], which itself builds on methods of Rubinstein and Scharlemann [19]. To a Heegaard splitting we associate a "sweepout" by parallel surfaces of the manifold minus a pair of spines, and given two surfaces Λ and Σ and associated sweepouts, we examine the way in which they interact. In particular under some natural genericity conditions we can assume that one of two situations occur:

(1) Λ F–*spans* Σ, or
(2) Λ F–*splits* Σ.

For precise definitions see Section 3. The case of F–spanning implies that, up to isotopy, there is a moment in the sweepout corresponding to Σ when the subsurface parallel to F lies in the upper half of $M \setminus \Lambda$, and a moment when it lies in the lower half. It then follows that Λ separates the product region between these two copies of F, and it follows fairly easily that it can be isotoped to contain F.

In the F–splitting case we are able to show that, for each moment in the sweepout corresponding to Σ, the level surface intersects Λ in curves that have essential intersection with F. By studying the way in which these intersections change during the course of the sweepout, we are able to use the topological complexity of Λ to control the subsurface distance of F.

Combining these two results and imposing the condition that $d_F(\Sigma)$ is greater than a suitable function of $g(\Lambda)$ forces the first, ie F–spanning, case to occur.

The discussion is complicated by some special cases, where F has particularly low complexity, in which the dichotomy between F–spanning and F–splitting doesn't quite hold. In those cases we obtain slightly different bounds.

Outline

In Section 2 we recall the definitions of Heegaard splittings, curve complexes and subsurface projections. In Section 3 we discuss sweepouts, pairs of sweepouts and the Rubinstein–Scharlemann graphic, and use this to define F–splitting and F–spanning and variations. In Lemma 3.7 we show that F–spanning and F–splitting are generically complementary conditions, and also work out the situation for the low-complexity cases.

In Section 4 we consider the F–spanning case, and show that it leads to the conclusion that Λ can be isotoped to contain F.

In Section 5 we give a more careful analysis of the local structure of nontransverse intersections of surfaces in a pair of sweepouts. In particular we use this to quantify the way in which intersection loops, and their projections to a subsurface F, can change as one moves the sweepout surfaces.

In Section 6 we show that the F–splitting condition leads to an upper bound on $d_F(\Sigma)$. The proof makes considerable use of the structure developed in Section 5, with a fair amount of attention being necessary to handle the exceptional low-complexity cases.

Finally in Section 7 we put these results together to obtain the proof of the main theorem.

Acknowledgements The authors are grateful to Saul Schleimer for pointing out an error in the original proof of Lemma 2.1.

The first author was partially supported by NSF MSPRF grant 0602368. The second author was partially supported by NSF grant 0504019. The third author would like to thank Yale University for its hospitality during a sabbatical in which the research was done.

2 Heegaard splittings and curve complexes

Heegaard splittings

A handlebody H is a 3–manifold homeomorphic to a regular neighborhood of a finite, connected graph in S^3. The graph will be called a *spine* of the handlebody. A Heegaard splitting of genus $g \geq 2$ for a closed 3–manifold M is a triple (Σ, H^-, H^+), where H^+, H^- are genus g handlebodies and $\Sigma = \partial H^+ = \partial H^- = H^+ \cap H^-$ is the Heegaard surface so that $M = H^+ \cup_\Sigma H^-$.

Curve complexes and Hempel distance

For any surface Σ there is an associated simplicial complex called the *curve complex* and denoted by $\mathcal{C}(\Sigma)$. An i–simplex in $\mathcal{C}(\Sigma)$ is a collection $([\gamma_0], \dots, [\gamma_i])$ of isotopy classes of mutually disjoint essential simple closed curves in Σ. On the 1–skeleton $\mathcal{C}^1(\Sigma)$ of the curve complex $\mathcal{C}(\Sigma)$ there is a natural path metric d defined by assigning length 1 to every edge. The subcollection $\mathcal{D}(H^\pm)$, of isotopy classes of curves in Σ, that bound disks in $H^\pm$ (also called *meridians*) is called the *handlebody set* associated with $H^\pm$ respectively.

Given a Heegaard splitting (Σ, H^-, H^+) we define the *Hempel distance* $d(\Sigma)$:

$$d(\Sigma) = d_{\mathcal{C}^1(\Sigma)}(\mathcal{D}(H^+), \mathcal{D}(H^-))$$

where distance between sets is always minimal distance.

When a surface F has boundary we can define the *arc and curve complex* $\mathcal{AC}(F)$, by considering isotopy classes of essential (nonperipheral) simple closed curves and properly embedded arcs. If F is an annulus, there are no essential closed curves, and the isotopy classes of essential arcs should be taken rel endpoints. As before an n–simplex is a collection of $n+1$ isotopy classes with disjoint representative loops/arcs. We denote by d_F the path metric on the 1–skeleton of $\mathcal{AC}^1(F)$ that assigns length 1 to every edge.

If F is a connected proper essential subsurface in Σ, there is a map

$$\pi_F \colon \mathcal{C}^0(\Sigma) \to \mathcal{AC}^0(F) \cup \{\varnothing\}$$

defined as follows (see Masur–Minsky [16] and Ivanov [8; 9]): First assume F is not an annulus. Given a simple closed curve γ in Σ, isotope it to intersect ∂F minimally. If the intersection is empty let $\pi_F(\gamma) = \varnothing$. Otherwise consider the isotopy classes of components of $F \cap \gamma$ as a simplex in $\mathcal{AC}(F)$, and select (arbitrarily) one vertex to be $\pi_F(\gamma)$. If F is an annulus, let $\widehat{F} \to \Sigma$ be the annular cover associated to F, compactified naturally using the boundary at infinity of $\mathbb{H}^2$, lift γ to a collection of disjoint properly embedded arcs in $\widehat{F}$, and arbitrarily select an essential one (which must exist if γ crosses F essentially) to be $\pi_F(\gamma)$. If γ does not cross F essentially we again let $\pi_F(\gamma) = \varnothing$. By abuse of notation we identify $\mathcal{AC}(\widehat{F})$ with $\mathcal{AC}(F)$.

Now if (Σ, H^-, H^+) is a Heegaard splitting of a 3–manifold M and $F \subset \Sigma$ is a connected proper essential subsurface, let $\mathcal{D}_F(H^-) = \pi_F(\mathcal{D}(H^-))$ and $\mathcal{D}_F(H^+) = \pi_F(\mathcal{D}(H^+))$ – the projections to F of all loops that bound essential disks in H^- and H^+, respectively. The *F–distance* of Σ, which we will write $d_F(\Sigma)$, is the distance between these two sets,

$$d_F(\Sigma) = d_{\mathcal{AC}^1(F)}(\mathcal{D}_F(H^+), \mathcal{D}_F(H^-)).$$

We will have use for the following fact, which is a variation on a result of Masur and Schleimer:

Lemma 2.1 *Let Σ be the boundary of a handlebody H of genus $g \geq 2$. Let $F \subset \Sigma$ be an essential connected subsurface of Σ. If $\Sigma \smallsetminus F$ is compressible in H then $\pi_F(\mathcal{D}(H))$ comes within distance 2 of every vertex of $\mathcal{AC}(F)$, provided F is not a 4–holed sphere. If it is a 4–holed sphere the distance bound is 3.*

Proof If F is a 3–holed sphere the diameter of $\mathcal{AC}(F)$ is 1, hence the lemma follows trivially. We next give the proof in the nonannular case. We claim that $\Sigma \smallsetminus F$ must contain a meridian $\delta \in \mathcal{D}(H)$ such that F can be connected by a path to either side of δ: Compressibility of $\Sigma \smallsetminus F$ yields some meridian, which has the desired property if it is nonseparating in Σ. If it is separating, then on the side complementary to F we can find a nonseparating meridian.

Let γ_1 and γ_2 be components of ∂F which can be connected by paths to the two sides of δ (possibly $\gamma_1 = \gamma_2$). If α is any essential embedded arc in F with endpoints on γ_1 and γ_2 then α can be extended to an embedded arc meeting δ on opposite sides. A band sum between two copies of δ along this arc yields a new meridian δ'.

