ELSEVIER

Topology and its Applications 69 (1996) 251–264

TOPOLOGY AND ITS APPLICATIONS

Least weight injective surfaces are fundamental

Anneke Bart [1], Martin Scharlemann *,[2]

Department of Mathematics, University of California, Santa Barbara, CA 93106, USA

Received 8 June 1995; revised 18 October 1995

Abstract

To detect if there is an injective surface in a compact irreducible 3-manifold it suffices to triangulate the manifold and check only the fundamental surfaces (Jaco and Oertel, 1984). Here we show that this is true simply because an injective surface of least weight will be fundamental.

Keywords: Haken manifold; Injective surface; Normal surface; Fundamental surface; Incompressible surface

AMS classification: Primary 57M50, Secondary 57Q15

1. Introduction

In [2] Jaco and Oertel show there is an algorithm to decide if an irreducible 3-manifold is Haken. The critical step is to show that in any closed 3-manifold there is a finite constructible set of surfaces in M so that M contains an injective surface (different from S^2) if and only if one of the members of this finite set is injective.

A central ingredient is Haken's theory of normal surfaces [1]. Haken shows that each normal surface can be constructed from a finite set of "fundamental" surfaces. The main advance in [2, Theorem 2.2], states that if a normal surface F in a closed irreducible manifold is two-sided and incompressible, then either it is fundamental or it can be constructed from surfaces of smaller complexity which are also incompressible.

In this paper we prove that the same result holds if we drop the hypothesis that F is two-sided and replace incompressible with injective. This shows directly that, if M contains an injective surface F, then the fundamental surfaces used to construct F are

* Corresponding author.

[1] Partially supported by a UCSB Graduate Research Mentorship Program grant.

[2] Partially supported by NSF grant DMS 9203522.

SSDI 0166-8641(95)00107-7

also injective. So (as also shown in [2, 3.5] by a somewhat more complicated route) to decide if M is Haken, it suffices to check if any fundamental surface is injective. We also generalize (to nonorientable 3-manifolds) a finiteness result [2, 2.3] on incompressible surfaces in atoroidal 3-manifolds.

Here is an outline: In Section 2 we review the theory of normal surfaces from the viewpoint of [4]. In Section 3 we prove some preliminary results, mostly mild generalizations and reformulations of proofs in [2]. In Section 4 we prove our main theorem, which says that if M is a closed irreducible manifold, $F \subset M$ is an injective minimal weight surface, and $F = F_1 + F_2$ is in reduced form, then both F_1 and F_2 are injective. The proof is modeled on, but extends, [2, 2.2]. Section 5 contains the applications, including the finiteness result: in any closed 3-manifold with no injective tori or Klein bottles, there are at most a finite number of injective surfaces of a given genus.

2. The theory of normal surfaces

Let M be a compact triangulated 3-manifold with a fixed triangulation $\mathcal{T}$. Let T^i denote the i-skeleton. Suppose F is a properly embedded surface in M.

Recall that a surface $F \subset M^3$ is *injective* if $(\mathrm{incl})_* : \pi_1(F) \to \pi_1(M)$ is injective. A surface F is *compressible* if

(1) $F = S^2$ and bounds a 3-ball,

(2) there is an embedded disk $D \subset M$ such that $D \cap F = \partial D$, and ∂D is essential in F.

We now give a brief description of normal surfaces based on a more detailed review in [4]. An isotopy of M is called a *normal isotopy* (with respect to $\mathcal{T}$) if it leaves the various simplices of $\mathcal{T}$ invariant. A properly embedded arc in a 2-simplex σ is called *spanning* if its ends lie on different sides of the triangle. A (simple) closed curve in the boundary $\partial\tau$ of a tetrahedron τ is called a *curve type* of $\mathcal{T}$ if it meets the faces of τ in spanning arcs and meets any given face at most once. A tetrahedron has up to normal isotopy precisely seven curve types. There are four curve types with three sides and three curve types with four sides.

If α is a curve type in τ, and p is a point in the interior of τ, then the cone $p^*\alpha$ of α to p is called a *disk type* of τ. Hence a tetrahedron has up to normal isotopy precisely seven disk types.

$F \subset M$ is a *normal surface* if F intersects each tetrahedron of $\mathcal{T}$ in a (necessarily pairwise disjoint) collection of these disktypes.

A normal surface is determined, up to normal isotopy, by the number of each curve type in which it meets the boundaries of the various tetrahedra. Let $\mathcal{C}_1, \ldots, \mathcal{C}_n$ be an ordering of the curve types. Then the surface F determines (and is itself determined) by an n-tuple $(x_1, \ldots, x_n)$, where x_i denotes the number of representatives of $\mathcal{C}_i$ which F induces in the tetrahedra of $\mathcal{T}$.

