ILLINOIS JOURNAL OF MATHEMATICS
Volume 39, Number 3, Fall 1995

ALGORITHMS FOR THE COMPLETE DECOMPOSITION OF A CLOSED 3-MANIFOLD

WILLIAM JACO AND JEFFREY L. TOLLEFSON

0. Introduction

Let F be a properly embedded normal surface in a compact, triangulated 3-manifold M. The *projective class* of F is a rational n-tuple lying in the solution space of a finite linear system of normal equations defined in terms of the triangulation of M. We refer to this compact, convex linear cell in $\mathbf{R}^n$ as the *projective solution space*. A *vertex surface* in M is a connected, two-sided, normal surface whose projective class is a vertex in the projective solution space. We show that the finite collection of vertex surfaces carries a significant amount of information about the topology of M and a variety of interesting surfaces can always be found among the vertex surfaces. The construction of the vertex surfaces is routine and the results we obtain lead to decision and decomposition algorithms based on procedures using vertex surfaces. Among these algorithms are improvements of earlier algorithms of Haken $[H_1]$, $[H_2]$, and Jaco and Oertel [JO].

The theory of normal surfaces was developed by Haken in the early 1960's and he used it to solve a number of decision problems. In this theory each normal surface F corresponds to a unique integral n-tuple $\mathcal{N}_F$ which is a solution to a finite linear system of matching equations. The normal equations are obtained from these matching equations by the addition of a normalizing equation. The projective class of F is the unit vector in the direction of $\mathcal{N}_F$. A *fundamental surface* is a normal surface whose coordinate $\mathcal{N}_F$ is not the sum of two integral solutions to the matching equations and every normal surface can be obtained as a finite sum of fundamental surfaces. There are only a finite number of fundamental surfaces and these can be found algorithmically. Haken's algorithms are generally based on constructing the set of all fundamental surfaces and looking for surfaces from among this set which shed information on the question being considered. In our algorithms it is the vertex surfaces that provide a source of readily constructed surfaces of significance that can be used to carry out certain decision procedures. While all connected vertex surfaces are either fundamental surfaces or doubles of fundamental surfaces we give examples in §3 that show there are many fundamental surfaces which are not vertex sur-

Received November 13, 1990
1991 Mathematics Subject Classification. Primary 57 M99;
Secondary 57 M10.

faces. It is a simpler procedure to list the vertex surfaces than it is to list the fundamental surfaces.

In §4 we give a geometric characterization of those normal 2-spheres and properly embedded disks which are vertex surfaces. Using this characterization, we show in §5 that a non-irreducible, closed 3-manifold with a given triangulation can be completely decomposed by a system $\Sigma = \{F_1, \ldots, F_n\}$ of pairwise disjoint, normal 2-spheres, each of which is a vertex surface. Moreover, this system of 2-spheres can be chosen such that the projective classes of the vertex surfaces in Σ are affinely independent and span an $(n-1)$-dimensional simplex which is a face of the projective solution space. This leads in §7 to an algorithm to decompose a closed triangulated 3-manifold into irreducible 3-manifolds.

In §6 we consider compression disks for the boundary of a compact, irreducible 3-manifold M with compressible boundary. We show that there exists a complete system $\mathscr{D} = \{D_1, \ldots, D_n\}$ of pairwise disjoint, normal, essential compression disks such that each disk D_i is a vertex surface and splitting M along $\mathscr{D}$ yields a 3-manifold with incompressible boundary. In the special case that $M = F \times [-1, 1]$, where F is a compact surface with boundary, we can impose the additional requirement that each ∂D_i meets both ends $F \times \{-1\}$ and $F \times \{1\}$ in an essential arc. As a simple application, we describe an algorithm to decide if a knot K in S^3 is unknotted. Assume S^3 has been triangulated in such a way that K is contained in the 1-skeleton. Let M denote the complement of a regular neighborhood of K with a triangulation of M obtained from subdividing the induced cell decomposition. List the finite set of vertex surfaces in M which are disks and test each such disk D to see if it is essential. This can be done by calculating the Euler characteristic of the components of $\partial M - \partial D$. The knot K is nontrivial if and only if all the vertex disks D tested are inessential.

The first significant result involving vertex surfaces was obtained in [JO]. Suppose F is a least weight, two-sided, incompressible surface in a closed irreducible 3-manifold M. It is shown that if F_1 and F_2 are normal surfaces such that $\mathscr{N}_F = \mathscr{N}_{F_1} + \mathscr{N}_{F_2}$ then both F_1 and F_2 are injective. In particular, every vertex surface in the *face* carrying F is injective. In §6 this theorem is extended to include least weight incompressible, ∂-incompressible surfaces in compact irreducible, ∂-irreducible 3-manifolds with boundary. If $M = F \times [-1, 1]$, where F is a closed surface, then it follows that there exists an essential two-sided annulus which is a vertex and spans the two boundary components. More generally, if F is an essential annulus in a compact, irreducible, ∂-irreducible 3-manifold M then each vertex surface carried by the face of $\mathscr{N}_F$ is either an essential annulus or an essential torus. In view of this theorem we need no additional surfaces besides vertex surfaces to decide whether or not a compact, sufficiently large, irreducible 3-manifold is a product or Seifert fiber space M and for the decomposition of M into its characteristic fibered submanifold and a simple 3-manifold. Complete details of these algorithms are given in §8 and §9.

