J. DIFFERENTIAL GEOMETRY
39 (1994) 343-357

THIN POSITION AND HEEGAARD SPLITTINGS OF THE 3-SPHERE

MARTIN SCHARLEMANN & ABIGAIL THOMPSON

We present here a simplified proof of the theorem, originally due to Waldhausen [7], that a Heegaard splitting of S^3 is determined solely by its genus. The proof combines Gabai's powerful idea of "thin position" [2] with Johannson's [4] elementary proof of Haken's theorem [3] (Heegaard splittings of reducible 3-manifolds are reducible). In §3.1, 3.2 & 3.8 we borrow from Otal [6] the idea of viewing the Heegaard splitting as a graph in 3-space in which we seek an unknotted cycle.

Along the way we show also that Heegaard splittings of boundary reducible 3-manifolds are boundary reducible [1, 1.2], obtain some (apparently new) characterizations of graphs in 3-space with boundary-reducible complement, and recapture a critical lemma of [5]. We are indebted to Erhard Luft for pointing out a gap in the original argument.

1. Heegaard splittings: a brief review

1.1. All surfaces and 3-manifolds will be compact and orientable. A *compression body* H is constructed by adding 2-handles to a (surface) $\times$ 1 along a collection of disjoint simple closed curves on (surface) $\times\{0\}$, and capping off any resulting 2-sphere boundary components with 3-balls. The component (surface) $\times\{1\}$ of ∂H is denoted $\partial_+ H$ and the surface $\partial H - \partial_+ H$, which may or may not be connected, is denoted $\partial_- H$ (Figure 1a, next page). If $\partial_- H = \emptyset$, then H is a *handlebody*. If $H = \partial_+ H \times 1$, H is called a *trivial* compression body. A *spine* for H is a properly imbedded 1-complex Q such that H collapses to $Q \cup \partial_- H$ (Figure 1b).

1.2. Spines are not unique, but can be altered by *edge slides*, as follows: Choose an edge e in Q and let $\overline{Q}$ be the graph $Q - e$. Let $\overline{H}$ denote a regular neighborhood of $\partial_- H \cup \overline{Q}$. Then H is the union of $\overline{H}$ and a 1-handle h attached to $\partial_+ \overline{H}$. The core of h is the edge e, with its ends

Received January 15, 1993. Authors are supported in part by a National Science Foundation grant. The second author is a Sloan Foundation Fellow.

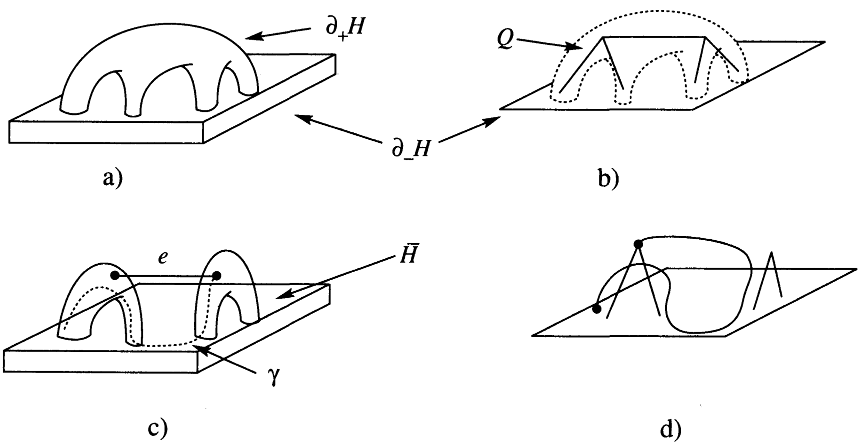


FIGURE 1

in $\overline{H}$ deleted so that $\partial e \subset \partial_* \overline{H}$. Suppose γ is a path on $\partial_+ \overline{H}$ which begins at an end of e (equivalently, a path in $\partial_+ H$ which begins at a meridian of h but never crosses the meridian) (Figure 1c). Then the end of e can be isotoped along γ before h is attached. The effect on Q is to replace e with the union of e and a copy of γ pushed slightly away from $Q \cup \partial\_H$ (Figure 1d).

Here is an apparent generalization: Suppose x is a point in the interior of e, dividing it into two segments e' and e''. Suppose γ is a path in $\partial_+ H$ which begins and ends at the meridian of e at x and never intersects a meridian of e'. Then introduce a new vertex at x, perform an edge slide of e' using γ as above, and then reamalgamate e' and e'' at x. This operation will be called a *broken edge slide.* A broken edge slide can be viewed as a series of standard edge slides: slide an edge incident to the other end of e'' down e'' to dx, introducing a vertex at x, perform the (now standard) edge slide of e', then unto the original slide.

1.3. Let F be a closed connected surface imbedded in a 3-manifold N. F is a *splitting surface* for a Heegaard splitting if F divides N into two compression bodies H_1 and H_2 with $\partial_+ H_1 = F = \partial_+ H_2$. An *elementary stabilization* of F is the splitting surface obtained by taking the connected sum of pairs $(N, F)\#(S^3, T)$, for T the standard unknotted torus in S^3. A Heegaard splitting is *stabilized* if it is an elementary stabilization of another splitting. This is equivalent to the existence of proper disks $D_1 \subset H_1$ qnd $D_2 \subset H_2$ with $\partial D_1 \cap \partial D_2$ a single point in F.