If we start with an n-tuple of nonnegative integers, then we can construct a normal surface in M corresponding to this n-tuple if it satisfies the following constraints:

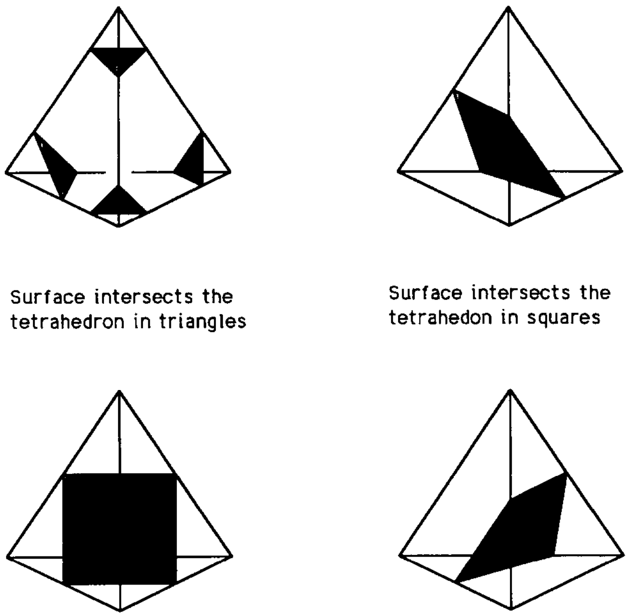

Fig. 1.

(1) We can't have two 4-sided disks from distinct normal isotopy classes in the same tetrahedron.

(2) Edges of disktypes on corresponding faces of incident tetrahedra have to match. Namely, if F intersects one face of a tetrahedron in p representatives of a certain arc type, then F also has to intersect the corresponding face of the incident tetrahedron in p representatives of the same arc type.

A normal surface F in M is *straight* if it satisfies the following conditions:

(1) For any 2-simplex σ in T^2, $\sigma \cap F$ consists only of straight spanning arcs (called *chords*).

(2) In each tetrahedron τ any 3-sided disk in $\tau \cap F$ is the triangle given by the convex hull of its vertices.

(3) Any 4-sided disk in $\tau \cap F$ is the cone to the barycenter of its four vertices.

Clearly any normal surface can be normally isotoped to be straight. Now consider how two straight normal surfaces F_1 and F_2 intersect. First move them slightly so that $F_1 \cap F_2 \cap T^1 = \emptyset$ and so that no barycenter of a 4-sided disk in F_2 lies in F_1 (and vice versa). Then

Lemma 2.1. *In each tetrahedron τ, $F_1 \cap F_2$ consists of proper arcs, each of which has its ends on distinct 2-simplices. Each end is a point in a 2-simplex $\sigma \prec \tau$ where a chord of $F_1 \cap \sigma$ and a chord of $F_2 \cap \sigma$ intersect.*

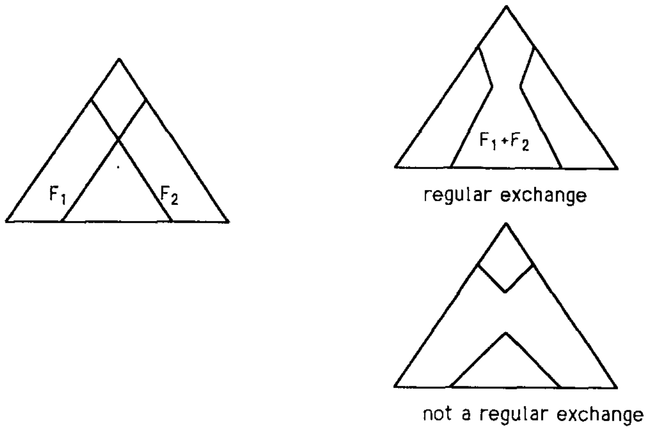

Fig. 2.

Consider how chords in a 2-simplex σ can intersect. Let p be the intersection point. There is a unique way to remove an X neighborhood of p and rejoin the endpoints of the X by two disjoint arcs so that the result gives two spanning arcs in σ. This process is called a *regular exchange* at p. The two opposite quadrants of X which are not connected by this operation are called *good corners* of the resultant spanning arcs at that point. The other two quadrants of X are called *bad corners*.

Now consider extending this regular exchange along an arc component C of $F_1 \cap F_2$ inside a tetrahedron. That is, given two straight disks in a tetrahedron which intersect along an arc C, try to remove a neighborhood of C from both F_1 and F_2 and reattach the sides so that the result is a regular exchange at the ends of C. It is easy to see that this is possible, *unless* the disk types are distinct and both 4-sided.

We say that normal surfaces F_1 and F_2 are *compatible* if, in each tetrahedron, the four-sided curve types of F_1 and F_2 (if any) are the same. If F_1 and F_2 are compatible then, after they are straightened, we have seen that in a neighborhood of each curve of $F_1 \cap F_2$ it is possible to perform a regular exchange to eliminate the curve of intersection. The result of this operation on all intersection curves is a normal surface called the *geometric sum* of F_1 and F_2. Denote this surface by $F_1 + F_2$.

There are several interesting properties which are additive with respect to the geometric sum operation.

If F_1 and F_2 are compatible normal surfaces, then $F_1 + F_2$ is defined and

(1) $\chi(F_1 + F_2) = \chi(F_1) + \chi(F_2)$, where χ is Euler Characteristic,

(2) if F_1 corresponds to $(x_1, \ldots, x_n)$ and F_2 corresponds to $(y_1, \ldots, y_n)$, then $F_1 + F_2$ corresponds to $(x_1 + y_1, \ldots, x_n + y_n)$,

(3) $w(F_1 + F_2) = w(F_1) + w(F_2)$, where $w(F) =$ weight of $F = |F \cap T^1|$.

It is then a theorem of Haken [1] that every normal surface can be built up from a finite set of "fundamental" surfaces by addition (always of compatible surfaces).