In §1 we review the basic definitions of normal surface theory. Some of the combinatorics of normal surfaces are discussed in §2. Throughout it is to be understood that a 3-manifold M always comes equipped with a fixed triangulation $\mathscr{T}$ and that a normal surface under consideration is embedded in M and defined relative to this fixed triangulation.

1. Normal surfaces and the projective solution space

Let M denote a compact 3-manifold with a fixed triangulation $\mathscr{T}$ in which there are t tetrahedron. A surface F properly embedded in M is called a *normal surface* (relative to $\mathscr{T}$) if F meets the 2-skeleton $\mathscr{T}^{(2)}$ transversally and meets each tetrahedron Δ in a collection of pairwise disjoint elementary disks. An *elementary disk* in a tetrahedron Δ is a disk that is properly embedded in Δ and is only allowed to intersect a 2-face of Δ in an arc spanning distinct edges of the 2-face as shown in Figure 1.1. A *normal isotopy* of M (relative to $\mathscr{T}$) is an isotopy which is invariant on each simplex of $\mathscr{T}$. We call the normal isotopy class of an elementary disk a *disk type*. The normal isotopy class of the boundary of an elementary disk is called a *curve type*. The normal isotopy class of an arc in which an elementary disk meets a 2-face of Δ is called an *arc type*.

In each tetrahedron Δ there are seven disk types, four of which consist of triangles and three consisting of quadrilaterals. If we fix once and for all an ordering $d_1, \ldots, , \ldots, d_{7t}$ of the disk types in $\mathscr{T}$ then we can assign a $7t$-tuple $\mathscr{N}_F = (x_1, \ldots, x_{7t})$, called the *normal coordinates* of F, to a normal surface F by letting x_i denote the number of elementary disks in F of type d_i. The normal surface F is uniquely determined, up to normal isotopy, by $\mathscr{N}_F$.

Among all $7t$-tuples of non-negative integers $\vec{x} = (x_1, \ldots, x_{7t})$, those corresponding to normal surfaces are characterized by two constraints. The first constraint is that it must be possible to realize the required 4-sided disk types d_i corresponding to nonzero x_i's by disjoint elementary disks. This is equiva-

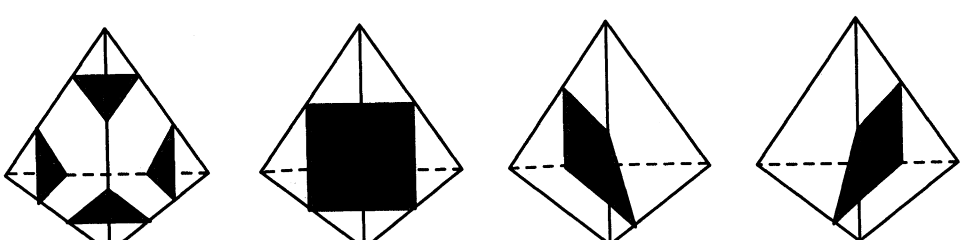


FIG. 1.1 The seven elementary disk types

lent to allowing no more than one 4-sided disk type to be represented in each tetrahedron. The second constraint concerns the matching of the edges of elementary disks along incident 2-faces of tetrahedron. Consider two tetrahedron meeting along a common 2-face and fix an arc type in this 2-face. There are exactly two disk types from each of the tetrahedron whose elementary disks meet this 2-face in arcs of the given arc type. If the $7t$-tuple is to correspond to a normal surface then there must be the same number of elementary disks on both sides of the incident 2-face meeting it in arcs of the given type. This constraint can be given as a system of $6t$ *matching* equations, one equation for each arc type in the 2-simplexes of $\mathcal{T}$ interior to M.

Matching Equations

$$x_i + x_j = x_k + x_1 \tag{1}$$
$$0 \le x_i,\ 1 \le i \le 7t.$$

The non-negative solutions to the matching equations (1) form an infinite linear cone $\mathcal{S}_{\mathcal{T}} \subset \mathcal{R}^{7t}$. A normalizing equation is added to form the system of *normal equations* for $\mathcal{T}$. The solution space $\mathcal{P}_{\mathcal{T}} \subset \mathcal{S}_{\mathcal{T}}$ becomes a compact, convex, linear cell and is referred to as the *projective solution space* for $\mathcal{T}$.

Normal Equations for $\mathcal{T}$

$$x_i + x_j = x_k + x_l$$
$$\sum_{i=1}^{7t} x_i = 1$$
$$0 \le x_i,\ 1 \le i \le 7t. \tag{2}$$

The *projective class* of F, denoted by $\overline{\mathcal{N}_F}$, is the image of $\mathcal{N}_F$ under the projection $\mathcal{S}_{\mathcal{T}} \to \mathcal{P}_{\mathcal{T}}$. If F is a connected normal surface, a typical normal surface corresponding to $\overline{\mathcal{N}_F} \in \mathcal{P}_{\mathcal{T}}$ may consist of normal isotopic copies of a one-sided surface G and normal isotopic copies of a two-sided surface H (where $H = 2G$ if G exists) such that $\overline{\mathcal{N}_G} = \overline{\mathcal{N}_H} = \overline{\mathcal{N}_F}$. A rational point $\vec{z} \in \mathcal{P}_{\mathcal{T}}$ is said to be an *admissible* solution if corresponding to each tetrahedron there is at most one of the quadrilateral variables which is nonzero. Every admissible solution is the projective class of an embedded normal surface.