If δ is not isotopic to either γ_1 or γ_2, then the essential intersection of δ' with F is two copies of the arc α, so that $\pi_F(\delta') = \alpha$. If it is isotopic to γ_1 or γ_2 (or both), then $\pi_F(\delta')$ is an arc or closed curve contained in a regular neighborhood of $\alpha \cup \gamma_1 \cup \gamma_2$.

We now need to show that for every vertex $\beta \in \mathcal{AC}(F)$, there is a path of length at most 2 from β to $\pi_F(\delta')$, for some choice of α.

Suppose first that $\gamma_1 = \gamma_2$. In this case δ cannot be isotopic to γ_1, because γ_1 cannot be connected to both sides of δ without going through another boundary component of F. Hence in this case $\pi_F(\delta') = \alpha$. Since F is not an annulus, cutting along β gives a surface F_β no component of which is a disk. If the component of F_β containing γ_1 is not an annulus then an essential arc α disjoint from β and with endpoints on γ_1 exists in F. This gives $d_{\mathcal{AC}(F)}(\beta, \pi_F(\mathcal{D}(H))) \le 1$. If the component of F_β containing γ_1 is an annulus let β' be an arc or curve in the other component. Apply the previous case to β' to obtain $d_{\mathcal{AC}(F)}(\beta, \pi_F(\mathcal{D}(H))) \le 2$.

Suppose now $\gamma_1 \neq \gamma_2$. If β is disjoint from γ_1 and γ_2 and doesn't separate them, then we can choose α connecting γ_1 and γ_2 and disjoint from β, so $d_{\mathcal{AC}(F)}(\beta, \pi_F(\delta')) \le 1$.

If β does separate γ_1 from γ_2, let X_1 and X_2 be the components of $F \setminus \beta$. If one of them is nonplanar, it contains a nonseparating closed curve β', therefore $d_{\mathcal{AC}(F)}(\beta', \pi_F(\delta')) \le 1$ and hence $d_{\mathcal{AC}(F)}(\beta, \pi_F(\delta')) \le 2$.

If both X_i are planar, and F is not a 4–holed sphere, then at least one, say X_1, has two boundary components other than β and γ_1 or γ_2. An arc β' connecting those boundary components is disjoint from γ_1 and γ_2 and does not separate them, so again we are done using the previous cases.

Now suppose β is an arc with endpoint on γ_1 or γ_2, or both. If it has endpoints on both, then choose $\alpha = \beta$, and note that β is disjoint from $\pi_F(\delta')$ in all cases. If β meets just γ_1, say, let α be an arc connecting γ_1 and γ_2 and disjoint from β, and consider a regular neighborhood N of $\gamma_1 \cup \gamma_2 \cup \alpha \cup \beta$. Then $\partial N \setminus \partial F$ is either an arc or a pair of closed curves which is disjoint from $\beta, \gamma_1, \gamma_2$ and α. The arc, and at least one of the closed curves, is essential as long as F is not a 4–holed sphere. Let β' be this curve or arc, and again we see that the distance from β to $\pi_F(\delta')$, via β', is 2.

We have yet to handle the case that F is a 4–holed sphere, and β separates γ_1 from γ_2. Choose α to be an arc intersecting β exactly once. In this case, $\pi_F(\delta')$ is either α or a closed curve separating $\gamma_1 \cup \gamma_2$ from the other two boundary components, and one can check that $d_{\mathcal{AC}(F)}(\beta, \pi_F(\delta')) \le 3$.

Finally we consider the case that F is an annulus. As before there is a meridian δ that separates the boundaries of F in $\Sigma \setminus F$. We may assume that δ is not isotopic to ∂F, because if it were then its complement would contain other meridians disjoint from F. Let α' be an embedded arc connecting the two sides of δ and passing essentially through F, and let δ' be the result of the band sum, as before.

Let $\widehat{F} \to \Sigma$ be the compactified annular cover associated to F. There is a lift $\widehat{\alpha}$ of α' to $\widehat{F}$ connecting lifts $\widehat{\delta}_1$ and $\widehat{\delta}_2$ of δ which meet opposite sides of $\partial\widehat{F}$. There is then a lift $\widehat{\delta}'$ of δ' disjoint from $\widehat{\delta}_1 \cup \widehat{\alpha} \cup \widehat{\delta}_2$, and connecting opposite sides of $\widehat{F}$. This is a representative of $\pi_F(\delta')$.

A Dehn twist of α' around F has the following effect on $\widehat{\delta}'$: it performs a single Dehn twist about the core of $\widehat{F}$, as well as a homotopy of the endpoints of $\widehat{\delta}'$ which stays outside of the disks D_i cobounded by $\widehat{\delta}_i$ and $\partial\widehat{F}$ $(i = 1, 2)$.

Now let β be an arc representing a vertex of $\mathcal{AC}(\widehat{F})$. There exists a disjoint arc β' whose endpoints lie in D_1 and D_2. It follows that after n Dehn twists on α' we obtain an arc α_n whose associated δ'_n has lift $\widehat{\delta}'_n$ which is disjoint from β'. We conclude that $d_{\mathcal{AC}(F)}(\beta, \pi_F(\delta'_n) \leq 2$. □

3 Sweepout pairs

In this section we discuss *sweepouts* of a 3–manifold representing a Heegaard splitting, and consider the interaction of pairs of sweepouts using the *Rubinstein–Scharlemann graphic* and generalizations of the notions of *mostly above* and *mostly below* from Johnson [10]. We will formulate relative versions of these which allow us to consider subsurfaces, and from this develop a relative version of the *spanning* and *splitting* relations from [10].

Sweepouts

Given a closed 3–manifold M, a *sweepout* of M is a smooth function $f\colon M \to [-1,1]$ so that each $t \in (-1,1)$ is a regular value, and the level set $f^{-1}(t)$ is a Heegaard surface. Furthermore each of the sets $\Gamma^+ = f^{-1}(1)$ and set $\Gamma^- = f^{-1}(-1)$ are spines of the respective handlebodies. When this happens we say the sweepout *represents* the Heegaard splitting associated to each level surface. It is clear (with a bit of attention to smoothness at the spines) that every Heegaard splitting can be represented this way.

Two sweepouts f and h of M determine a smooth function $f \times h\colon M \to [-1,1] \times [-1,1]$. The differential $D(f \times h)$ has rank 2 (or dim $Ker(D(f \times h)) = 1$) wherever the level sets of f and h are transverse. Thus we define the *discriminant set* Δ to be the set of points of M for which dim $Ker(D(f \times h)) > 1$. The discriminant, and its image in $[-1,1] \times [-1,1]$, therefore encode the configuration of tangencies of the level sets of f and h.

A smooth function $\varphi\colon M \to N$ between smooth manifolds M, N is *stable* if there is a neighborhood U of φ in $C^\infty(M,N)$ such that any map $\eta \in U$ is *isotopic* to φ through a family of maps in U. (We say that η and φ are isotopic when there are diffeomorphisms $\beta\colon M \to M$ and $\alpha\colon N \to N$, isotopic to the identity, such that $\alpha \circ \eta \circ \beta = \varphi$.) Here the topology on $C^\infty(M,N)$ is the Whitney topology, also known as the strong topology on C^∞. This topology differs from the weak (compact-open) topology on C^∞ for noncompact domains, which is relevant here since we consider stability on the complement of the spines.

Kobayashi and Saeki [13] show that, after isotopies of f and h, one can arrange that $f \times h$ is stable on the complements of the four spines. When that holds, the kernel of $D(f \times h)$ (off the spines) is always of dimension at most 2 and it follows from Mather [17] that Δ is a smooth manifold of dimension one.

The image $(f \times h)(\Delta)$ is a graph in $[-1,1] \times [-1,1]$ with smooth edges, called the *graphic*, or the *Rubinstein–Scharlemann graphic* (see Rubinstein–Scharlemann [19]).