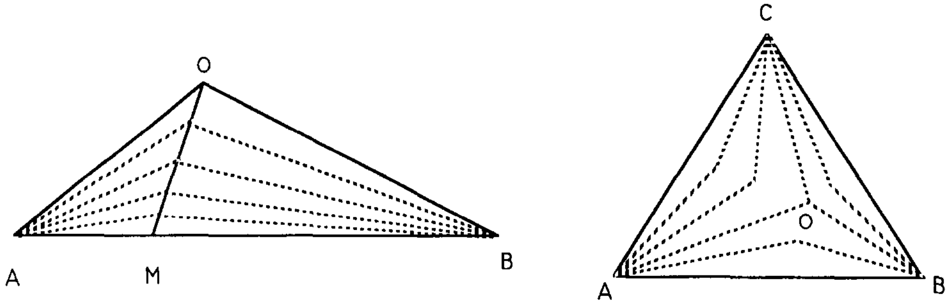

Fig. 3.

We will occasionally need to do a regular exchange along just a subset of $F_1 \cap F_2$ and need to understand the consequences, so we examine spanning arcs in a 2-simplex σ more carefully. Recall that a spanning arc that is a straight line is a *chord*.

Lemma 2.2. *Suppose Γ and Λ are two families of disjoint chords in σ and let Θ be the (not necessarily embedded) collection of arcs obtained by a regular exchange at some of the points in $\Gamma \cap \Lambda$. Then in fact each arc of Θ is an embedded spanning arc.*

Proof. The central idea is the following: In the case of two (but not more) families of disjoint chords it is always possible to find in σ a train-track which carries both families. Even after some regular exchanges, each arc component will remain transverse to the I-fibers of the train-track and so will be an embedded spanning arc.

Here is a more detailed account of this argument, stripped to its elementary core: Observe that the interior of any triangle $\triangle AOB$ can be foliated by proper open intervals, tangent to the sides of the triangle, so that

(1) any chord in $\triangle AOB$ with one end on AB is transverse to the foliation,

(2) any proper arc in $\triangle AOB$ that is transverse to the foliation has precisely one end on AB.

For example one way to construct such a foliation is to choose a point M in the side AB, then in each of $\triangle AOM$ and $\triangle BOM$ choose the fibers of the projection from A and B to OM. Finally, connect the two by rounding corners near OM (see Fig. 3).

Now consider Γ and Λ in $\sigma = \triangle ABC$. Assume, with no loss, that each contains at least one chord between any two sides of $\triangle ABC$. Then for each of Γ and Λ there is a unique hexagonal complementary component in $\triangle ABC$ and these intersect. (This intersection might be empty for more than two families of chords.) Choose O to be a point in the intersection and build the foliation above in each of the three triangles $\triangle OAB$, $\triangle OBC$ and $\triangle OCA$. This produces a proper foliation $\mathcal{F}$ of the interior of $\triangle ABC$, with a 3-pronged singularity at O, so that each chord of Γ and Λ is transverse to $\mathcal{F}$ and any 1-manifold in $\triangle ABC$ that is everywhere transverse to $\mathcal{F}$ is a spanning arc. A regular exchange for Γ and Λ can then be defined locally as that exchange which retains transversality to $\mathcal{F}$. Hence any number of regular exchanges will not destroy transversality, and the resulting arcs are embedded spanning arcs. $\square$

An important application of this lemma is the following result:

Proposition 2.3. *Suppose F_1 and F_2 are normal surfaces in M and $F_1 + F_2$ is defined. Suppose an embedded closed surface T is among the surfaces created from F_1 and F_2 by doing regular exchanges along some subcollection of the collection of curves $F_1 \cap F_2$. Then $w(T) > 0$.*

Proof. The alternative is that T is disjoint from the 1-skeleton, yet must still intersect the 2-skeleton since no intersection curve is entirely contained in a tetrahedron. Let σ be a 2-simplex which T intersects. Then the simple closed curves $T \cap \sigma$ are obtained from the families of disjoint spanning chords $\Gamma = F_1 \cap \sigma$ and $\Lambda = F_2 \cap \sigma$ by regular exchanges at some intersection points. This contradicts Lemma 2.2 above. $\square$

There is also a version of Lemma 2.2 for Γ and Λ compatible sets of disjoint disk types in a tetrahedron.

3. Preliminary results

Theorem 3.1. *Suppose M is a closed irreducible 3-manifold and $F \subset M$ is incompressible. Then F is isotopic to a normal surface of no greater weight.*

The proof can be found in [4].

Theorem 3.2. *Suppose M is a closed irreducible 3-manifold and $F \subset M$. Then*
(1) *F injective $\Rightarrow$ F incompressible,*
(2) *if F is 2-sided then F is injective $\Longleftrightarrow$ F is incompressible,*
(3) *F is injective $\Longleftrightarrow$ $\widetilde{F} = \partial(\eta(F))$ is incompressible.*

Proof. An easy exercise. $\square$

A *patch* for $F_1 + F_2$ is a component of $F_1 \setminus F_2$ or $F_2 \setminus F_1$.