The *carrier* of a normal surface F, denoted by $\mathcal{C}_{\mathcal{T}}(F)$, is the unique minimal face of $\mathcal{P}_{\mathcal{T}}$ that carries $\overline{\mathcal{N}_F}$. A normal surface G is said to be *supported* by $\mathcal{C}_{\mathcal{T}}(F)$ if $\overline{\mathcal{N}_G} \in \mathcal{C}_{\mathcal{T}}(F)$. Every rational point in $\mathcal{C}_{\mathcal{T}}(F)$ is an admissible solution. In particular, if $\vec{v}$ is a vertex of $\mathcal{C}_{\mathcal{T}}(F)$ then $\vec{v}$ is an admissible solution since it has rational coordinates which are zero in any variable corresponding to a disk type not represented in F. If k is the smallest non-negative integer such that $k\vec{v}$ is integral then we call $k\vec{v}$ a *vertex* *solution* of $\mathcal{S}_{\mathcal{T}}$. An integral solution $\vec{z} \in \mathcal{S}_{\mathcal{T}}$ is a vertex solution if and only if integral multiples of $\vec{z}$ are the only integral points $\vec{x}, \vec{y} \in \mathcal{S}_{\mathcal{T}}$ satisfying an equation of the form $n\vec{z} = \vec{x} + \vec{y}$ for n a positive integer. If F is a connected, two-sided, normal surface such that $\overline{\mathcal{N}_F}$ is a vertex of $\mathcal{P}_{\mathcal{T}}$ then we call F a *vertex surface*. Hence $\mathcal{N}_F$ will be either a vertex solution or twice a vertex solution of $\mathcal{S}_{\mathcal{T}}$. The finite set of vertex surfaces can be explicitly constructed from the system of normal equations using elementary methods of linear algebra.

The vertex surfaces are the basis for numerous algorithms for 3-manifolds since they include so many important and interesting surfaces. For example, a corollary to the following theorem is that if F is a least weight, two-sided, incompressible surface then every rational point in $\mathcal{C}_{\mathcal{T}}(F)$ (including the vertex points) is the projective class of an injective normal surface in M.

THEOREM 1.1 [JO]. *Let M be a closed, irreducible 3-manifold with a triangulation $\mathcal{T}$. Suppose that F is a least weight (or least complexity) normal surface and $F = F_1 + F_2$. If F is two-sided and incompressible then both F_1 and F_2 are injective.*

A consequence is that in order to decide whether or not M contains injective surfaces one has only to check a finite number of vertex surfaces for injectivity.

2. Some combinatorics of normal surfaces

Our model for a normal surface F is one in which F intersects the 2-skeleton of $\mathcal{T}$ transversely and intersects each tetrahedron Δ in linear triangles or quadrilaterals which are the union of two linear triangles. In practice, we often vary from this model up to normal isotopy. Since each elementary disk is determined up to normal isotopy by its vertices in $\mathcal{T}^{(1)}$, a normal surface F is determined by the finite set of points $F \cap \mathcal{T}^{(1)}$. The *weight* of a normal surface F, denoted by $wt(F)$, is defined to be $\sharp(F \cap \mathcal{T}^{(1)})$, the number of intersection points between F and the 1-skeleton of $\mathcal{T}$. The notion of least weight in normal surfaces has played a key role in the work of [JO] and [JR]. We say that a normal surface F is *least weight* if $wt(F)$ is a minimum value for values of $wt(F')$ where F' ranges over normal surfaces isotopic to F. (The range of F' may vary in certain contexts.) Another important measure in working with vertex surfaces is the number of disk types represented by the elementary disks present in F. The *size* $\sigma(F)$ of F is the number of nonzero coordinates in $\mathcal{N}_F$, that is, the number of distinct disk types represented in F. Vertex solutions correspond to local minima relative to size.

When we say that two elementary disks E_1, E_2 in a tetrahedron Δ intersect *transversely* we have in mind the above models with straight edges and linear triangles. In particular, each component of $E_1 \cap E_2$ should be an arc α properly embedded in Δ that spans the interiors of distinct 2-faces of Δ. The intersection $E_1 \cap E_2$ is always connected except possibly when E_1 and E_2 are both quadrilateral disks of the same type, in which case there may be two components. We say that α is a *regular arc* of intersection if there exists a pair of disjoint elementary disks having the same disk types as E_1 and E_2, or equivalently, if the union of the vertices of E_1 and E_2 span a disjoint pair of elementary disks. This is always the case except when E_1 and E_2 are quadrilateral disks of different disk types. Two normal surfaces F and G are said to intersect *transversely* if each pair of elementary disks from F and G, respectively, intersect transversely. Suppose, in addition, that each intersection curve of $F \cap G$ is *regular* in the sense that it is a union of regular arcs. In this case it follows that there is a unique (embedded) normal surface $F + G$, called the *geometric sum* of F and G, determined by the points $(F \cup G) \cap \mathscr{T}^{(1)}$.