We call $f \times h$ *generic* if it is stable away from the spines and each arc $[-1,1] \times \{s\}$ in the square intersects at most one vertex of the graphic. The following lemma of Kobayashi and Saeki [13] justifies this term:

Lemma 3.1 *Any pair of sweepouts can be isotoped to be generic.*

Suppose therefore that $f \times h$ is generic. Points in the square are denoted by (t,s), and we define the surfaces $\Lambda_s = h^{-1}(s)$ and $\Sigma_t = f^{-1}(t)$. If the vertical line $\{t\} \times [-1,1]$ meets no vertices of the graphic, then $h|_{\Sigma_t}$ is Morse, and its critical points are $\Sigma_t \cap \Delta$. In particular the fact that a Morse function has at most one singularity at any level corresponds to the fact that the map from Δ to the graphic is one-to-one over the smooth points of the graphic. If $v = (t_0, s_0)$ is a vertex, we see certain transitions as t passes through t_0.

(1) *Cancelling pair*: If v has valence 2, then the edges of the graphic incident to v are either both contained to the left of $\{t_0\} \times (0,1)$, or both to the right of it. As t passes through t_0, a pair of singularities of $h|_{\Sigma_t}$, one saddle and one min or max (which we call central), is created or annihilated.

(2) *Simultaneous singularities*: If v has valence 4, it is the intersection of the images of two disjoint arcs of Δ. Hence there are two singularities whose relative heights are exchanged as t passes through t_0. These can be either saddle or central singularities. Note that the singularities cannot coalesce, for example in a monkey saddle, as they pass each other, since that would produce a vertex in the discriminant set Δ, contradicting the fact that it is a smooth 1–manifold.

Vertices on the boundary of the square can have valence either one or two; we will however not need to consider these.

Above and below

Let f and h be sweepouts for a manifold M, representing the Heegaard splittings (Σ, H^-, H^+) and (Λ, V^-, V^+), respectively. For each $s \in (-1,1)$, define $V_s^- = h^{-1}([-1,s])$ and $V_s^+ = h^{-1}([s,1])$. Note that $\Lambda_s = \partial V_s^- = \partial V_s^+$ Similarly, for $t \in (-1,1)$, $H_t^- = h^{-1}([-1,t])$, $H_t^+ = h^{-1}([t,1])$ and $\Sigma_t = \partial H_t^- = \partial H_t^+$.

Throughout the rest of the paper, let $F \subset \Sigma$ denote a compact, connected, essential subsurface, where “essential” means that no boundary component of F bounds a disk. We exclude 3–holed spheres. Let F_t be the image of F under the identification of Σ with Σ_t determined up to isotopy by the sweepout.

Definition 3.2 We will say that Σ_t is *mostly above* Λ_s *with respect to* F, denoted

$$\Sigma_t \succ_F \Lambda_s,$$

if $\Sigma_t \cap V_s^-$ is contained in a subsurface of Σ_t that is isotopic into the complement of F_t (or is just contained in a disk, when $F = \Sigma$). Similarly, Σ_t is *mostly below* Λ_s *with respect to* F, or

$$\Sigma_t \prec_F \Lambda_s,$$

if $\Sigma_t \cap V_s^+$ is contained in a subsurface that is isotopic into the complement of F_t (or contained in a disk).

The case that $F = \Sigma$ corresponds to the notion of *mostly above* and *mostly below* from [10]. We will mostly be concerned with the case that F is a proper subsurface.

Define R_a^F (respectively R_b^F) in $(-1,1) \times (-1,1)$ to be the set of all values (t,s) such that $\Sigma_t \succ_F \Lambda_s$ (respectively $\Sigma_t \prec_F \Lambda_s$). When $F = \Sigma$ these are the sets R_a and R_b of [10].

Figure 1 illustrates some typical configurations of R_a^F and R_b^F. Their basic properties are described in the following lemma.

Lemma 3.3 *Let $f \times h$ be a generic sweepout pair, and let F be an essential subsurface of Σ. If F is an annulus assume it is not isotopic into Λ.*

The sets R_a^F and R_b^F are disjoint, open and bounded by arcs of the graphic, so that all interior vertices appearing in ∂R_a^F or ∂R_b^F have valence 4.

Moreover R_a^F and R_b^F intersect each vertical line in a pair of intervals as follows: For each $t \in (-1,1)$, there are values $x \le y \in (-1,1)$ such that

$$(t,s) \in R_a^F \iff s \in (-1,x)$$

and

$$(t,s) \in R_b^F \iff s \in (y,1).$$

Proof We assume F is a proper subsurface of S (the case $F = \Sigma$ was done in Johnson [10], and our argument is a direct generalization). Openness of R_a^F and R_b^F is clear from the definition. If the sets intersect then we may select a point (t,s) in the intersection which is not in the graphic, so that $\Sigma_t \cap \Lambda_s$ is a 1–manifold. This

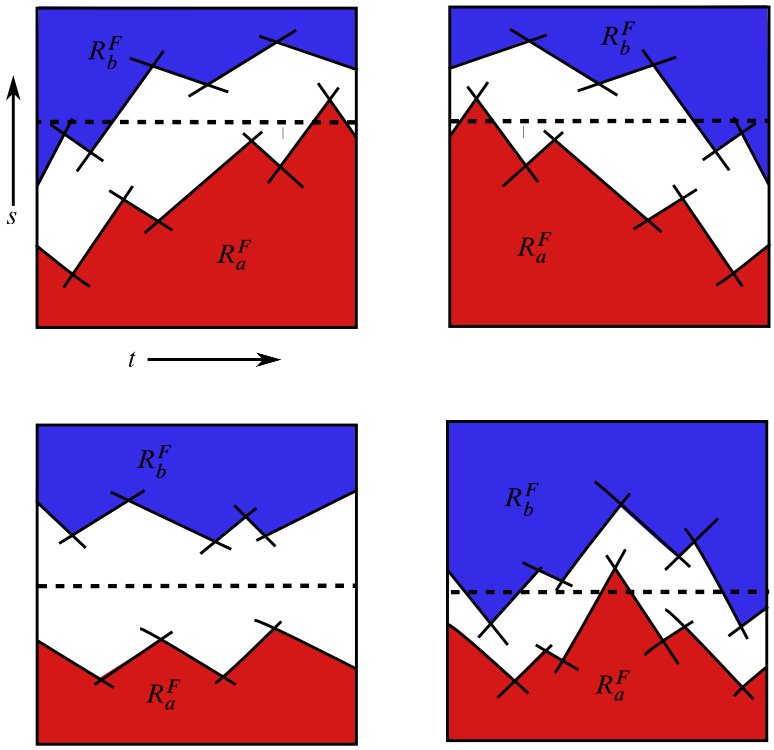

Figure 1: Four partial graphics of pairs $f \times h$ of sweepouts. In the top two graphics, h F–spans f. In the bottom left, h F–splits f. In the bottom right, h F–spans f again but the switch between mostly above and mostly below happens twice along the horizontal arc shown.

1–manifold divides Σ_t into two subsurfaces $X_- = \Sigma_t \cap V_s^-$ and $X_+ = \Sigma_t \cap V_s^+$, each of which can be isotoped off F_t. But this is impossible: If X_- can be isotoped off F_t, then after isotopy, X_+ contains F_t and hence can't be isotoped away from F_t, unless F_t is an annulus. If F_t is an annulus, this means it is isotopic to a neighborhood of the common boundary between X_+ and X_-, which is isotopic into Λ_s contradicting the assumption. We conclude that R_a^F and R_b^F are disjoint.

Now let $(t,s) \in (0,1) \times (0,1)$ be any point which is not on the graphic. Again Σ_t and Λ_s are transverse, and hence a small perturbation of (t,s) would yield an isotopic intersection pattern. In particular a small neighborhood of (t,s) is either contained in R_a^F or in its complement (and similarly for R_b^F). It follows that the boundary must be contained in the graphic.

Let $(t,s) \in R_a^F$. Then $\Sigma_t \cap V_s^-$ can be isotoped out of F_t. Since V_s^- monotonically increases with s, the same must hold for each $s' < s$. Hence the set $\{s : (t,s) \in R_a^F\}$

must, if nonempty, be an interval of the form $(-1, x)$. It is nonempty because, for small enough s, V_s^- is a small regular neighborhood of a spine and so intersects Σ_t in a union of disks. Similarly we find that $\{s : (t, s) \in R_b^F\}$ has the form $(y, 1)$. Finally, $x \le y$ follows from the disjointness of R_a^F and R_b^F.