The Heegaard splitting is *reducible* if there exists an essential simple closed curve $c \subset F$ which bounds imbedding disks in both H_1 and H_2. Equivalently, there is a sphere in N which intersects F in a single circle which is essential in F. A stabilized Heegaard splitting F with genus$(F) > 1$ is reducible, for in this case the boundary of a regular neighborhood of $\partial D_1 \cup \partial D_2$ is essential in F, yet bounds a disk in both H_1 and H_2. A Heegaard splitting is ∂*-reducible* if there is a ∂-reducing disk for N which intersects F in a single circle.

1.4. Reducible and ∂-reducible Heegaard splittings have a particularly nice property. Suppose S is a sphere intersecting F in a single essential circle c. Remove a neighborhood $S \times 1$ of S from N and cap off $S \times \partial 1$ with two 3-balls, creating a new 3-manifold N'. Simultaneously compress F along the disk in H_1 (or H_2) which c bounds. Then F becomes a Heegaard splitting surface for N'. The same thing happens if N is ∂-reduced along a disk which is also ∂-reducing for the Heegaard splitting.

1.5. Suppose Δ is a properly imbedded family of disks in a 3-manifold M, and D is a disk in M whose interior is disjoint from Δ and whose boundary either lies entirely on Δ or is the union of two arcs, one on Δ and one on ∂M. Then the boundary of a regular neighborhood of $\Delta \cup D$ has two parts, one isotopic to Δ and the other a new set of properly imbedded disks Δ' (together with a 2-sphere if $\partial D \subset \Delta$). We say Δ' is obtained from Δ by a *disk-swap* along D. If $\partial D \subset \Delta$, then Δ' has the same number of disks. Otherwise it has one more.

A property of the family Δ which is always preserved by disk-swaps is said to be *swap-preserved.* Here are three examples of such properties:

(a) Δ contains a ∂-reducing disk for M.

(b) Δ contains a complete collection of ∂-reducing disks for M.

(c) For κ a given normal subgroup of $\pi_1(\partial M)$, $\partial\Delta$ contains a component representing a class not in κ.

2. Sliding spines around: Haken's theorem

2.1. Let N be a compact orientable 3-manifold, and H a properly imbedded compression body in N, with spine Q. For a compression body H, properly imbedded means $\partial N \cap H = \partial\_H$. Let $(T, \partial T) \subset (N, \partial N)$ be a properly imbedded surface in N, and $(\Delta, \partial\Delta)$ a family of properly imbedded disks in $M = N - H$. Extend Δ via the retraction $H \to \partial\_H \cup Q$ so that it becomes a collection of disks in N whose imbedded interior is disjoint from Q and whose (singular) boundary lies on $\partial\_H \cup Q$. Put T

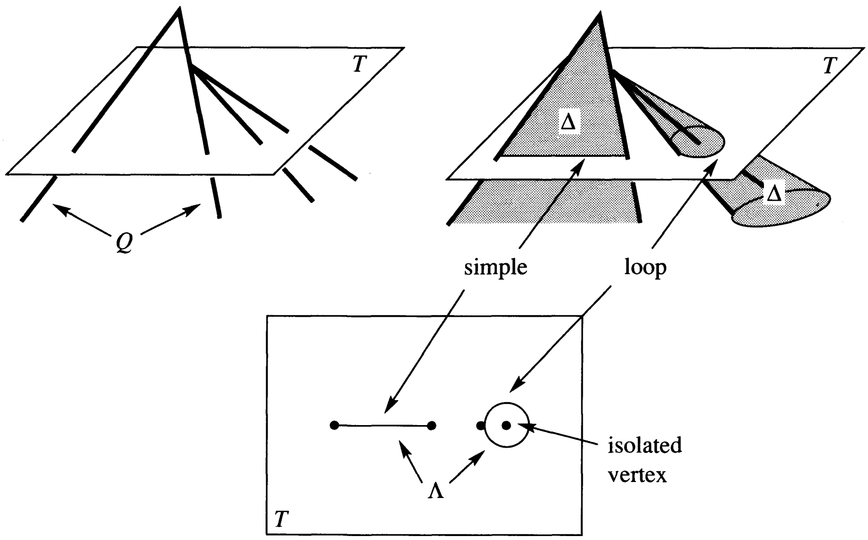


FIGURE 2

in general position with respect to Q and Δ. Then $Q \cap T$ is a finite set of points, and $(\Delta \cap T) - Q$ is the interior of a 1-manifold whose boundary is incident to $Q \cap T$. Ignoring closed components of $\Delta \cap T$, we can view the result as a graph Λ in T, with vertices the points $Q \cap T$, and edges the arc components of $\Delta \cap T$. An edge in Λ is *simple* if its ends lie on different vertices, otherwise it is a *loop* based at the vertex common to both its ends. A loop is inessential if it bounds a disk in T disjoint from Q, otherwise it is *essential.* A vertex in Λ is *isolated* if it is incident to no edge. Such a vertex in Λ represents a point in an edge of Q which is incident to no 2-disks of Δ. See Figure 2.