This geometric sum of two normal surfaces can also be approached by standard cut-and-paste operations along the regular curves of intersection. Let F_1 and F_2 be two normal surfaces intersecting transversely. We consider the possible cut-and-paste operations along $F_1 \cap F_2$ as viewed locally in a single tetrahedron along the intersections of the elementary disks of the two surfaces. A component α of $F_1 \cap F_2$ is composed of a union of *elementary arcs* arising from the pairwise intersection of elementary disks in F_1 and F_2. We call α a *singular curve* if at least one of the elementary arcs along α arises from the intersection of a pair of 4-sided disks of different disk types. Otherwise, α is called a *regular curve*.

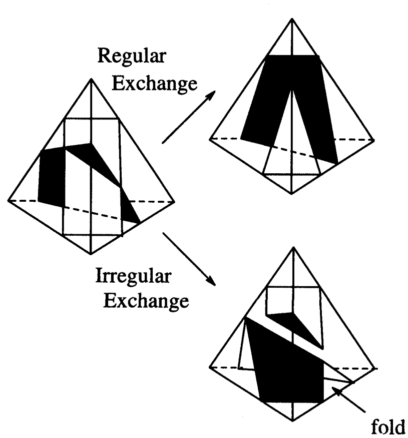


FIG. 2.1 Regular and irregular exchanges

Suppose α is a curve of intersection between F_1 and F_2. Then α meets the 2-skeleton of $\mathcal{T}$ in a finite set of points, each of which can be viewed as the point of intersection between two straight spanning arcs λ_1 and λ_2 in a 2-simplex σ. A component of $\sigma - (\lambda_1 \cup \lambda_2)$ disjoint from the vertices of σ is called a *face-fold* (for α) between λ_1 and λ_2. Let E_1, E_2 be elementary disks in a tetrahedron Δ such that $E_1 \cap E_2$ contains an arc a of α. A *fold* between E_1 and E_2 (along a) is a component V of $\Delta - (E_1 \cup E_2)$ containing a face-fold. If the intersection arc a is regular then each fold V contains two face-folds. If E_1 and E_2 are 4-sided disks of different types then each fold V contains only one face-fold.

Suppose α is a regular intersection curve. As one moves along α, folds between pairs of disks in adjacent tetrahedron must be compatible in that the face-folds created by the edges of the elementary disks in the incident 2-face must coincide. By using the folds to keep track of orientation, one can see that a regular curve α is always orientation preserving in M and therefore has a solid torus or 3-cell regular neighborhood $N(\alpha)$. If we let $A_i = N(\alpha) \cap F_i$ then it follows that A_1 and A_2 are both annuli, both moebius bands, or both disks. There are always two possible ways to define a cut-and-paste operation between F_1 and F_2 along α, although only one of these will preserve (locally) the existing disk types present in $F_1 \cup F_2$. If α is orientation preserving in F_i, then we replace $A_1 \cup A_2$ in $F_1 \cup F_2$ by B, where B is the union of one of two pairs of annuli. In either case, this cut-and-paste operation replaces $F_1 \cup F_2$ by

$$(F_1 - A_1) \cup (F_2 - A_2) \cup B.$$

Viewed locally along an arc a of α in a tetrahedron this corresponds to a normal isotopy defined by pulling one of the elementary disks across a fold along a and thus eliminating a as an arc of intersection. This unique cut-and-paste operation is called a *regular exchange* along α (see Figure 2.1).

A regular exchange does not alter the number of elementary disks of each type in $F_1 \cup F_2$. If every component of $F_1 \cap F_2$ is a regular curve then performing a regular exchange along each component produces the normal surface $F_1 + F_2$, where $\mathcal{N}_{F_1+F_2} = \mathcal{N}_{F_1} + \mathcal{N}_{F_2}$. A useful observation, which follows immediately from our first description, is that the geometric sum on compatible normal surfaces is an associative and commutative operation [JR].

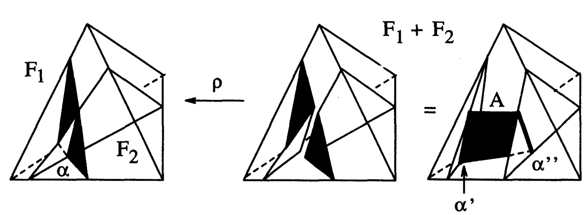


FIG. 2.2 Intersection curve α, trace curves α', α'' and exchange annulus A

Consider a normal surface $F = F_1 + F_2$, where $F_1 \cap F_2 \neq \emptyset$. Each component α of $F_1 \cap F_2$ is a regular curve of intersection along which we perform the above cut-and-paste operation in the formation of F. The identified cut curves along which the components of $(F_1 \cup F_2) - (F_1 \cap F_2)$ are pasted together are referred to as *trace curves*. Corresponding to each component α of $F_1 \cap F_2$ is a single trace curve α' if α is one-sided in both F_1 and F_2, and two trace curves α', α'' if α is two-sided in both F_1 and F_2. One can define an identification map