It remains to show that only 4–valent vertices appear in the boundary of R_a^F and R_b^F. Suppose v is a 2–valent vertex in ∂R_a^F, say. Then there is a neighborhood U of v which intersects the graphic in a pair of arcs incident to v, which lie on one side of the vertical line through v. Such a pair separates U into two pieces, and hence is equal to $\partial R_a^F \cap U$. This contradicts what we have just proven about the intersection of R_a^F with vertical lines. □

Relative spanning and splitting

The relations of *spanning* and *splitting* were introduced in [10] for pairs of sweepouts. Here we will extend this notion to spanning and splitting relative to a subsurface F.

Definition 3.4 We say h *F–spans* f if there is a horizontal arc $[-1, 1] \times \{s\}$ in $(-1, 1) \times (-1, 1)$ that intersects both R_a^F and R_b^F.

In other words, h F–spans f if there are values $s, t_-, t_+ \in [-1, 1]$ such that $\Sigma_{t_-} \prec_F \Lambda_s$ while $\Sigma_{t_+} \succ_F \Lambda_s$.

Figure 1 also shows examples of graphics of pairs of sweepouts that don't span, or that span with two distinct such arcs.

The complementary situation is the following:

Definition 3.5 We say h *weakly F–splits* f if there is no horizontal arc $[-1, 1] \times \{s\}$ that meets both R_a^F and R_b^F.

A somewhat stronger condition which under some circumstances is equivalent, is the following:

Definition 3.6 We say h *F–splits* f if, for some $\{s\} \in (-1, 1)$, the arc $[-1, 1] \times \{s\}$ is disjoint from the closures of both R_a^F and R_b^F.

We extend these notions to pairs of Heegaard splittings with no sweepout specified: That is, if Λ and Σ are Heegaard surfaces and there exist some sweepouts h and f representing Λ and Σ, respectively, such that $f \times h$ is generic and h F–splits f for some $F \subset \Sigma$, we say that Λ F–splits Σ; and similarly for F–spanning and weak F–splitting.

Dichotomy

By definition F–spanning and weak F–splitting are complementary conditions. Generically, F–spanning and F–splitting are complementary as well, ie weak F–splitting implies F–splitting. The proof proceeds along the lines of [10], except when F is one of a short list of low-complexity surfaces, when weak F–splitting becomes a distinct case:

Lemma 3.7 *Let f and h be sweepouts for M, and F a connected, essential subsurface of the Heegaard surface Σ represented by f. Suppose that $f \times h$ is generic.*

If F is not an annulus, 4–holed sphere or a 1–holed or 2–holed torus, then

$$h \text{ weakly } F\text{–splits } f \quad \Longrightarrow \quad h \ F\text{–splits } f.$$

Equivalently, either h F–spans f, or h F–splits f.

In the exceptional cases, if h weakly F–splits f but does not F–split, then there exists a unique horizontal line $[-1,1] \times \{s\}$ which meets both closures $\overline{R_a^F}$ and $\overline{R_b^F}$ in a single point, which is a vertex of the graphic.

Proof By Lemma 3.3 the set of heights of horizontal lines meeting R_a^F has the form $(-1,x)$ and the set of heights of horizontal lines meeting R_b^F has the form $(y,1)$. If h does not F–span f, then these sets are disjoint so $x \le y$. If $x < y$ then any line with height $s \in (x,y)$ misses both $\overline{R_a^F}$ and $\overline{R_b^F}$, so that h F–splits f, and we are done.

If $x = y$ then the line $[-1,1] \times \{x\}$ meets a maximum-height point of $\overline{R_a^F}$ as well as a minimum-height point $\overline{R_b^F}$. Such a point lies on the graphic by Lemma 3.3 and it must be a vertex, for the only other possibility is an interior horizontal tangency of the graphic. Such a tangency corresponds to a critical point of h away from the spine, which is ruled out by definition of a sweepout (see [11] for details).

Since $f \times h$ is generic, a horizontal line can only meet one vertex, so the maximum of $\overline{R_a^F}$ and the minimum of $\overline{R_b^F}$ are the same vertex (t,x) of the graphic. It remains to show that this can only happen in the given exceptional cases.

The fact that (t,x) is a vertex (of valence 4 by Lemma 3.3) means that the function $h_t \equiv h|_{\Sigma_t}$ has two critical points on $\Gamma \equiv h_t^{-1}(x)$. Choose $x_- < x_+$ so that x is the unique critical value in $[x_-, x_+]$, and let

$$Q = h_t^{-1}([x_-, x_+]),$$

which is a regular neighborhood of Γ in Σ_t. Since $(t, x_-) \in R_a^F$, the subsurface $Y_- \equiv h_t^{-1}([-1, x_-])$ can be isotoped out of F_t. Since $(t, x_+) \in R_b^F$, the subsurface $Y_+ \equiv h_t^{-1}([x_+, 1])$ can also be isotoped out of F_t. Equivalently F_t can be isotoped out of each of Y_- and Y_+, and it follows that this can be done simultaneously (this is a general property of surfaces). Therefore after this isotopy we find that $F_t \subset Q$.

However Q cannot be very complicated – since $h_t^{-1}(x)$ has two critical points, Q is a disjoint union of annuli, disks, 3–holed spheres and possibly one component that is a 4–holed sphere or 2–holed torus. (The detailed analysis of possibilities for Q, which we will need in Section 6, will be done in Section 5.) The only essential subsurfaces of Σ_t that can embed in Q are therefore annuli, 3–holed spheres, 4–holed spheres, and 1– and 2–holed tori. This takes care of all the exceptional cases. □

4 Spanning implies coherence

In this section we consider the F–spanning case, for which we can show that the surface F is isotopic into Λ.

Proposition 4.1 *Let Σ and Λ be Heegaard surfaces in a closed 3–manifold M and let $F \subset \Sigma$ be a proper connected essential subsurface. If Λ F–spans Σ then after isotoping Λ we obtain a surface whose intersection with Σ contains F.*

Proof Let f and h be sweepouts representing Σ and Λ, respectively, such that h F–spans f and $f \times h$ is generic. Hence there is a level surface Λ_s and values $t_-, t_+ \in (-1,1)$ such that $\Sigma_{t_-} \prec_F \Lambda_s$ and $\Sigma_{t_+} \succ_F \Lambda_s$. We may assume without loss of generality that $t_- < t_+$.

We may identify the complement of the spines of f, namely $f^{-1}((-1,1))$, with $\Sigma \times (-1,1)$ in a way that is unique up to level-preserving isotopy. Consider the submanifold $F \times J \subset M$, where $J = [t_-, t_+]$. Because $\Sigma_{t_-} \prec_F \Lambda_s$, the handlebody $V_s^+ = h^{-1}([s,1])$ intersects $\Sigma \times \{t_-\}$ in a set that can be isotoped outside of $F \times \{t_-\}$; equivalently, the subsurface $F \times \{t_-\}$ can be isotoped within $\Sigma \times \{t_-\}$ so that it is contained in V_s^-. Similarly, since $\Sigma_{t_+} \succ_F \Lambda_s$, we can isotope the surface $F \times \{t_+\}$ inside $\Sigma \times \{t_+\}$ so that it is contained in V_s^+. After a level-preserving isotopy of $\Sigma \times (-1,1)$, we may therefore assume that $F \times \{t_-\}$ and $F \times \{t_+\}$ are contained in V_s^- and V_s^+, respectively. The surface $S = \Lambda_s \cap F \times J$ therefore separates $F \times \{t_-\}$ from $F \times \{t_+\}$ within $F \times J$, since Λ_s separates V_s^- from V_s^+.

Note that a surface in $F \times J$ with boundary in $\partial F \times J$ separates $F \times \{t_-\}$ from $F \times \{t_+\}$ if and only if its homology class in $H_2(F \times J, \partial F \times J)$ is nonzero.

Compressing S if necessary, we obtain an incompressible surface in the same homology class. Let S' be a connected component which is still nonzero in $H_2(F \times J, \partial F \times J)$. Then S' separates $F \times \{t_-\}$ from $F \times \{t_+\}$ and hence (up to orientation) must be homologous to $F\times\{t_-\}$. The vertical projection of S' to $F\times\{t_-\}$ is therefore a proper map which is π_1–injective and of degree ±1. By a covering argument it must also be π_1–surjective, and hence Theorem 10.2 of [6] implies that S' is isotopic to a level surface $F\times\{t\}$.