An edge α in Λ with an end at a vertex w is called an edge *at* w. The edge α divides the disk $D \in \Delta$ in which it lies into two disks. Suppose one of them, E, contains no arc of intersection with T corresponding to an edge of Λ at w. Then we say that α is *outermost for* w and that E is the corresponding outermost disk. Note that E may still contain many components of intersection with T, but none will be edges at w.

2.2. Proposition. *For a given swap-preserved property of properly imbedded disk families in* M, *let* Q *be a spine of* H, *and* Δ *a disk family in* M *with the given property, chosen so that the pair* $(|Q \cap T|, |\Delta \cap T|)$ *is minimized. Then each vertex of the corresponding graph* Λ *in* T *is either isolated or the base of an essential loop i* T.

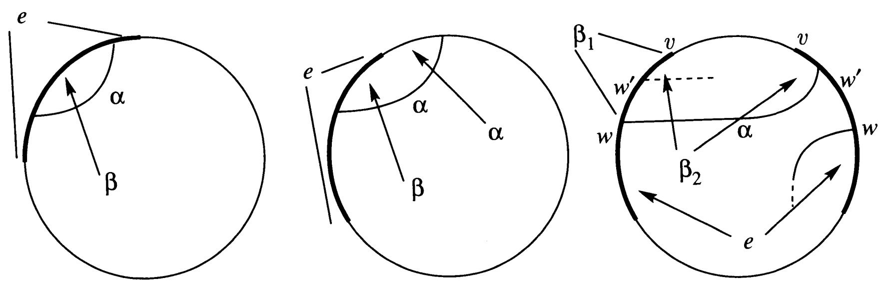


FIGURE 3

Proof. The alternative is that there is a vertex w of Λ incident to some simple edges and possibly some inessential loops. In fact there can be no loops, because a disk-swap along the disk cut off in T by an innermost inessential loop (or a disk component of $T - \Delta$ within it) would reduce $|\Delta \cap T|$. Thus w is incident to some simple edges, but no loops. Let α be an outermost edge in Λ at w, and E the corresponding outermost disk.

Let e be the edge in Q on which w lies. Since $\partial\Delta$ comes from a normal family of simple closed curves in $\partial\eta(Q)$, the subarcs of $\partial\Delta$ lying on $\eta(e)$ can be thought of as copies of e lying in $\partial\Delta$. Since α is outermost for w, no complete copy of e can lie in $\partial E \cap \partial\Delta$. There are then three possible ways in which the arc $\partial E \cap \partial\Delta$ could intersect copies of e in $\partial\Delta$ (see Figure 3):

1) $\partial E \cap \partial\Delta$ could be a subsegment of e or
2) one end of $\partial E \cap \partial\Delta$ could lie in a copy of e or
3) each end of $\partial E \cap \partial\Delta$ could lie in a copy of e.

In each case we can reduce $|Q \cap T|$ (see Figure 4, next page).

1) When $\partial E = \alpha \cup \beta$, β a subsegment of e, then E describes an isotopy of β to α which eliminates both w and w'.

2) When $\partial E \cap \partial\Delta$ is the union of an end segment β of e running from w to an end vertex v of e in Q, and a path γ from v to the other end of α in $Q - e$, then γ describes a path on which to slide the end of e at v. The slide reduces the problem to the previous case.

3) In the last case, $\partial E \cap \partial\Delta$ is the union of three segments: an end segment β_1 of e running from w to an end vertex v of e in Q, a path γ from v to an end vertex v' of e, and a segment β_2 of e running from v' to the other end of α, which we call w'. β_2 cannot contain w, since α is outermost for w. In the case that v' and v are different ends of e, the argument of case 2 applies, using $\gamma \cup \beta_2$ instead of γ. When

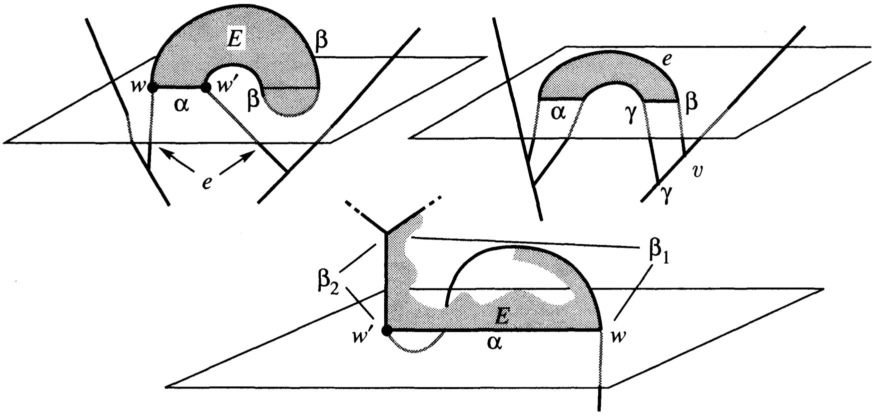


FIGURE 4

$v' = v$, as illustrated in Figure 3, we have $\beta_2 \subset \beta_1$, Break the edge e into β_2 and $e - \beta_2$ by introducing a new vertex at w'. Then as in case 2), E describes an edge slide of $e - \beta_2$ which moves the segment $\beta_1 - \beta_2$ to α. Then reamalgamate $e - \beta_2$ and β_2 at w'. This is a broken edge slide (see 1.2) which removes the point w, as well as any points of $(\beta_1 - \beta_2) \cap T$ from $Q \cap T$, thereby reducing $|Q \cap T|$.