Lemma 3.3. *If a patch is a disk, then its weight is greater than zero.*

Proof. (See also [4, p. 164].) Let $C = \partial D$, where D is a patch. Since $F_1 \cap F_2$ intersects each tetrahedron in proper arcs, $C \cap T^2 \neq \emptyset$. If $D \cap T^1 = \emptyset$ then $D \cap (T^2 \setminus T^1)$ is a collection of proper arcs in D. Consider an outermost arc in D, lying entirely in a 2-simplex σ. Then the segment of C cut off by the outermost arc is a component of $F_1 \cap F_2 \cap$ (tetrahedron) with both ends on the same simplex σ. This contradicts Theorem 2.1. $\square$

Lemma 3.4. *Even if M is nonorientable, a neighborhood of an intersection curve C is orientable. Thus C is either 1-sided in both F_1 and F_2, or 2-sided in both.*

Proof. Arbitrarily pick a surface, F_1 say. Given a point p in C and a vector η_1 normal to C in F_1 at p, there is a unique normal η_2 to C in F_2 at p so that η_1 and η_2 are adjacent to the same good corner. The orientation $\langle\eta_1, \eta_2\rangle$ is independent of the choice of $\pm\eta_1$ since $-\eta_1$ and $-\eta_2$ also abut a good corner. Hence this rule determines a continuous well-defined normal orientation to all of C. □

Corollary 3.5. *If M is not RP^3 and is irreducible, then the boundary of any disk-patch of $F_1 + F_2$ is 2-sided.*

Proof. Suppose D is a disk-patch of $F_1 \setminus F_2$. If ∂D is 1-sided in F_1, then the component of F_1 containing D is RP^2. We also know from Lemma 3.4 that RP^2 is 1-sided. Hence $\eta(RP^2) = RP^3 \setminus B^3$. □

We say a surface $F \subset M$ is of *minimal weight* if it cannot be isotoped to a surface with lower weight. By Theorem 3.1 a minimal weight incompressible surface is normal. A surface is *least weight injective* if it is of lowest weight among all injective surfaces. $F = F_1 + F_2$ is in *reduced form*, if the value of $|F_1 \cap F_2|$ is minimal amongst all normal surfaces F_1' and F_2' isotopic to F_1 and F_2 such that $F = F_1' + F_2'$.

Lemma 3.6. *Let F be a minimal weight normal surface in a closed irreducible 3-manifold M. If $F = F_1 + F_2$ is in reduced form and F is incompressible (respectively injective), then*

(1) *each patch is incompressible (respectively injective),*

(2) *no patch of $F_1 + F_2$ is a disk.*

Proof that (2) implies (1). Suppose that P is a compressible patch of $F = F_1 + F_2$. A compressing disk D for P is also a compressing disk for F unless ∂D bounds a disk D' in F. But even then there is a diskpatch since

$$\sum \chi(\text{patches in } D') = \chi(D') > 0.$$

To prove the injective case, suppose P is a patch which is not injective. As in Theorem 3.2 let $\widetilde{F} = \partial\eta(F)$ and $\widetilde{F}_i = \partial\eta(F_i)$. Note that $\widetilde{F} = \widetilde{F}_1 + \widetilde{F}_2$, and that each double curve of $\widetilde{F}$ corresponds to a double curve of F. We know that the map $p: \widetilde{F} \to F$ is a two-fold cover and that, except over small annulus or Möbius band neighborhoods of $F_1 \cap F_2$, p maps patches to patches. Let D be a compressing disk for $\widetilde{P} = p^{-1}(P)$. Since $\widetilde{F}$ is incompressible $\Rightarrow \partial D$ bounds a disk Δ in $\widetilde{F}$ and $\partial\Delta$ lies entirely in $\widetilde{P}$. There must be double curves in Δ, otherwise ∂D would be inessential in $\widetilde{P}$. Since $\chi(\Delta) > 0$, there is a patch P_0 in $\widetilde{F}$ with $\chi(P_0) > 0$, i.e., P_0 is a disk-patch of $\widetilde{F}$. Then $p(P_0)$ is a disk-patch in F itself.

Proof of (2). Suppose D is a disk-patch of $F_1 \setminus F_2$ which has least weight $w(D)$ (> 0 by Lemma 3.3) among all disk patches in F_1 or F_2. If $C = \partial D$ were 1-sided in F_1 then

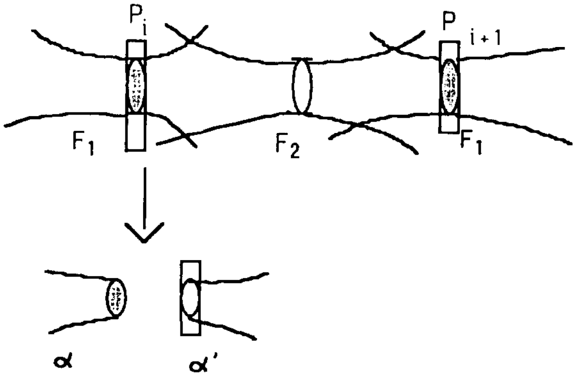

Fig. 4.

$F_1 = RP^2$ and from Corollary 3.5, $M = RP^3$. But F injective in RP^3 implies that F is isotopic to F_1, whereas $w(F) = w(F_1) + w(F_2)$, contradicting F minimal weight. Thus we know that C produces two curves α and α' in F. One bounds the patch D in F_1, the other bounds a disk D' in F, since F is incompressible. D' must contain a disk-patch since $\sum \chi(\text{patches}) = \chi(D') > 0$.