$$\rho\colon F_1 + F_2 \to F_1 \cup F_2$$

which identifies the trace curves in a (locally) two-to-one fashion. There is a 0-weight annulus, moebius band, or disk band $A \subset N(\alpha)$ spanning the trace curve(s) corresponding to α such that $\rho^{-1}(\alpha) = A$ (assume the identification ρ is defined carefully). The union $\mathscr{A} = \rho^{-1}(F_1 \cap F_2)$ of all such 0-weight surfaces is called a *proper exchange system* of surfaces for the sum $F_1 + F_2$. Given a proper exchange system spanning a normal surface F one can always reconstruct the normal surfaces which sum to F and give rise to the proper exchange system.

Our characterizations of vertex surfaces in Section 4 are formulated in terms of the less restricted notions of *exchange surfaces* and *systems*. The simplest example is a component of a proper exchange system for a sum $F = F_1 + F_2$. More generally, we say that an annulus, moebius band or disk A embedded in M is an *exchange surface* for the normal surface F provided: (1) $fr(A) = A \cap F$, (2) A has an orientable regular neighborhood $N(A)$, and (3) for every tetrahedron Δ, each component of $\Delta \cap A$ is a 0-weight disk L spanning two distinct elementary disks E_1, E_2 of F such that $\partial L = L \cap (E_1 \cup E_2 \cup \partial\Delta)$ and $L \cap E_i$ is an arc joining the interiors of two distinct 2-faces of Δ. An *exchange system* is a finite union of a pairwise disjoint collection of exchange surfaces.

If $\mathscr{A}$ is an exchange system for a normal surface F then we can construct a normal surface S (possibly connected) with one self-intersection curve for each component of $\mathscr{A}$. Each intersection curve is composed of a union of elementary arcs arising from the pairwise intersection of elementary disks in S and such that "regular exchanges" produce F. Two elementary disks in S may intersect many times since we do not necessarily have transverse intersection among the elementary disks. In particular, $S = (F - F \cap N(\mathscr{A})) \cup \mathscr{A}' \cup \mathscr{A}''$, where $\mathscr{A}', \mathscr{A}''$ are two copies of $\mathscr{A}$ in $N(\mathscr{A})$ spanning $fr(F \cap N(\mathscr{A}))$ and intersecting transversely such that for each component A of $\mathscr{A}$, $A' \cap A'' = \alpha$ is the core of A. Thus, as for a proper exchange system, we can define an identification map $\rho\colon F \to S$ which identifies the trace curves in a (locally) two-to-one fashion and such that $\rho^{-1}(\mathscr{A}' \cap \mathscr{A}'') = \mathscr{A}$. The construction of S can be carried out locally in each tetrahedron Δ along one component of $\mathscr{A} \cap \Delta$ at a time and is independent of the order in which it is done. At each step in the construction of $S \cap \Delta$, a cut-and-paste operation is performed on two elementary disks to obtain two new elementary disks of the same disk type(s). Thus, the inverse operation of performing "regular exchanges" on $S \cap \Delta$ along $\mathscr{A}' \cap \mathscr{A}'' \cap \Delta$ to produce $F \cap \Delta$ can also be carried out along one arc of $\mathscr{A}' \cap \mathscr{A}'' \cap \Delta$ at a time. Clearly this is independent of order and each such "regular exchange" between two elementary disks produces two new elementary disks of the same type. Indeed, the final outcome is already completely determined by $S \cap \mathscr{T}^{(1)}$.

Let $\mathscr{A}$ be an exchange system for the normal surface F. A *patch* relative to $\mathscr{A}$ is a connected subsurface $P \subset F$ whose frontier $fr(P)$ consists of trace curves from $\partial\mathscr{A}$ but otherwise P is disjoint from $\mathscr{A}$. One can think of a patch minus its frontier as one of the components of $S - (\mathscr{A} \cap S)$. Let A be a component of $\mathscr{A}$ and consider patches P', P'' each containing a component of ∂A in their frontiers. Let $N(A)$ denote a small regular neighborhood of A such that the closure W of the component of $N(A) - F$ containing $\mathring{A}$ is an I-bundle over A. We say that P' and P'' are *adjacent along* A if $P' \cap W$ and $P'' \cap W$ both meet the same side of A in W. We say that a patch P' *lies on a face-fold* along A if there exists a 2-simplex σ in $\mathscr{T}$ such that $\rho(P' \cap \sigma)$ lies on an innermost face-fold of $S \cap \sigma$ in σ. That is, there exists another patch P'' such that arc components p', p'' of $P' \cap \sigma$, $P'' \cap \sigma$, respectively, each span $A \cap \sigma$ and one edge γ of σ, so that $\partial p' \cup \partial p'' = (p' \cup p'') \cap (A \cup \gamma)$.