The contradiction completes the proof of the proposition.

2.3. Corollary. *If Q is a properly imbedded graph in a reducible or ∂-reducible 3-manifold N, and $N - \eta(Q)$ is irreducible but ∂-reducible, then, after some edge slides of Q, there is a ∂-reducing disk for $N - \eta(Q)$ whose boundary is disjoint from some edge in Q.*

Proof. Apply the proposition, letting T be a reducing sphere or ∂-reducing disk of N, H a regular neighborhood of $\partial N \cup Q$, and Δ a family of disks containing a ∂-reducing disk. If Q is disjoint from T, then T is a ∂-reducing disk in $N - \eta(Q)$ and we are done. Otherwise, some vertex w of Λ must be isolated, since an innermost loop in T would otherwise be inessential. But this implies that $\partial\Delta$ may be isotoped off the edge of Q containing w.

2.4. Corollary. (a) *Any Heegaard splitting of a reducible 3-manifold is reducble.*

(b) *Any Heegaard splitting surface of a ∂-reducible 3-manifold is ∂-reducible.*

Proof. (a) is essentially [3] and (b) essentially [1, 1.2]. The following alternative proofs are really a reformulation of [4, 3.2] that exploits 2.2.

First observe that it will suffice to prove a weaker proposition. A Heegaard splitting of a reducible or ∂-reducible 3-manifold is either reducible or ∂-reducible. To see that this suffices, suppose, for example, that N is reducible. We would know, then, that the splitting is either reducible (and we are done) or ∂-reducible. Maximally ∂-reduce N along disks that are also ∂-reducing for the Heegaard splitting. The result is still a Heegaard splitting for the new and still reducible manifold N'. Since no further ∂-reductions of the new Heegaard splitting F' on N' are possible, we conclude that the splitting F' must be reducible. But a reducing sphere for F' is also one for F. A symmetric argument, using maximal reductions of F, applies if instead we are initially given that N is ∂-reducible.

So we proceed with the proof of the weaker assertion, given that N is either reducible or ∂-reducible. Apply the proposition with the following data: T is a reducing sphere or ∂-reducing disk for N, H is one of the two compression bodies in the Heegaard splitting of N, and Δ is a family of disks in the other compression body H' which contains a complete collection of ∂-reducing disks for H'.

The argument of the previous corollary shows that either a ∂-reducing disk for N is disjoint from Q, so the splitting is ∂-reducible, or some edge e of the spine Q of H is disjoint from the boundary of a complete collection of ∂-reducing disks for H'. In the latter case the boundary μ of a meridian of e is parallel in H' to a circle c in $\partial_- H'$, or, if H' is a handlebody so $\partial_- H'$ is empty, μ bounds a disk in H'. If c is essential in $\partial_- H'$, we have a ∂-reducing disk intersecting the splitting surface in the single circle μ. If c is inessential in $\partial_- H'$, then μ bounds a disk in H' as well, giving a sphere intersecting the splitting surface in the single circle μ. So in every case the splitting is either reducible or ∂-reducible.

3. Thin position of graphs in 3-space

Let Γ be a finite graph in S^3 in which all vertices have valence 3. Let $h\colon S^3 \to R$ denote projection of $S^3 \subset R^4$ onto a coordinate, so that, besides the two poles, the level sets $h^{-1}(t)$ of h are concentric spheres in S^3. Alternatively, we can think of Γ as lying in R^3 and set $h\colon R^3 \to R$ to be distance from the origin. Let V denote the set of vertices of Γ, and $S(t)$ the sphere $h^{-1}(t)$.

3.1. Definition. A graph Γ in S^3 is in *Morse position* with respect to h if

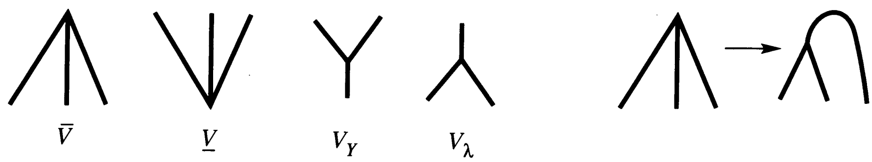


FIGURE 5

(a) on any edge e of Γ, the critical points of $h|e$ are nondegenerate and lie in the interior of e,

(b) the critical points of $h|\Gamma - V$ and the vertices V all occur at different heights.