Now $w(D') \leqslant w(D)$ (otherwise trade D' in for D and get a lower weight for F). In particular, the disk-patches in D' have total weight $\leqslant w(D)$. Since $w(D)$ is minimal among disk-patches, and D' must contain at least one disk patch there is exactly one disk-patch in D' and its weight is equal to $w(D)$. Then every other patch in D' must be an annulus of weight zero

Suppose D' itself is a patch (and recall that $w(D) = w(D')$). Construct F from F_1 and F_2 in two steps: first, perform a regular switch on C only (call this $F_1' \cup F_2'$), followed by doing a regular switch along all other double curves (call this final result $F_1' + F_2'$). Note that F_1' and F_2' are F_1 and F_2 with D and D' switched. Since M is irreducible, $F_1' \sim F_1$ and $F_2' \sim F_2$, and note that $F_1' + F_2'$ is in a more reduced form. This contradicts the assumption that F is in reduced form. So D' is not itself a patch. (But D' does contain a disk-patch.)

Let $\{P_1, P_2, \ldots, P_n\}$ be the set of all the disk-patches in F_1 or F_2 that have weight $w(D)$. We found a way to assign to each P_i a P_j: After doing the regular switch $C_i = \partial P_i$ is represented by two curves: α_i which bounds P_i and α_i' which bounds a disk D_i', containing P_j. But there are only finitely many P_i, hence we must cycle back at some point. (Note that we could have a cycle consisting of only one P_i.) In general we will have some cycle $P_1 \to P_2 \to \cdots \to P_k \to P_1$, where $P_i \to P_{i+1}$, $\partial P_i = \alpha_i$ and α_i' is parallel to ∂P_{i+1} in F. Moreover, the annulus A_i in F between α_i' and ∂P_{i+1} has $w(A_i) = 0$.

Let T be the union of the A_i along the α_i, a union of weight zero annuli, hence a torus or Klein bottle T with $w(T) = 0$. But T is a surface obtained by double curve sum on all curves except $C_1, \ldots, C_k$. This contradicts Proposition 2.3. $\square$

4. Main result

Theorem 4.1. *Let M be a closed irreducible manifold. If F is an injective, minimal weight surface in M and $F = F_1 + F_2$ is in reduced form, then F_1 and F_2 are injective.*

Proof. Suppose not, and assume with no loss of generality that F is connected and F_1 is not injective. Then $\widetilde{F_1} = \partial(\eta(F_1))$ is compressible by Theorem 3.2. Let $(D, \partial D) \subset (M \backslash \eta(F_1), \widetilde{F_1})$ be a compressing disk for $\widetilde{F_1}$ chosen so that $| D \cap F_2 |$ is minimal.

We will look at the following 3 cases and show that each leads to a contradiction:

Case 1: $D \cap F_2 = \emptyset$.

Case 2: $D \cap F_2$ contains simple closed curves.

Case 3: $D \cap F_2$ contains only arcs.

Case 1: $D \cap F_2 = \emptyset$. Since ∂D is nontrivial in $\pi_1(F_1)$, it's nontrivial in $F_1 \setminus F_2$, so the patch of F_1 in which ∂D lies is not injective. This contradicts Lemma 3.6.

Case 2: $D \cap F_2$ contains simple closed curves. Pick an innermost simple closed curve, say α, so that α bounds a disk $B \subset D$ and $B \cap F_2 = \alpha$. By Lemma 3.6, ∂B bounds a disk Δ in the patch of F_2 in which it lies. But then by replacing a subdisk of D (possibly B) with an innermost disk of $\Delta \setminus D$ we can lower $|D \cap F_2|$.

We are then in Case 3, where each component of $F_2 \cap D$ is then a proper arc in D. Call the closure of a component of $D \setminus (D \cap F_2)$ a *region* in D. Let γ be a component of $D \cap F_2$. The endpoints of γ (labeled t_0 and t_1) are in $\widetilde{F_1} \cap F_2$. Clearly γ separates a small neighborhood of t_i into two regions, where one region corresponds to a good corner and the other corresponds to a bad corner. (Note: possibly t_0 and t_1 correspond to the same point in $F_1 \cap F_2$ since $\widetilde{F_1}$ double covers F_1.)

Claim 4.1.1. *There is a region in D with at most one bad corner.*

Proof. (From [2] and attributed to Haken.) If there are n spanning arcs, then they cut D into $(n + 1)$ regions. Furthermore, the labeling introduces $2n$ bad corners. Hence at least one region contains less than 2 bad corners. This proves the claim. □

Then Case 3 naturally divides into two subcases:

Subcase 3a: $D \cap F_2$ contains only arcs and some region $E \subset D \setminus F_2$ has only good corners. Let E be a region with only good corners, $\partial E \subset \widetilde{F}$. Then ∂E bounds a disk $\Delta \subset \widetilde{F}$ since $\widetilde{F}$ is incompressible. $\partial E = \partial \Delta$ can be decomposed into arcs alternately lying in $\widetilde{F_1}$ and $\widetilde{F_2}$. So Δ consists of disks from $\widetilde{F_1}$ or $\widetilde{F_2}$ separated by arcs of $\widetilde{F_1} \cap \widetilde{F_2}$. Consider a disk S cut out from Δ by an outermost arc of $\widetilde{F_1} \cap \widetilde{F_2}$. There are naturally two subcases:

(i) The disk S lies in $\widetilde{F_1}$. Let $\beta = S \cap \partial \Delta$ be the subarc of $\partial \Delta$, cut off by the outermost arc γ. Then we can isotope β across S in $\widetilde{F_1}$. This isotopy reduces $|\partial D \cap F_2|$ by two, and hence either reduces $|D \cap F_2|$ by one or creates a closed curve of intersection which we can remove as in Case 2. This contradicts our choice of D.