Suppose that the disk $D \subset F$ is a patch for F relative to the exchange system $\mathscr{A}$. If ∂D is a simple closed trace curve then $D \subset \mathring{F}$ and we say that D is a *disk patch*. If $D \cap \partial M \neq \emptyset$ then we will only call D a *disk patch* if $D \cap \partial M$ is an arc. This is equivalent to the existence of only one trace curve for D. The following elementary Euler characteristic argument is given in [JR] to show that disk patches cannot have 0-weight. Suppose that D is a patch. The trace curves cut the elementary disks into 2-cell pieces which define a cell decomposition for each patch. Since a 0-weight patch D is the union of such 2-cell pieces not containing any vertices of the elementary disks, we have a cell decomposition of D into 4-, 6-, and 8-sided disks. Using this decomposition to compute the Euler characteristic, it follows that $\chi(D) = -f_8 - \frac{1}{2}f_6 + \frac{1}{2}b$, where f_i denotes the number of i-sided disks in the cell decomposition of D and b is the number of components of $D \cap \partial(M)$. Thus if b is 0 or 1 we must have $\chi(D) \neq 1$.

3. Examples

The normal 2-spheres and disks obtained by taking the link of a vertex in $\mathscr{T}$ are the simplest examples of vertex surfaces. Although a vertex surface is a fundamental surface, the converse is not true. The next two examples illustrate a method to construct fundamental surfaces that are not vertex surfaces.

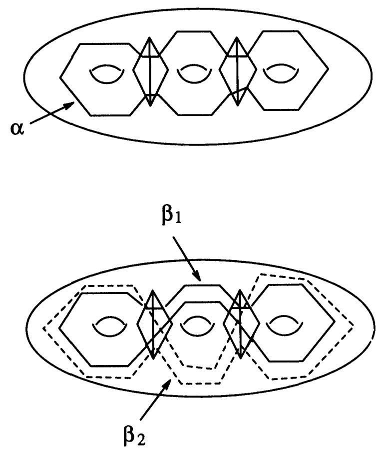


FIG. 3.1 (a) $F = \alpha \times S^1$ (b) $2F = (\beta_1 \times S^1) + (\beta_2 \times S^1)$

Example 3.1. A normal surface F of genus g in $M = T_g \times S^1$ that is a fundamental surface but not a vertex surface.

Let $M = T_g \times S^1$, where T_g is a triangulated surface of genus g. Let τ_1, τ_1', τ_2, and τ_2' denote four triangles in the triangulation such that the only pairwise intersection among them are the disjoint 1-simplexes $e_1 = \tau_1 \cap \tau_1'$ and $e_2 = \tau_2 \cap \tau_2'$. Let α denote an essential, normal, simple closed curve in T_g that is the union of elementary arcs as depicted in Figure 3.1(a). We require that α meet each of the triangles τ_1, τ_1', τ_2, and τ_2' in two elementary arcs of distinct arc types with each having one end point on e_1 or e_2 and that these are the only triangles in the triangulation meeting α in more than one arc. View $M = T_g \times S^1$ as the union of two copies of $T_g \times I$ and let $\mathscr{T}$ be a triangulation of M obtained by triangulating the induced cell complex structure without introducing new vertices. Let F denote the normal surface $\alpha \times S^1$ in M. With a properly chosen order of the disk types we have $\mathscr{N}_F = (1, \ldots, 1, 0, \ldots, 0)$.

A key property possessed by F is that it meets each 2-simplex in $\mathscr{T}$ in a single elementary arc except for those 2-simplexes which lie along the two annuli $A_i = e_i \times S^1$, $i = 1, 2$. Observe that F meets each 2-simplex of $A_1 \cup A_2$ in two arcs of the same arc type. Each tetrahedron having a face in $A_1 \cup A_2$ intersects F in one elementary 3-sided disk and one elementary 4-sided disk. Substituting $\mathscr{N}_F$ into the matching equations gives equations of the forms $0 = 0$, $1 = 1$, and $2 = 2$. To see that $\mathscr{N}_F$ is a fundamental solution, observe that if any nonzero coordinate in $\mathscr{N}_F$ is changed to 0 then the matching equations force the remaining 1's to be 0.

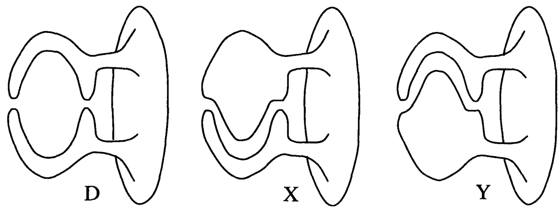


FIG. 3.2 A disk D with $2D = X + Y$

The normal surface $2F$ can be expressed as the sum of two surfaces F_1 and F_2 carried by proper faces of $\mathscr{C}_{\mathscr{T}}(F)$. To construct F_1 and F_2, take two normal simple closed curves β_1, β_2 as shown in Figure 3.1(b). If we let $F_i = \beta_i \times S^1$ then we obtain normal surfaces with the property that $2F = F_1 + F_2$. Therefore F is a fundamental surface but is not a vertex surface. □

Example 3.2. A least weight essential compression disk D that is a fundamental solution but is not a vertex solution.