The set of heights at which either there is a critical point of $h|\Gamma - V$ or a vertex of V is called the set of *critical heights* for Γ. The vertices V of Γ then can be classified into four types (see Figure 5):

$\overline{V} = \{v$ in V so that all ends of incident edges lie below $v\}$,

$\underline{V} = \{v$ in V so that all ends of incident edges lie above $v\}$,

$V_Y = \{v$ in V so that exactly two ends of incident edges lie above $v\}$,

$V_\lambda = \{v$ in V so that exactly two ends of incident edges lie below $v\}$.

We will further simplify the local picture by isotoping a neighborhood of a vertex in $\overline{V}$ (resp. $\underline{V}$), transforming it into a vertex in V_λ (resp. V_Y) and a nearby maximum (resp. minimum). Then all vertices are of type V_Y or V_λ. Such a graph is said to be in *normal form.*

3.2. A regular neighborhood $\eta(\Gamma)$ of Γ can be viewed as the union of 0-handles, each a neighborhood of a vertex, and 1-handles, each a neighborhood of an edge. A simple closed curve in $\partial\eta(\Gamma)$ is in *normal form* if it intersects the boundary $\partial B^2 \times 1$ of each 1-handle in 1-fibers and intersects the boundary ∂B^3 of each 0-handle in arcs essential in the complement of the three attaching disks for the 1-handles.

A disk $(D, \partial D) \subset (S^3 - \eta(\Gamma), \partial\eta(\Gamma))$ is in normal form if

(a) ∂D is in normal form on $\partial\eta(\Gamma)$,

(b) each critical point of h on D is nondegenerate,

(c) no critical point of h on $\text{int}(D)$ is a critical height of Γ,

(d) no two critical points of h on $\text{int}(D)$ occur at the same height,

(e) the minima of $h|\partial D$ at Y-vertices, the minima of Γ, the maxima of $h|\partial D$ at λ-vertices and the maxima of Γ are also local extrema of h on D, i.e., "half-center" singularities. The maxima of $h|\partial D$ at Y vertices and the minima of $h|\partial D$ at λ-vertices are, on the contrary, "half-saddle" singularities of h on D (see Figure 6).

Standard Morse theory ensures that any properly imbedded essential disk $(D, \partial D) \subset (S^3 - \eta(\Gamma), \partial\eta(\Gamma))$ can be put in normal form. The image

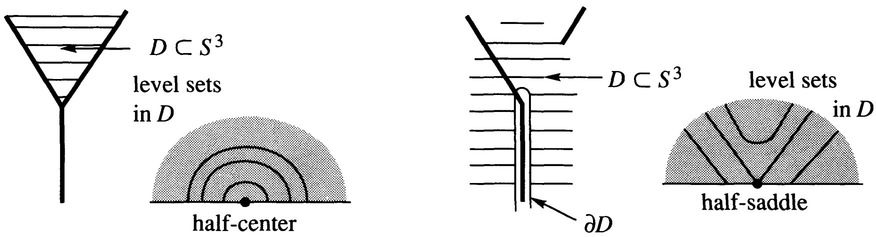


FIGURE 6

of such a normal form disk under the retraction $(S^3, \eta(\Gamma)) \to (S^3, \Gamma)$, will, with a slight abuse of terminology, be called a normal form disk $(D, \partial D) \subset (S^3, \Gamma)$.

Suppose $(D, \partial D) \subset (S^3, \Gamma)$ is a normal form disk. For t a noncritical value of $h|D$, $S(t)$ intersects D in a disjoint union of proper arcs and circles. At a critical height of $h|\Gamma - V$, the intersection may also include a finite collection of points on ∂D, corresonding to half-saddles, may be the endpoints of two arc components of $S(t) \cap D$.

3.3. For each value of t, let $w(t) = |\Gamma \cap S(t)|$. Then w is a function of t which increases by 2 at a minimum of $\Gamma - V$ and by 1 at a Y-vertex, and decreases by 2 at a maximum of $\Gamma - V$ and by 1 at a λ-vertex. Let W denote the largest value of $w(t)$, and n the number of times W appears as a local maximum of $w(t)$. Γ is in *thin position* if among all normal-form graphs obtained from Γ by isotopies and edge slides, $w(\Gamma) = (W, n)$, lexicographically ordered, is minimized.

3.4. Proposition. *Suppose a graph Γ in S^3 is in thin position and there is some nonempty disk family $(\Delta, \partial\Delta) \subset (S^3, \Gamma)$ with a given swap-preserved property. Then either there is such a disk family whose boundary is disjoint from an edge of Γ or, after at most two edge slides, Γ contains an unknotted cycle.*

Before proving the proposition, we demonstrate its utility.

3.5. Corollary. *Suppose Γ is a graph in S^3, and $S^3 - \eta(\Gamma)$ has a ∂-reducing disk whose boundary is nontrivial in $\pi_1(\Gamma)$. Then some edge slides will convert Γ to a graph containing an unknotted cycle.*

Proof of 3.5. Define the swap-preserved property of disk families in $S^3 - \eta(\Gamma)$ to be the property of containing a disk whose boundary is essential in Γ. (An alternative description: take 1.5 example (c), with κ the normal subgroup generated by meridians of $\eta(\Gamma)$.) Note that if $\partial\Delta$ is disjoint from an edge e of Γ, then Δ satisfies the same property for the graph $\Gamma - e$. So, of all normal-form graphs obtained from Γ by edge slides and edge deletions, choose Γ' to have a minimal number of edges subject to the requirement that some such family Δ exists for Γ'. Perform isotopies and edge slides of Γ' until it is in thin position. Since Γ' contains a minimal number of edges, $\partial\Delta$ must intersect all edges of Γ'. It follows from 3.4 that, after at most two edge slides, Γ' will contain an unknotted cycle.