Making the isotopy ambient and keeping track of the 1–handles corresponding to the compressions that gave us S', we can obtain an isotopic copy of Λ which contains $F\times\{t\}$ minus attaching disks for the 1–handles. Now we can slide these disks outside of $F\times\{t\}$. Hence Λ itself is isotopic to a surface containing $F\times\{t\}$, as claimed. □

5 Saddle transitions

In this section we examine more carefully the intersections $\Sigma_t\cap\Lambda_s$, and the relationship between their regular neighborhoods in the two surfaces.

Fix (t,s) for the rest of this section. The interesting case is when (t,s) lies in the graphic, and hence the surfaces are not transversal, or equivalently, t is a critical value of $f|_{\Lambda_s}$ and s is a critical value of $h|_{\Sigma_t}$. As in the proof of Lemma 3.7, take an interval $[s_-,s_+]$ in which s is the only critical value of $h|_{\Sigma_t}$ (if any), and let

$$Q=(h|_{\Sigma_t})^{-1}([s_-,s_+])\subset\Sigma_t.$$

Similarly choose an interval $[t_-,t_+]$ containing t with no other critical values of $f|_{\Lambda_s}$, and define

$$Z=(f|_{\Lambda_s})^{-1}([t_-,t_+])\subset\Lambda_s.$$

Then Q is a regular neighborhood of $\Gamma\equiv\Sigma_t\cap\Lambda_s$ in Σ_t, and Z is a regular neighborhood of Γ in Λ_s.

To compare Q and Z, fix an identification of $f^{-1}((-1,1))$ with $\Sigma\times(-1,1)$ such that f becomes projection to the second factor, and let

$$p\colon f^{-1}((-1,1))\to\Sigma$$

denote the projection to the first factor. For convenience identify Σ with Σ_t so that $p\colon\Sigma_t\to\Sigma$ can be taken to be the identity. Now we can compare $p(Z)$ to Q within Σ.

The trivial case is that of a component of Γ that contains no critical points; in this case the components of Q and Z that retract to it are both annuli and the restriction of p is homotopic to a diffeomorphism between them. Let us record this observation:

Lemma 5.1 *Two level curves of $f|_\Lambda$ which are not separated by critical points have homotopic p–images in Σ.* □

Another simple case is of a component of Γ which is an isolated (hence central) singularity, and the corresponding components of Q and Z are disks.

If Γ_0 is a component containing one saddle singularity, then it is a figure–8 graph, and both associated components $Q_0 \subset Q$ and $Z_0 \subset Z$ must be 3–holed spheres, or pairs of pants. In this case we have:

Lemma 5.2 *If Γ_0 is a figure–8 component of Γ then $p|_{Z_0}$ is homotopic to a diffeomorphism $Z_0 \to Q_0$. In particular $p(\partial Z_0)$ is homotopic to a collection of simple disjoint curves in Σ.*

Recall that for any point of the graphic that is not a vertex, Γ contains exactly one singularity, so if it is a saddle we have exactly one such figure–8 component of Γ.

Proof Because each boundary curve of Z_0 lies on a level surface Σ_{t_-} or Σ_{t_+}, it is embedded by p. On the other hand it is homotopic to an essential curve in Γ_0, and since p is continuous the same is true for its image in Σ. The only essential simple curves in a 3–holed sphere are parallel to boundary components, so $p(\partial Z_0)$ is homotopic to ∂Q_0. The lemma follows. □

The most complicated case is that of a component $\Gamma_0 \subset \Gamma$ that contains two saddle singularities. There are only two possible isomorphism types for Γ_0 as a graph with two vertices of valence 4, and a total of three ways that Γ_0 can embed as a *level set* of a function with nondegenerate singularities. These are indicated in Figure 2, which applies to both Q_0 and Z_0.

To understand how p looks in this case, note first that at the vertices of Γ_0 the surfaces are tangent so p is a local diffeomorphism. Hence it either preserves or reverses orientation at these points. This determines the local picture of Q_0 and Z_0 at each tangency, and the rest of the configuration is determined by how Z_0 attaches along the edges of Γ_0. Either the two orientations match or they do not. When they match, one sees that $p|_{Z_0}$ is homotopic to a diffeomorphism – we call this the untwisted case. If they do not match – the twisted case – then in fact Q_0 and Z_0 may not even be diffeomorphic. The three twisted cases are depicted in Figure 3, where in the first case Q_0 is a 4–holed sphere and Z_0 a 2–holed torus, in the middle case they change roles, and in the third case both are 4–holed spheres but p is not homotopic to a diffeomorphism.

Note as in Lemma 5.2 the boundary curves of Z_0 are mapped to simple curves; however in the twisted case their images are not disjoint. Their intersections are prescribed by Figure 3, and we wish to record something about how they look from the point of

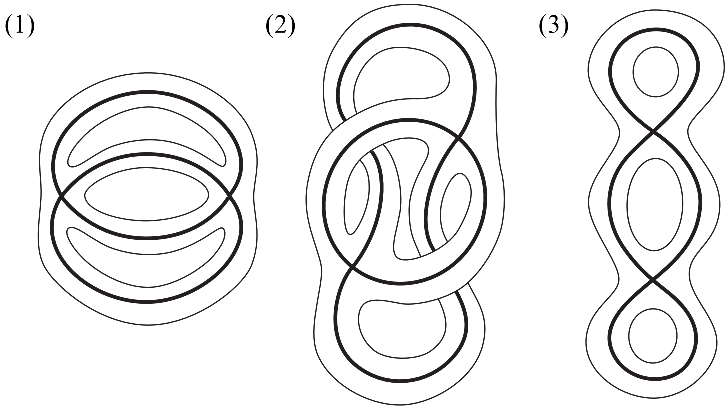

Figure 2: The three possible types for Q_0 or Z_0, depicted as immersed surfaces in the plane with the level set Γ_0 in heavy lines. The first and third are 4–holed spheres and the second is a 2–holed torus.

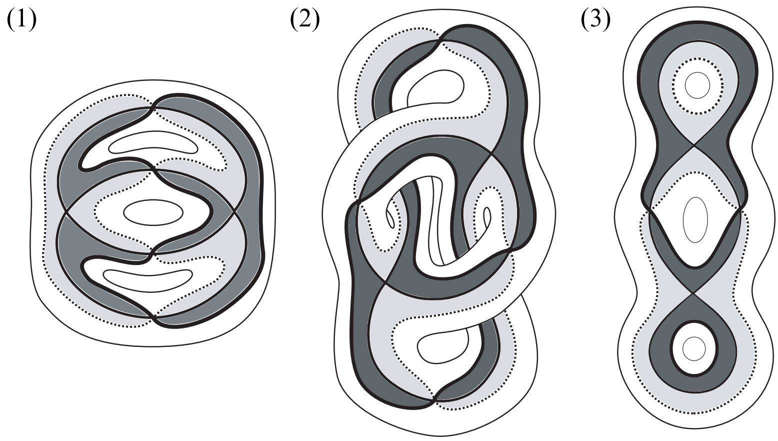

Figure 3: The twisted cases of Q_0 and Z_0. In each case Q_0 is depicted immersed in the plane, while Z_0 intersects it. The parts of Z_0 above the plane are shaded darker than the parts below. $\partial_+ Z_0$ is a heavy line, $\partial_- Z_0$ is dotted. The map p is essentially projection to the plane.

view of our subsurface F. In fact the only time that this double-saddle case actually occurs is when h weakly F–splits f, and (t, s) is the vertex in the intersection of $\overline{R_a^F}$ and $\overline{R_b^F}$, as discussed in Lemma 3.7. In this case F (up to isotopy) lies in Q_0, which allowed us to conclude in Lemma 3.7 that it must be in one of the exceptional cases.

Now divide up ∂Z_0 as $\partial_+ Z_0 = Z_0 \cap f^{-1}(t_+)$ and $\partial_- Z_0 = Z_0 \cap f^{-1}(t_-)$, each of which is embedded by p.