Using the method of Example 3.1, Figure 3.2 suggests how to construct examples of least weight essential compression disks D where $2D = X + Y$. □

4. A characterization of disk and 2-sphere vertex surfaces

We assume throughout that all surfaces are embedded in a 3-manifold M with a fixed triangulation $\mathscr{T}$. Recall that a vertex surface is a connected, two-sided normal surface F where either $\mathscr{N}_F$ is a vertex solution or there exists a one-sided normal surface X such that $\mathscr{N}_X$ is a vertex solution and $F = 2X$. The goal in this section is to find a relatively simple property related to exchange surfaces that characterizes vertex surfaces when they are disks or two-spheres. Let $\mathscr{A}$ be an exchange system of F and let P_1, P_2 be two patches relative to $\mathscr{A}$. We say that P_1 and P_2 are *normal isotopic along* $\mathscr{A}$ if there exists a sequence of compatible normal isotopies of the elementary disks of P_1 leaving $\mathscr{A}$ invariant and carrying P_1 to P_2. It is apparent that P_1 and P_2 must be adjacent along each component of $\mathscr{A}$. At intermediate stages in the deformation of P_1 onto P_2 there may be self-intersections of ∂P_1 in $\mathscr{A}$.

THEOREM 4.1. *A normal two-sphere F is a vertex surface if and only if F has the property that whenever there exists an annulus A which is an exchange surface for F then the two disjoint disks in F bounded by ∂A are normal isotopic along A.*

COROLLARY 4.2. *If a normal two-sphere F is not a vertex surface then* $2F = X + Y$, *where neither X nor Y is a multiple of F.*

THEOREM 4.3. *A properly embedded, normal, compression disk F is a vertex surface if and only if F satisfies the following properties*:

(a) *If there exists an annulus A which is an exchange surface for F then ∂A bounds disjoint disks in F which are normal isotopic along A.*

(b) *If there exists a disk A which is an exchange surface for F then the disjoint disks in F with frontiers in $fr(A)$ are normal isotopic along A.*

Since the proof of Theorem 4.3 is parallel to the proof of Theorem 4.1, we shall omit it.

Example 4.4. A two-sphere F expressed as the sum of two projective planes P_D and P_E.

Consider a two-sphere F for which there exists a moebius band exchange surface A^* spanning F. Let D and E denote the disjoint disks in F bounded by ∂A^*. We have the two projective planes $P_D = D \cup A^*$ and $P_E = E \cup A^*$. Observe that A^* is an exchange system for the sum $F = P_D + P_E$ and we can regard ∂A^* as the trace curve in F corresponding to the one-sided intersection curve $P_D \cap P_E$ (assume that P_D and P_E have been normal isotoped to intersect transversely along a one-sided curve in A^*).

Let us assume that A^* can be chosen such that the two disks D and E in F bounded by ∂A^* are not normal isotopic along A^*. Under this assumption we can show that P_D is not normal isotopic to P_E and thus F is not a vertex surface. Suppose there does exist a normal isotopy from P_D to P_E. Consider a tetrahedron Δ meeting $P_D \cap P_E$. A normal isotopy between connected surfaces must preserve the relative arrangement in Δ of the elementary disks from $P_D \cap \Delta$. Since $P_D \cap P_E$ is a single simple closed curve, it follows that either (i) the normal isotopy can be chosen to leave the intersection curve $P_D \cap P_E$ invariant or (ii) there exists a component D^* of $P_D - (P_D \cap P_E)$ that is sandwiched between two families of parallel elementary disks which are related in pairs by the given normal isotopy. Whenever the latter case occurs, we can define a normal isotopy between D^* and a component E^* of

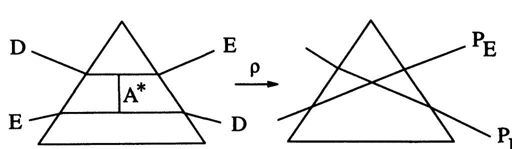


FIG. 4.1 $P_D + P_E$

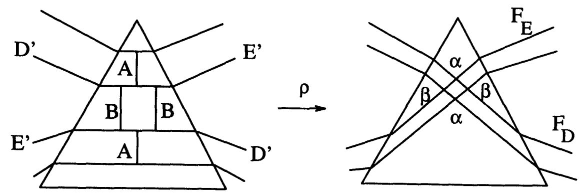


FIG. 4.2 $S_E + S_D$ where $S_D = 2P_D$ and $S_E = 2P_E$

$P_E - (P_D \cap P_E)$ which fixes $P_D \cap P_E$. But D^* is the only component and hence we have that D and E are normal isotopic along A^* to D^* and E^*, respectively. This implies that D and E are normal isotopic along A^*, a contradiction.