3.6. Corollary. *Suppose Γ is a graph in S^3, and $S^3 - \eta(\Gamma)$ is ∂-reducible. Then some edge slides will convert Γ to a graph containing either an unknotted cycle or a split link.*

Proof. Following the previous case, either Γ contains an unknotted cycle, or there is a ∂-reducing disk D so that ∂D is inessential in $\eta(\Gamma)$. Then ∂D bounds a disk E in $\eta(\Gamma)$. A series of edge slides will transform Γ to a graph Γ' in which E is the meridian of some edge e. Thus $\Gamma' - e$ is split by the sphere $D \cup E$, so it contains a split link.

3.7. Corollary [5, Lemma 3.1]. *Suppose $\Gamma \subset S^3$ is a connected graph which is not a simple circuit, and suppose $S^3 - \eta(\Gamma)$ is ∂-reducible. Then, after some edge slides, there is a ∂-reducing disk for $S^3 - \eta(\Gamma)$ whose boundary is disjoint from some edge of Γ.*

Proof. Since Γ is connected, $S^3 - \eta(\Gamma)$ is irreducible. Let $\gamma \subset \Gamma$ be either an unknotted cycle or split link, given by 3.6. Apply 2.3 to the reducible or ∂-reducible manifold $S^3 - \eta(\gamma)$ with Q the nonempty graph $\Gamma - \gamma$.

3.8. Corollary. *Any Heegaard splitting of S^3 is standard.*

Proof. Suppose Γ is the spine of one side of a positive-genus Heegaard splitting F or S^3. By induction, it suffices to show that the Heegaard splitting is reducible or stabilized. Apply the proposition to Γ and the swap-preserved property that Δ contain a complete set of ∂-reducing disks for the handlebody $S^3 - \eta(\Gamma)$. If some edge e of Γ is disjoint from $\partial\Delta$, then the boundary of a meridian μ of e lies on the boundary of the 3-ball $S^3 - \eta(\Gamma \cup \Delta)$, so it bounds a disk E in $S^3 - \eta(\Gamma)$. Thus the union of E and the meridian of e is a sphere intersecting the Heegaard splitting precisely in μ, and therefore the splitting is reducible.

If, on the other hand, Γ contains an unknotted cycle γ, then the complement of a small tubular neighborhood of γ in a larger regular neighborhood of Γ is still a compression body H, with $\partial_- H$ the torus $\partial\eta(\gamma)$, and $F = \partial_+ H = \partial\eta(\Gamma)$. Hence F gives a Heegaard splitting of the solid torus $S^3 - \eta(\gamma)$. From 2.4 it follows that the Heegaard splitting of $S^3 - \eta(\gamma)$ is ∂-reducible, so that γ bounds a disk whose interior is disjoint from Γ. That disk and a meridian of γ define a stabilization of the original splitting of S^3.

Proof of 3.4. We suppose that Γ is in thin position, and that for every disk family Δ with the swap-preserved property, $\partial\Delta$ runs over every edge of Γ. We will show that, after at most two edge slides, Γ contains an unknotted cycle.

For any regular height t, the arc components of $\Delta \cap S(t)$ and the points of $\Gamma \cap S(t)$ create as before a graph $\Lambda(t)$ in $S(t)$. Since $\partial\Delta$ runs over every edge of Γ, no vertex of $\Lambda(t)$ is isolated. If a vertex were incident to a single edge of $\Lambda(t)$, then Γ would contain an edge e over which the boundary of a disk $D \in \Delta$ runs exactly once. Thus $\partial D - e$ is a path in $\Gamma - e$, and an edge slide of e along this path would convert e into an unknotted cycle. So we further assume that every vertex of $\Lambda(t)$ has valence at least 2.

Now choose Δ to minimize $|\partial\Delta \cap \{\text{meridia of } \Gamma\}|$. A loop in $\Lambda(t)$ divides $S(t)$ into two disks. Both must contain vertices of $\Lambda(t)$, for otherwise we could reduce $|\partial\Delta \cap \{\text{meridia of } \Gamma\}|$ by a disk swap along the disk in $S(t)$ bounded by an innermost inessential loop. Likewise, any edge in $\Lambda(t)$ at a vertex inside an innermost loop must be simple.

Suppose α is an outermost arc for a vertex w in $\Lambda(t)$, and E is the corresponding outermost disk. We say E is upper or lower according as, near α, it lies just above or below α for the given height function h on S^3. If w lies inside an innermost loop, then all edges are simple, and some simple edge is outermost for w. Hence we conclude that $\Lambda(t)$ is either empty or it contains an outermost simple edge.