(ii) The disk S lies in $\widetilde{F_2}$. Let $\beta = S \cap \partial \Delta$ be the subarc of $\partial \Delta$, cut off by the outermost arc γ (γ is an arc in the adjoining patch from $\widetilde{F_1}$).

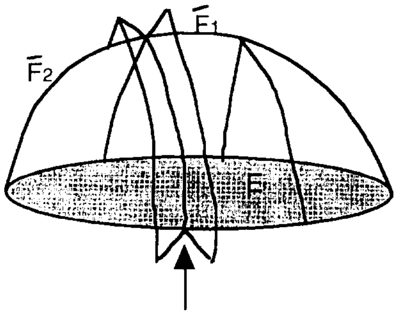

Fig. 5.

If we do a boundary compression of D along S, then one of the new disks we get is a compressing disk for $\widetilde{F}_1$ with fewer components of intersection with F_2. This again contradicts our choice of D.

Subcase 3b: $D \cap F_2$ contains only arcs and some region $E \subset D$ has exactly one bad corner. It will be helpful to consider E in two other contexts. The neighborhood $\eta(F_1)$ can be thought of as the mapping cylinder of the double cover $\widetilde{F}_1 \to F_1$. If we attach to D the annular image of $\partial D \subset \widetilde{F}_1$ in this mapping cylinder, the result is a disk $\overline{D}$ which is embedded except on $\partial \overline{D} \subset F_1$. Similarly E extends to a disk $\overline{E}$ with $\partial \overline{E} \subset F$ a possibly singular curve. The bad corner b in E lies at the end of a spanning arc of $F_2 \cap D$ in ∂D which extends to a spanning arc of $F_2 \cap \overline{D}$. Let α be the curve in $F_1 \cap F_2$ which contains the endpoint of this spanning arc. The regular exchange at α creates one or two curves in F. These bound a Möbius band or annulus B in $M \setminus F$ with α the core of B. After this cut we have $\partial \overline{E} =$ (spanning arc of $B) \cup \beta$, where β is a possibly singular arc in F which is the projection of an embedded arc in $\partial E \cap \widetilde{F}$.

Suppose first that $B \subset M \setminus F$ is a Möbius band, so α is 1-sided in both F_1 and F_2 and B is 1-sided in M (see Lemma 3.4). Consider the corresponding regular exchanges of $\widetilde{F}_1$ and $\widetilde{F}_2$ near α which give rise to $\widetilde{F}$. All this can be understood locally: The single curve α lifts to two curves of intersection of $\widetilde{F}_1$ and $\widetilde{F}_2$ each giving rise to two curves in $\widetilde{F}$. One of these curves bounds a copy of the Möbius band B in $M \setminus \eta(\widetilde{F})$ (we will continue to call it B) and two of the others bound an annulus $\widetilde{B}$ which is the boundary of a regular neighborhood $\eta(B)$ of B in $M \setminus \eta(\widetilde{F})$.

Now $\widetilde{B}$ is ∂-compressed to $\widetilde{F}$ via $\widetilde{E} = E \setminus (\eta(F_2) \cup \eta(B))$. The result is a disk whose boundary lies on $\widetilde{F}$. Since $\widetilde{F}$ is incompressible, and M is irreducible, we can conclude that $\widetilde{B}$ is parallel in $M \setminus \eta(B \cup F)$ to an annulus on $\widetilde{F}$. It follows that $\widetilde{F}$ in fact bounds a solid torus (whose core is $\alpha \subset \eta(B)$). This contradicts the incompressibility of $\widetilde{F}$.

So B must be a 2-sided annulus. Just as above, a copy of B lies in $M \setminus \eta(\widetilde{F})$ and, via $\widetilde{E}$, is parallel to an annulus A in $\widetilde{F}$. If A in fact projects homeomorphically to an annulus $\overline{A}$ in F then β is imbedded, $\overline{E}$ is an embedded disk, and $\overline{A}$ is parallel to B in $M \setminus F$. There is then an isotopy of F in M moving $\overline{A}$ to B. This new surface has lower weight