Lemma 5.3 *Suppose that h weakly F–splits f, (t,s) is the vertex comprising $\overline{R_a^F} \cap \overline{R_b^F}$, and Γ_0 is the component of $\Gamma = \Sigma_t \cap \Lambda_s$ whose regular neighborhood Q_0 in Σ_t contains F. Let Z_0 be the regular neighborhood of Γ_0 in Λ_s.*

If two components $\alpha_+ \subset p(\partial_+ Z_0)$ and $\alpha_- \subset p(\partial_- Z_0)$ intersect $F \subset Q_0$ essentially, then

$$d_F(\alpha_+, \alpha_-) \leq 3.$$

Proof In the untwisted case all boundary components of Z_0 map to boundary components of Q_0, and hence are peripheral and the lemma holds vacuously. The same occurs (via Lemma 5.2) if Q_0 is a 3–holed sphere and F is an annulus. We therefore consider the twisted cases from now on.

Consider first the case that F is actually isotopic to Q_0. Note that since F is essential it can only be a proper subsurface of Q_0 if F is an annulus, or if F is a 1–holed torus and Q_0 is a 2–holed torus. We return to these cases in the end.

We can read off α_+ and α_- from the diagrams in Figure 3: they are components of the thickened curves and the dotted curves, respectively. In case (1), they intersect each other four times, and one can see from Figure 4(a) that they are connected by a chain of length 3 in which the two middle vertices are the arcs $b^{\pm}$ in the figure. In case (2)

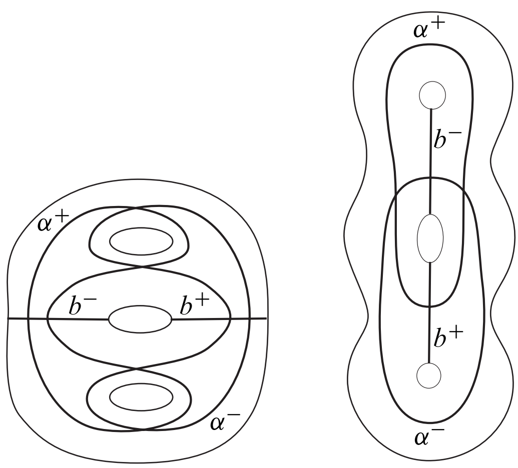

Figure 4: Checking that $d_{\mathcal{AC}(Q_0)}(\alpha^+, \alpha^-) = 3$

each possible α_- intersects each possible α_+ exactly once, and it is easy to see that the distance is 2. In case (3) they intersect twice (here two components are actually inessential in Q_0, so we must consider the others), and again the distance is 3.

Consider now the case that Q_0 is a 2–holed torus – case (2) of Figure 3 – and F is a 1–holed torus. Again α_+ and α_- intersect exactly once, therefore their intersections with F (if any) intersect at most once. Hence their distance in $\mathcal{AC}(F)$ is at most 3.

Finally we consider the case that F is an annulus. We can assume F is not boundary-parallel in Q_0, because otherwise it would not intersect ∂Z_0 essentially. The possibilities for Q_0 are again enumerated in Figure 3.

In case (1), let $(\alpha^+, b^+, b^-, \alpha^-)$ be the sequence from Figure 4(a). We claim that if F intersects α^+ and α^- essentially then it must intersect b^+ and b^- essentially as well. This is because if F misses b^- it must be parallel to α^-, and similarly for b^+ and α^+. This will imply that the same sequence gives a distance bound of 3 in $\mathcal{AC}(F)$, provided we understand how to lift the arcs $b^\pm$ to the annulus complex.

The boundary components of Q_0 lift to a union Θ of boundary-parallel arcs in the annulus $\widehat{F}$ (see Figure 5). The arcs $b^\pm$ lift to disjoint arcs connecting components

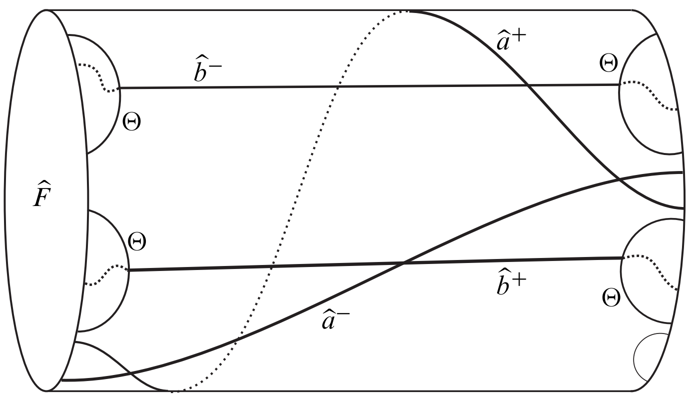

Figure 5: Lifting curves and arcs to the annulus $\widehat{F}$

of Θ, and since they cross F essentially, we can choose lifts with endpoints on arcs of Θ adjacent to opposite sides of $\widehat{F}$. Given such a lift, we can join its endpoints to $\partial\widehat{F}$ using arcs in the disks bounded by the components of Θ that it meets. This results in disjoint properly embedded arcs $\hat{b}^\pm$, such that $\hat{b}^+$ is disjoint from the lifts of α^+ because the latter never cross Θ, and similarly for α^-. We conclude that $d_{\mathcal{AC}(F)}(\alpha^+, \alpha^-) \le 3$.

In case (2), Figure 3 shows that any choice of α^+ and α^- intersect exactly once. If they intersect F essentially, by the argument given in [4, Section 2.1] their lifts to $\widehat{F}$ cross at most twice. We conclude that the distance in $\mathcal{AC}(F)$ is at most 3.

In case (3) we can use exactly the same argument as for case (1), using the arcs in Figure 4(b). □

6 Splitting bounds distance

In this section we put together the results obtained so far to show that F–splitting gives a bound on $d_F(\Sigma)$.

Proposition 6.1 *Let Σ and Λ be Heegaard surfaces for M, and let $F \subset \Sigma$ be a proper connected essential subsurface. If Λ F–splits Σ then*

$$d_F(\Sigma) \le 2g(\Lambda).$$

If Λ weakly F–splits Σ then

$$d_F(\Sigma) \le 2g(\Lambda) + 2.$$

Proof Choose sweepouts h and f with $f \times h$ generic, such that h F–splits f or weakly F–splits f, according to our hypothesis.

In the F–splitting case there is a value s such that $[-1,1] \times \{s\}$ is disjoint from the closures of R_a^F and R_b^F. In particular this means that there is an interval of such values and we may choose one such that $[-1,1] \times \{s\}$ meets no vertices of the graphic. It follows that the function $f|_{\Lambda_s}$ is a Morse function, having at most one critical point per critical value. We fix this s and henceforth identify Λ with Λ_s.

In the weak F–splitting case there is a value of s such that $[-1,1] \times \{s\}$ intersects the closures of R_a^F and R_b^F only at their unique intersection point, which is a vertex (t,s) that we will call the *weak splitting vertex*. We again fix this s and let $\Lambda = \Lambda_s$.

The proof proceeds along the following lines: The intersections $\Lambda \cap \Sigma_t$ are the level sets of $f|_\Lambda$, and the idea is to consider them as curves on Σ, argue using the F–splitting property that they intersect F essentially, and then use the topological complexity of Λ to bound the diameter in $\mathcal{AC}(F)$ of the corresponding set. This kind of argument originates with Kobayashi [12]. It was extended in various ways by Hartshorne [5], Bachman and Schleimer [1], Scharlemann and Tomova [20] and Johnson [10]. Our argument is a direct extension of the one used in [10].

We use the map $p\colon f^{-1}((-1,1)) \to \Sigma$, as in Section 5, to map level curves of $f|_\Lambda$ into Σ. The lemmas in Section 5 will help us control what happens as we pass through critical points of $f|_\Lambda$.

One important complication is the possibility that curves that are essential in one surface are trivial in the other. To handle this we adapt an argument of [10]:

Lemma 6.2 *If $d_F(\Sigma) > 3$, then there is some nontrivial interval $[u,v] \subset (-1,1)$ such that for each regular $t \in [u,v]$, every loop of $\Sigma_t \cap \Lambda$ that is trivial in Λ is trivial in Σ_t.*

Here and below when we say t is regular we mean it is a regular value for $f|_\Lambda$.