Suppose e is the edge of Γ which intersects $S(t)$ at w in $\Lambda(t)$ as above, α is a simple edge in $\Lambda(t)$ which is outermost for w, and E is the corresonding outermost disk. Just as in 2.2, E can be used to perform a (possibly broken) edge-slide and/or isotopy of the segment β of e. If E is an upper (resp. lower) disk, the isotopy can be used to replace β with an arc just below (resp. above) α.

In order to exploit thin position, we need to do this procedure simultaneously to a pair of simple edges.

3.9.1 *Claim.* Suppose that there are simple edges which are outermost for vertices w and w' in $\Lambda(t)$ and that there are no loops of $\Lambda(t)$ based at w. Then there are simple edges α and α' which are outermost for w and w' respectively so that either the outermost disk E' cut off by α has boundary disjoint from w' or the outermost disk E cut off by α' has boundary disjoint from w.

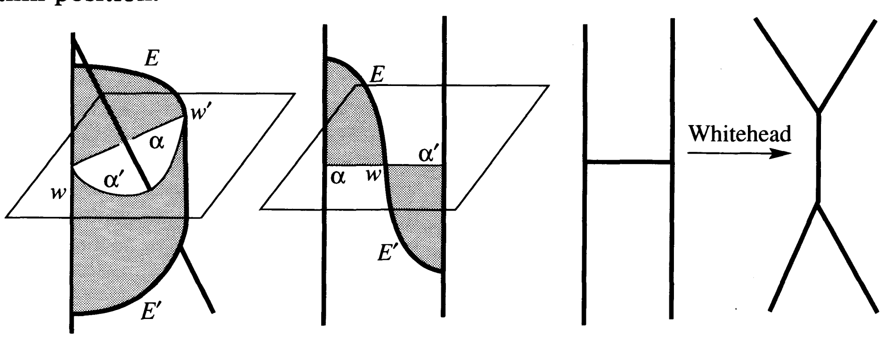
Proof. Suppose $\partial E'$ intersects w. Then an outermost arc for w in E' cuts off an outermost disk disjoint from w'.

3.9.2. *Claim.* Suppose E and E' are outermost disks for w and w' respectively, and ∂E is disjoint from w'. Let β and β' be the segments of ∂E and $\partial E'$ described above. If β and β' are disjoint, then isotopies and edge slides replace them respectively with α and α'. If not, then isotopy and edge slides replace them with α' and removes w'.

Proof. Since ∂E is disjoint from w', if β and β' are not disjoint, then $\beta \subset \beta'$. Apply the argument above first to E and then to E'. Since γ never passes through w', E' remains an outermost disk for w'.

Let t be a regular height where $w(t)$ achieves its maximum W. Then the first critical height t_- for Γ below t is either a maximum or a Y-vertex, and the first critical height t_+ above t is either a maximum or a λ-vertex.

Suppose there is both an upper and a lower outermost simple edge in $\Lambda(t)$, denoted α and α', with corresponding outermost disks E and E'. If α and α' are outermost for the same vertex w, lying on an edge e, then 3.9.2 shows that the parts of e lying on ∂E, and $\partial E'$ can be slid and isotoped to lie in $S(t)$. Unless this is the entire edge e, this move will immediately reduce the width at height t without increasing it elsewhere, contradicting thin position. So we conclude that all of e can be slid and isotoped to lie in $S(t)$. If α and α' each had their other ends at the same vertex, then e is a loop lying in $S(t)$, hence an unknotted cycle (Figure 7a). If the ends of e are at different vertices (Figure 7b), then does a Whitehead move on e, converting the horizontal edge e in Γ into a vertical edge (Figure 7c). This returns Γ to normal form, does not increase the width outside $[t_-, t_+]$ and reduces the maximal width in $[t_-, t_+]$ to at most $W-1$ (achieved perhaps at $t_\pm$). Again this contradicts thin position.



a) b) c)

FIGURE 7

So for each vertex in $\Lambda(t)$, either every outermost arc for the vertex is upper or they are all lower. Suppose w is a vertex without loops and every outermost arc for w is upper. Suppose there is anywhere in Λ an outermost simple lower edge α' for a vertex, with corresponding lower disk E'. Then by 3.9.1, either $\partial E'$ is disjoint from w or some outermost arc ∂ for w cuts off an upper disk E with boundary disjoint from w'. Now apply 3.9.2 to α and α'. This does not increase the width outside $[t_-, t_+]$ and again reduces the maximal width in $[t_-, t_+]$ to at most $W-1$.

So either every outermost simple arc in $\Lambda(t)$ is upper or they are all lower, say upper. A critical point of h on $\operatorname{int}(D^2)$ affects at most two arcs in Λ. Unless $W=2$, when Γ clearly contains the unknot, there are at least four simple outermost arcs—two at each of the two or more vertices in Λ without loops. It follows that at every regular value of t in $[t_-, t_+]$, every outermost simple arc in Λ is upper. In particular, the critical height t_- must correspond to a Y-vertex v in Γ, for a regular value just above a minimum always cuts off a lower disk.