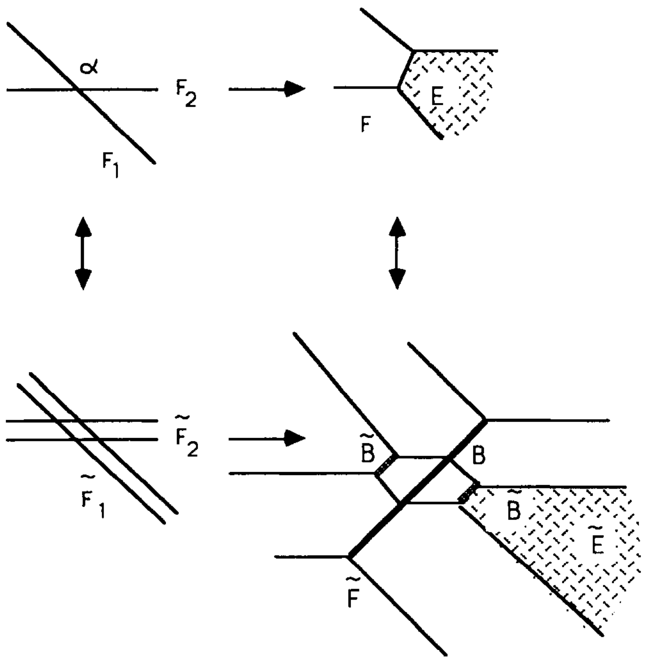

Fig. 6.

than F, since $\overline{A}$ has positive weight (via Lemma 2.2) and B does not, but may not be normal (it may contain a fold). But when it is isotoped to have minimal weight, it will be both normal (by Theorem 3.1) and still have lower weight than F. This contradicts the hypothesis. We may therefore assume that $A \subset \widetilde{F}$ does not project injectively to its image in F.

Claim 4.1.2. *The annulus A is divided up into patches $P_1, \ldots, P_n$ which are annuli.*

Proof. $\chi(A) = 0 \Rightarrow \sum_{i=1}^{n} \chi(P_i) = 0$. But there are no disk-patches in $\widetilde{F}$ (by Lemma 3.6), hence $\chi(P_i) \leqslant 0$ for all i. Hence all P_i must be annuli. □

Next consider how $\partial\widetilde{E}$ crosses A:

Claim 4.1.3. *$\partial\widetilde{E}$ intersects each patch of A in a spanning arc.*

Proof. If not, we can apply the proof of Subcase 3a. □

Continue to let $\overline{A}$ denote the image of A in F. Each patch of A covers a patch in F either homeomorphically or as a two-fold cover, so $\overline{A}$ is the union of annuli and Möbius bands. If the projection $p: A \to \overline{A}$ is not injective over any patch in $\overline{A}$ then $\overline{E}$ abuts both sides of the patch.

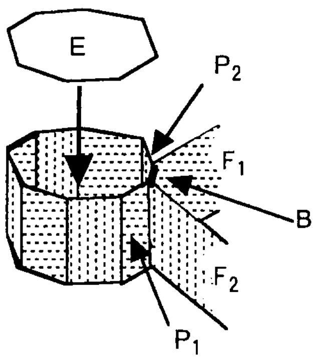

Fig. 7.

Claim 4.1.4. *The projection $p: A \to \overline{A}$ is injective over any patch in $\overline{A}$ which is not adjacent to $\partial \overline{A}$.*

Proof. Let $\overline{P}$ be a patch in $\overline{A} \cap F_1$, say, and suppose $\overline{P}$ is not adjacent to $\partial \overline{A}$. Then, since ∂E has only one bad corner, and it's adjacent to ∂A, we know that $\partial \overline{E}$ crosses $\partial \overline{P}$ at good corners. This means that there is a spanning arc γ of $\overline{P}$ and a normal direction to $\overline{P}$ along γ so that the normal vector near each end of γ lies in a good corner. Symmetrically, the other normal direction gives normal vectors at each end that lie in bad corners. The distinction globally defines distinct sides of $\overline{P}$, so $\overline{P}$ is 2-sided and lifts to two patches in $\widetilde{F}$. Note that $\overline{E}$ abuts only the side of $\overline{P}$ on which the good corners lie. Hence one lift of $\overline{P}$ in $\widetilde{F}$ lies in A and the other doesn't. □

There are two patches $P_i \subset \widetilde{F}_i$, $i = 1, 2$, adjacent to ∂A in the annulus A. Let α_i denote the boundary component of A which abuts P_i. Let $\overline{P}_i$ denote the image of P_i in $\overline{A}$.

Claim 4.1.5. *If $p|A$ is not injective over $\overline{P}_i$ then $\overline{P}_i$ is a 1-sided Möbius band double-covered by P_i.*

Proof. In place of the spanning arc γ in the proof of Claim 4.1.4 use $\beta_i = \partial \overline{E} \cap \overline{P}_i$. Each normal direction points into a good corner at one end and a bad corner at the other. In particular this is true of the normal direction pointing into $\overline{E}$. This means that the covering translation must take P_i to itself, since there is no other patch in A which lies over $\overline{P}_i$ and has a spanning arc whose normal points into a good corner at one end and a bad corner at the other. Thus $\overline{P}_i$ is 1-sided in M. Since $\overline{P}_i$ is 1-sided there is a proper isotopy of β_i in $\overline{P}_i$ which carries β_i back to itself but reverses the direction of a normal field to $\overline{P}_i$ along β_i. But directly across $\overline{P}_i$ from the good corner of β_i is a bad corner. This means that the isotopy reversing normal direction will also switch ends of β_i. This means $\overline{P}_i$ has one edge, so it's a Möbius band. □

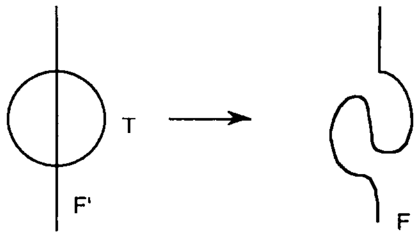

Fig. 8.