Proof As in Lemma 5.1, the isotopy classes in Σ of level sets of $f|_\Lambda$ are constant over an interval between two critical values. Thus it suffices to find a single regular value t such that every loop of $\Sigma_t \cap \Lambda$ that is trivial in Λ is trivial in Σ_t.

Note also that if $\Sigma \smallsetminus F$ is compressible to at least one side of Σ then $d_F(\Sigma) \leq 3$ by Lemma 2.1. Hence we may assume that every disk in H^+ or H^- intersects F.

Assume, seeking a contradiction, that for every regular t there is a loop of $\Sigma_t \cap \Lambda$ that is essential in Σ_t and trivial in Λ. We can assume this loop is innermost within Λ among intersection loops that are essential in Σ_t. Within the disk that it bounds in Λ there might still be intersection loops that are trivial in both surfaces, but after some disk exchanges we see that we have a loop of $\Sigma_t \cap \Lambda$ which bounds an essential disk in H_t^- or H_t^+. The loops of $\Sigma_t \cap \Lambda$, for regular values of t, are pairwise disjoint in Σ_t, so if there are loops of $\Sigma_t \cap \Lambda$ that bound disks on opposite sides of Σ_t then $d_F(\Sigma) \leq 1$. Note we are using here the fact that all such disks must meet F essentially.

If Λ F–splits Σ and s has been chosen as above, then for a nonregular value t_0, there is a unique critical point. If it is a saddle point then Lemma 5.2 tells us that the p–images of loops of intersection of $\Sigma_t \cap \Lambda$ for the regular values t just before t_0 and just after t_0 are also pairwise disjoint in Σ. We call this a *saddle transition*. So if the loops switch from bounding a disk on one side of Σ_{t_0} to the other, we again find that the distance $d_F(\Sigma)$ is at most one. If the point is a central singularity then (again referring to Section 5) it corresponds to the appearance or disappearance of a level curve which is trivial in *both* Λ and Σ, and so can be ignored. We call this a *central transition*.

If Λ *weakly* F–splits Σ and s has been chosen as above, then a nonregular value t_0 could occur at the weak splitting vertex, (t_0, s). In this case, recall that F is in one of the exceptional cases, and by Lemma 5.3, any level curves just before and just after t_0 which intersect F essentially have distance at most 3 in $\mathcal{AC}(F)$. Hence we obtain $d_F(\Sigma) \leq 3$ if these loops correspond to disks on opposite sides.

Now, for t near -1, all the loops of intersection bound disks on the negative side of Σ_t and for t near 1, they all bound disks on the positive side. Hence they must switch sides at some point, thus implying $d_F(\Sigma) \leq 3$. We conclude that there is some regular value t for which the assumption does not hold. Thus the first paragraph gives the desired interval $[u, v]$. □

For a regular value $t \in (-1, 1)$ of $f|_\Lambda$, let $\mathcal{L}_t$ denote the set of nontrivial isotopy classes in Σ of the p–images of the curves of $(f|_\Lambda)^{-1}(t)$. For an interval $J \subset (-1, 1)$ let $\mathcal{L}_J$ denote the union of $\mathcal{L}_t$ over regular $t \in J$.

We now observe that for each regular $t \in (-1,1)$, $\mathcal{L}_t$ contains a curve that intersects F essentially, and in particular $\pi_F(\mathcal{L}_t)$ is nonempty. This is because both F–splitting and weak F–splitting tell us that (t,s) is in neither R_a^F nor R_b^F (it cannot be the weak splitting vertex since t is regular), so that both $V_s^+ \cap \Sigma_t$ and $V_s^- \cap \Sigma_t$ cannot be isotoped off F_t in Σ_t, and hence neither can their common boundary.

We can now bound the diameter in $\mathcal{AC}(F)$ of $\mathcal{L}_{[u,v]}$ for $[u,v]$ satisfying the conclusion of Lemma 6.2.

Lemma 6.3 *Let $[u,v]$ be an interval such that for each regular $t \in [u,v]$, every loop of $\Sigma_t \cap \Lambda$ that is trivial in Λ is trivial in Σ_t. If $(u,v) \times \{s\}$ encounters no vertices of the graphic then*

$$\mathrm{diam}_F(\mathcal{L}_{[u,v]}) \le 2g(\Lambda) - 2.$$

If $(u,v) \times \{s\}$ meets the weak splitting vertex then

$$\mathrm{diam}_F(\mathcal{L}_{[u,v]}) \le 2g(\Lambda) - 1.$$

Proof The critical values of $f|_\Lambda$ cut $[u,v]$ into intervals. Let $t_0 < \cdots < t_n \in (u,v)$ be a selection of one regular point for each interval. Then $\bigcup_i \mathcal{L}_{t_i} = \mathcal{L}_{[u,v]}$.

Consider the transition from $\mathcal{L}_{t_i}$ to $\mathcal{L}_{t_{i+1}}$. If it is a central transition then only the loss or gain of a curve trivial in both Λ and Σ is involved. Hence $\mathcal{L}_{t_i} = \mathcal{L}_{t_{i+1}}$.

Suppose it is a saddle transition (as in the proof of Lemma 6.2), but that at least one of the curves of the associated 3–holed sphere is trivial in Λ. Then its p–image is trivial in Σ, and the remaining two curves are trivial or homotopic. Once more $\mathcal{L}_{t_i} = \mathcal{L}_{t_{i+1}}$.

If it is an *essential* saddle transition, meaning all three boundary curves are essential in Λ, then Lemma 5.2 implies that all the curves of $\mathcal{L}_{t_i}$ and $\mathcal{L}_{t_{i+1}}$ are pairwise disjoint.

Suppose we are in the first case, where $(u,v) \times \{s\}$ encounters no vertices of the graphic. Then these transitions are the only possibilities.

Since every $\mathcal{L}_{t_i}$ intersects F essentially we find that each $\mu_i \equiv \pi_F(\mathcal{L}_{t_i})$ is nonempty, and two successive μ_i are equal unless they differ by an essential saddle, in which case they are at most distance 1 apart. Since the pants that occur among level sets of $f|_\Lambda$ are all disjoint, and since the number of disjoint essential pants in Λ is bounded by $-\chi(\Lambda) = 2g(\Lambda) - 2$, this gives the desired bound on $\mathrm{diam}(\pi_F(\mathcal{L}_{[u,v]}))$.

If $(u,v) \times \{s\}$ meets the weak splitting vertex then, as in the proof of Lemma 6.2, we apply Lemma 5.3 to show that the curves before and after have distance at most 3 in $\mathcal{AC}(F)$. On the other hand such a transition uses up two saddles. If we can show that both are essential in Λ, this will make our count higher by just 1, ie we get a bound of $2g(\Lambda)-1$. We therefore proceed to show this.

Note first that only the twisted case can occur here: If Q_0 and Z_0 are in the untwisted case then in fact $p(\partial Z_0)$ is isotopic to ∂Q_0 and hence does not intersect F. This means that for the level t_i just after (or before) the weak splitting vertex, $\mathcal{L}_{t_i}$ does not intersect F, which is a contradiction. Hence this case cannot happen.

Recall that in the weak splitting case F is isotopic into Q_0 (notation as in Section 5). Suppose Q_0 is an essential subsurface of Σ (which holds in particular whenever F is isotopic to Q_0, since F itself is essential). By the hypothesis on $[u,v]$ the boundary curves of Z_0 cannot be trivial in Λ since they map to nontrivial curves in Q_0. Hence Z_0 is essential in Λ so both saddles are essential.

Suppose that Q_0 is not essential in Σ, and hence F is a proper subsurface of Q_0. If Q_0 is a 4–holed sphere, as in cases (1) and (3) of Figure 3, then F must be an annulus. Since at least one of the boundary components of Q_0 is inessential in Σ, F must be isotopic to one of the remaining essential boundary components of Q_0. This means that F has no essential intersections with ∂Z_0, which as above contradicts the fact that $\mathcal{L}_{t_i}$ always intersects F essentially.