Now consider a regular t_{--} just below t_-. If any outermost simple arc is upper, then an isotopy or edge slide, not increasing the width outside $[t_-, t_+]$ would again reduce the maximal width in $[t_-, t_+]$ to $W-1$. So we conclude that as we descend below t_-, all outermost upper simple arcs disappear, and only lower outermost simple arcs are created. (See Figure 8.) Let w_0 be the vertex of $\Lambda(t_{--})$ corresonding to the intersection of the descending edge from v in Γ with $S(t_{--})$. As t rises through t_-, ends of arcs of $S(t) \cap D$ at w_0 merge in pairs to create new arcs or possibly simple closed curves. But an arc in $\Lambda(t)$ not incident to w_0 is unaffected. If it is outermost simple for one of its ends, it remains so after passing through t_-. Hence we conclude that all outermost simple arcs at t_{--} are incident to w_0.

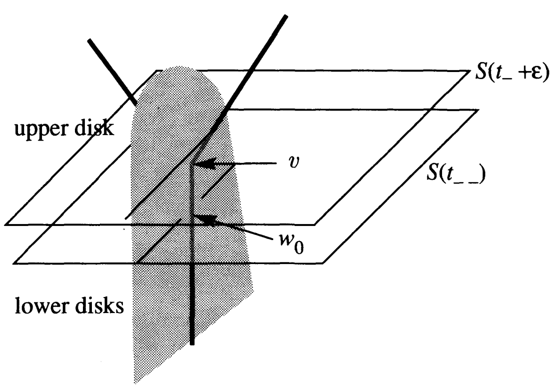


FIGURE 8

This implies in particular that there are no loops in $\Lambda(t_{--})$ at any vertex other than w_0. If the descending edges in Γ from the maximum or λ-vertex v_+ at t_+ are also the ascending edges from v, then they form an unknotted cycle and we are done (see Figure 9a). If one of the lower edges from v_+ coincides with an upper edge from v, let w be the point where the other descending edge e from v_+ first intersects $S(t)$. We have shown that there are no loops in Λ at w and that any outermost edge for w has its other end at w_0 and cuts off a lower disk. Now slide/isotope the rest of the edge e as in 2.2 to an outermost arc α of $\Lambda(t_{--})$ for w, creating an unknotted cycle in Γ (see Figure 9b). Note that in 2.2 a broken edge slide would be required only if w_0 were also on e and e pierced $S(t_{--})$ in the same direction at both w and w_0. Since the latter at least is visibly not the case, no broken edge slide is required.

If both lower edges e and e' from v_+ in Γ intersect $S(t_{--})$, let w and w' be their points of intersection. Again there are no loops in $\Lambda(t_{--})$ at w or w'. Apply 3.9.1 and 3.9.2 to outermost arcs α and α' of $\Lambda(t_{--})$ for w and w' respectively. Again the relevant outermost disks are lower disks, so we have replaced the rest of the edges e and e' by α and α'. This again creates an unknotted cycle in Γ (see Figure 9c).

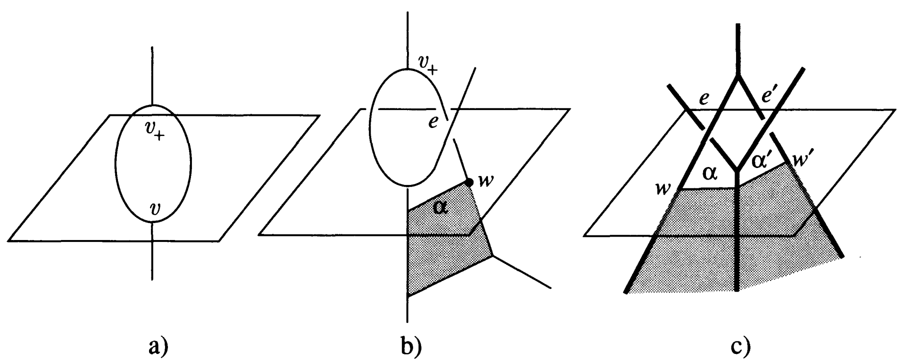


FIGURE 9

References

[1] A. Casson & C. McA. Gordon, *Reducing Heegaard splittings*, Topology Appl. **27** (1987) 275–283.

[2] D. Gabai, *Foliations and the topology of 3-manifolds*. III, J. Differential Geometry **26** (1987) 479–536.

[3] W. Haken, *Some results on surfaces in 3-manifolds*, Studies in Modern Topology, Math. Assoc. Amer., Prentice-Hall, Englewood Cliffs, NJ, 1968, 34–98.

[4] Klaus Johannson, *On surfaces and Heegaard surfaces*, Trans. Amer. Math. Soc. **325** (1991) 573–591 .

[5] W. Menasco & A. Thompson, *Compressing handlebodies with holes*, Topology **28** (1989) 485–494 .

[6] J.-P. Otal, *Sur les scindements de Heegaard de la sphere* S^3 , Topology **30** (1991) 249–258 .

[7] F. Waldhausen, *Heegaard-Zerlegungen der 3-Sphäre*, Topology **7** (1968) 195–203 .

UNIVERSITY OF CALIFORNIA, SANTA BARBARA
UNIVERSITY OF CALIFORNIA, DAVIS