We know that $p|A$ is not injective on at least one of the P_i. Suppose that it is injective on P_2 but not P_1 (or vice versa). Then by Claim 4.1.5 $\overline{P}_1$ is a Möbius band and the annulus $A' = A \setminus P_1$ projects homeomorphically to an annulus $\overline{A}'$ in F. Consider what happens if we do a regular exchange along all curves of $F_1 \cap F_2$ except α. Then $\overline{A}'$ is an annulus whose edges are identified at α to give a torus T, $\overline{P}_1$ is a 1-sided Möbius band in the surface F' obtained by switching on all curves but α, the edge of $\overline{P}_1$ lies on α, and the boundary P_1 of a regular neighborhood of $\overline{P}_1$ is parallel to $A' \subset \widetilde{F}$ hence to $\overline{A}' \subset F$. It follows that T bounds a solid torus W containing $\overline{P}_1$. (The meridian of W is the union of two copies of $\overline{E}$ glued together along opposite sides of the spanning arc β_1 of the Möbius band $\overline{P}_1$.) That is, W is just a regular neighborhood of $\overline{P}_1$. It's easy then to see that $F = F' + T$ is isotopic to F' by an isotopy contained in W. By Proposition 2.3 we know $w(T) > 0$ so $w(F') = w(F) - w(T) < w(F)$, contradicting our choice of F.

Now suppose that on both P_i the projection $p|A$ is not injective and let τ denote the covering translation in $\widetilde{F}$. Then $\tau|A$ identifies each P_i to itself, switching the edges since, by Claim 4.1.5, $\overline{P}_i$ is a Möbius band. It follows that $\widetilde{F}$ is the torus $A \cup_\tau \tau(A)$. (Since we assumed, at the beginning, that F is connected.) Then F is the quotient 1-sided Klein bottle obtained by attaching the homeomorphic image $\overline{A}'$ of $A' = A \setminus (P_1 \cup P_2)$ to the two Möbius bands $\overline{P}_i$. Since there are no disk patches, this means that all patches are annuli or Möbius bands. In particular this implies that F_1 also is the union of annuli and Möbius bands and the noninjective component contains the 1-sided Möbius band $\overline{P}_1$. Hence that component is a 1-sided Klein bottle and since it's not injective M is obtained by gluing on a solid torus. Then M is orientable and has a double-cover $\widetilde{M}$ which is just the union of two solid tori, so $\widetilde{M}$ has cyclic fundamental group. Then F couldn't be an injective surface, since the only surface with cyclic fundamental group is RP^2 and if $\chi(F) = \chi(RP^2) = 1$ then some patch of $F = F_1 + F_2$ must be a disk. □

5. Applications

Corollary 5.1. *Let M be a closed Haken 3-manifold. Then, for any triangulation of M, a least weight injective surface is a fundamental surface.*

Proof. Since M is Haken it contains an injective surface. Let F be a least weight injective surface. If F is not fundamental, then by [1] $F = F_1 + F_2$, where each F_i has lower weight. But then by Theorem 4.1, F_1 is injective and has lower weight. □

Corollary 5.2 [2, 4.3]. *There is an algorithm to decide if a compact* 3*-manifold is Haken.*

Proof. Triangulate M. Examine all fundamental surfaces which are 2-spheres, using [5] to determine if each bounds a 3-ball. This is so if and only if M is irreducible. If M is irreducible but not closed, then $\partial M \neq S^2$ if and only if M is Haken. If M is irreducible and closed then use [2, 4.2] (attributed to Haken) to check if any fundamental surface is injective. This suffices by Corollary 5.1. □

A manifold M is *atoroidal* if it contains no injective torus or Klein bottle.

Corollary 5.3. *If M is a closed, irreducible and atoroidal* 3*-manifold, then for any x_0 there are at most a finite number of injective surfaces* (*up to isotopy*) *with $\chi(F) \geqslant x_0$.*

Proof. The proof is by induction. Triangulate M, and note that the corollary is true for (a) $x_0 = 0$ (indeed there are none by assumption) and (b) fundamental surfaces, since there are only a finite number of them.

Assume the corollary is true for $x_0 + 1$. Suppose F is injective and $\chi(F) = x_0$.

(1) If F is fundamental, then it's already on our finite list.

(2) If F is not fundamental, then $F = F_1 + F_2$.

Since F is injective, F_1 and F_2 are injective so $\chi(F_1), \chi(F_2) < 0$. Since $x_0 = \chi(F) = \chi(F_1) + \chi(F_2)$, $\chi(F_1), \chi(F_2) > x_0$, so F_1 and F_2 are already on the finite list. □

References

[1] W. Haken, Theorie der Normalflächen, Acta. Math. 105 (1961) 245–375.

[2] W. Jaco and U. Oertel, An algorithm to decide if a 3-manifold is a Haken manifold, Topology 23 (1984) 195–209.

[3] W. Jaco and J.H. Rubinstein, PL minimal surfaces, J. Differential Geom. 27 (1988) 493–524.

[4] W. Jaco and J.H. Rubinstein, PL equivariant surgery and invariant decomposition of 3-manifolds, Adv. Math. 73 (1989) 149–191.

[5] A. Thompson, Thin position and the recognition problem for S^3, Math. Res. Lett. 1 (1994) 613–630.