If Q_0 is a 2–holed torus, as in case (2) of Figure 3, then we observe that none of the boundary curves of Z_0 can be inessential in Σ, since each one has exactly one intersection point with one of the others. Hence they must be nontrivial in Λ as well, so again we get two essential saddles, and therefore obtain a bound of $2g(\Lambda)-1$. □

We are now ready to complete the proof of Proposition 6.1.

If $d_F(\Sigma)\le 3$ then the claim of Proposition 6.1 follows immediately, since we assumed Λ has genus at least two. Hence we may assume that $d_F(\Sigma)>3$.

In particular we can obtain at least one interval $[u,v]$ satisfying the conclusion of Lemma 6.2. Note that if two such intervals intersect then their union also satisfies the conclusion of Lemma 6.2. Hence we can let $[u,v]$ be a *maximal* such interval. The endpoints u and v must be critical values, else we could enlarge the interval. Note also $u>-1$, since for t close to -1, Σ_t is the boundary of a small neighborhood of the spine $f^{-1}(-1)$ and so intersects Λ in small circles that bound disks in Λ but are essential in Σ_t. Similarly we see $v<1$. Let $u'<u$ be a regular value not separated from u by any critical values. Then $\Lambda\cap\Sigma_{u'}$ must contain a component β that is trivial in Λ but essential in $\Sigma_{u'}$.

Moreover β must in fact be trivial in the handlebody $f^{-1}([-1,u'])$ – in other words β determines an element of $\mathcal{D}(H^-)$. To see this, let E be the disk in Λ bounded by β.

Let $u'' \in (u, v)$ be a regular value, and let $E_1 = E \cap f^{-1}([-1, u''])$. Then E_1 contains a neighborhood of β, and any internal boundary component of E_1 is trivial in E, and hence in Λ. Since $u'' \in [u, v]$, such a curve is also trivial in $\Sigma_{u''}$, and so after a surgery and isotopy we can obtain a disk in $f^{-1}([-1, u'])$, whose boundary is β.

Similarly, if we consider a regular $v' > v$ not separated from v by critical values we obtain $\beta' \in \mathcal{D}(H^+)$ in the level set of v'.

As before, since $d_F(\Sigma) > 3$ we may assume by Lemma 2.1 that both β and β' intersect F essentially.

Now assume we are in the F–splitting case, so that all critical values have single critical points. In that case, by Lemma 5.2, β is disjoint from the regular curves at level just above u, and β' is disjoint from the regular curves at level just below v. It follows that $d_F(\beta, \mathcal{L}_{[u,v]}) \leq 1$, and similarly for β'. We conclude via Lemma 6.3 that $d_F(\beta, \beta') \leq 2g(\Lambda) - 2 + 2$, which is what we wanted to prove.

In the weakly F–splitting case, there could be one critical value corresponding to a vertex of the graphic. If this occurs inside (u, v) then we have already taken account of it in the bound of $2g(\Lambda) - 1$ in Lemma 6.3, so we get a final bound of $d_F(\beta, \beta') \leq 2g(\Lambda) + 1$. If the vertex occurs at u or at v, then we get contribution of 3 instead of 1 from one of the disks. Lemma 6.3 gives us $2g(\Lambda) - 2$ in that case, so we get $d_F(\beta, \beta') \leq 2g(\Lambda) - 2 + 4 = 2g(\Lambda) + 2$. That concludes the proof of Proposition 6.1. □

7 Proof of the main theorem

Proof of Theorem 1.1 Let f and h be sweepouts for Σ and Λ, respectively. Isotope f and h so that $f \times h$ is generic.

Suppose first that F is not an annulus, 4–holed sphere, or 1– or 2–holed torus. If $d_F(\Sigma) > 2g(\Lambda)$ then, by Proposition 6.1, h cannot F–split f. By Lemma 3.7, if h does not F–split f then h must F–span f, and then Proposition 4.1 implies that after isotopy the intersection of Λ and Σ contains F.

Now suppose F is an annulus, 4–holed sphere, or 1– or 2–holed torus. If $d_F(\Sigma) > 2g(\Lambda) + 2$ then, by Proposition 6.1, h cannot *weakly* F–split f. By definition this means h F–spans f, and again we are done by Proposition 4.1. □

References

[1] **D Bachman**, **S Schleimer**, *Surface bundles versus Heegaard splittings*, Comm. Anal. Geom. 13 (2005) 903–928 MR2216145

[2] **D Bachman**, **S Schleimer**, **E Sedgwick**, *Sweepouts of amalgamated* 3*–manifolds*, Algebr. Geom. Topol. 6 (2006) 171–194 MR2199458

[3] **J F Brock**, **R D Canary**, **Y N Minsky**, *The classification of Kleinian surface groups II: the ending lamination conjecture*, to appear in Ann. of Math. (2) `arXiv:math.GT/0412006`

[4] **B Farb**, **A Lubotzky**, **Y Minsky**, *Rank–*1 *phenomena for mapping class groups*, Duke Math. J. 106 (2001) 581–597 MR1813237

[5] **K Hartshorn**, *Heegaard splittings of Haken manifolds have bounded distance*, Pacific J. Math. 204 (2002) 61–75 MR1905192

[6] **J Hempel**, 3*–Manifolds*, Ann. of Math. Studies 86, Princeton Univ. Press (1976) MR0415619

[7] **J Hempel**, 3*–manifolds as viewed from the curve complex*, Topology 40 (2001) 631–657 MR1838999

[8] **N V Ivanov**, *The rank of Teichmüller modular groups*, Mat. Zametki 44 (1988) 636–644, 701 MR980584

[9] **N V Ivanov**, *Subgroups of Teichmüller modular groups*, Transl. of Math. Monogr. 115, Amer. Math. Soc. (1992) MR1195787 Translated from the Russian by E J F Primrose and revised by the author

[10] **J Johnson**, *Surface bundles with genus two Heegaard splittings*, J. Topol. 1 (2008) 671–692 MR2417449

[11] **J Johnson**, *Stable functions and common stabilizations of Heegaard splittings*, Trans. Amer. Math. Soc. 361 (2009) 3747–3765 MR2491898

[12] **T Kobayashi**, *Heights of simple loops and pseudo-Anosov homeomorphisms*, from: "Braids (Santa Cruz, CA, 1986)", (J S Birman, A Libgober, editors), Contemp. Math. 78, Amer. Math. Soc. (1988) 327–338 MR975087

[13] **T Kobayashi**, **O Saeki**, *The Rubinstein–Scharlemann graphic of a* 3*–manifold as the discriminant set of a stable map*, Pacific J. Math. 195 (2000) 101–156 MR1781617

[14] **M Lackenby**, *The Heegaard genus of amalgamated* 3*–manifolds*, Geom. Dedicata 109 (2004) 139–145 MR2113191

[15] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves. I. Hyperbolicity*, Invent. Math. 138 (1999) 103–149 MR1714338

[16] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves. II. Hierarchical structure*, Geom. Funct. Anal. 10 (2000) 902–974 MR1791145

[17] **J N Mather**, *Stability of C^∞ mappings. V. Transversality*, Advances in Math. 4 (1970) 301–336 (1970) MR0275461

[18] **Y Minsky**, *The classification of Kleinian surface groups. I. Models and bounds*, Ann. of Math. (2) 171 (2010) 1–107 MR2630036

[19] **H Rubinstein**, **M Scharlemann**, *Comparing Heegaard splittings of non-Haken* 3–*manifolds*, Topology 35 (1996) 1005–1026 MR1404921

[20] **M Scharlemann**, **M Tomova**, *Alternate Heegaard genus bounds distance*, Geom. Topol. 10 (2006) 593–617 MR2224466

Department of Mathematics, Oklahoma State University
401 MSCS, Stillwater OK 74078, USA

Department of Mathematics, Yale University
10 Hillhouse Ave, New Haven CT 06520-8283, USA

Department of Mathematics, Technion
32000 Haifa, Israel

jjohnson@math.okstate.edu, yair.minsky@yale.edu,
ymoriah@tx.technion.ac.il

http://www.math.okstate.edu/~jjohnson/, http://www.math.yale.edu/users/yair, http://www.math.technion.ac.il/~ymoriah/

Received: 25 February 2010 Revised: 26 July 2010