There are normal two-spheres arising as the boundaries of regular neighborhoods of the projective planes, namely $F_D = 2P_D$ and $F_E = 2P_E$. We have $2F = F_D + F_E$ and we may assume that there are two intersection curves in $F_D \cap F_E$, say α and β as shown in Figure 4.2. Both α and β intersect each 2-simplex of $\mathcal{T}$ in an even number of points. There is an exchange system $\mathcal{A}$ consisting of two disjoint annuli, denoted by A and B, such that $A \cap 2F = \partial A = \alpha' \cup \alpha''$ and $B \cap 2F = \partial B = \beta' \cup \beta''$ are the trace curves corresponding to α and β, respectively. One of these annuli, say B, is the closure of a component of $\partial(N(A^*)) - F$ for some solid torus regular neighborhood $N(A^*)$ of A^*. The disjoint disks D', E' in F bounded by ∂B are contained in D, E, respectively, and cannot be normal isotopic along B. Hence, it follows from Theorem 4.1 that F is not a vertex. □

LEMMA 4.5. *Suppose that A is an annulus or disk exchange surface for the normal surface F and let D_1, D_2 denote disks in F which are adjacent along A and bounded by $fr(A)$. If $D_1 \subset D_2$ then* $\mathrm{wt}(D_2 - D_1) > 0$ *and hence* $\mathrm{wt}(D_1) < \mathrm{wt}(D_2)$.

Proof. Let $X = \overline{D_2 - D_1}$ and assume that $wt(X) = 0$. Then $A \cup X$ is a 0-weight torus, Klein bottle, annulus or moebius band. Let σ be a 2-simplex and suppose C is an oriented component of $\sigma \cap (A \cup X)$. Observe that C is a simple closed curve which is a union of oriented arcs from $A \cap \sigma$ and $X \cap \sigma$ joined together in an alternating fashion. Let $\{a_1, \ldots, a_n = a_1\}$ denote the components of $A \cap \sigma \cap C$ and let $\{x_1, \ldots, x_n = x_1\}$ be the components of $X \cap \sigma \cap C$. Choose notation so that a_i joins the head of x_i to the tail of x_{i+1} as shown in Figure 4.3. Since D_1 and D_2 are adjacent along A it follows that x_i and x_{i+1} are not adjacent along a_i. Let λ_i be the elementary arc component of $F \cap \sigma$ containing x_i. The orientation on x_i induces an orientation on λ_i. Observe that each pair of elementary arcs λ_i, λ_{i+1} have either both tails or both heads on a common edge of $\partial\sigma$. Think of the direction of x_i as the edge of $\partial\sigma$ on which the head of λ_i lies. As one goes around C one complete revolution, the direction of the x_i must change three

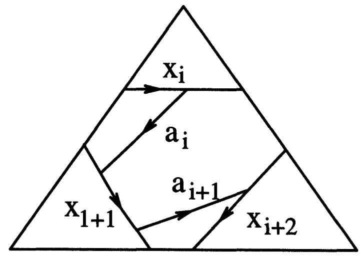


FIG. 4.3 x_i and x_{i+1} have different directions

times. However, it is easy to see that there can be at most one adjacent pair λ_i, λ_{i+1} which do not have the same direction. Thus it is impossible for $\text{wt}(X) = 0$. □

LEMMA 4.6. *Let A be an exchange surface for the normal surface Σ where each component of Σ is a two-sphere. Let D_1, D_2 denote disks with disjoint interiors in Σ bounded by ∂A. If D_1, D_2 are normal isotopic along A then there exists an I-bundle W in M such that W does not contain any vertices, $D_1 \cup D_2$ is the* 0*-bundle of W, and both A and $\mathcal{T}^{(2)} \cap W$ are vertical in W.*

Proof. It is sufficientto observe that there exists a suitable local product structure in each tetrahedron Δ. The normal isotopy in Δ between $D_1 \cap \Delta$ and $D_2 \cap \Delta$ along $A \cap \Delta$ allows one to construct the desired I-bundle structure for $W \cap \Delta$. □

Proof of Theorem 4.1. Suppose there exists an annulus A which is an exchange surface for F such that $\partial A = \alpha_1 \cup \alpha_2$ and α_1, α_2 bound adjacent disks D_1, D_2, respectively, in F that are not normal isotopic along A. If $D_1 \subset D_2$ then we can form the normal surfaces

$$X = D_1 \cup A \cup (F - D_2) \quad \text{and} \quad Y = (D_2 - D_1) \cup A.$$

The annulus A is a proper exchange system for the sum $F = X + Y$. In this case, illustrated in Figure 4.4(a), it is clear that neither X nor Y can be normal isotopic to F since both have a smaller weight than F.

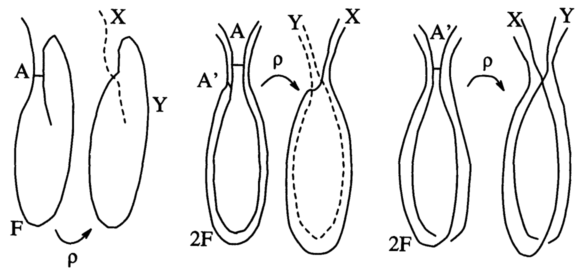
If $D_1 \cap D_2 = \emptyset$ then let $F \times I$ be a small collar on F in M with $F = F \times \{0\}$. We have two cases to consider which are illustrated in Figure 4.4(b) and (c).

First suppose that A meets only one side of F, say $A \cap (F \times I) \subset F \times \{0\}$. Let B be a collar neighborhood of ∂D_1 in $\overline{F - (D_1 \cup D_2)}$ and let $\beta = \partial B - \partial D_1$. Let $A' = \beta \times I \subset F \times I$. We form the normal surfaces

$$X = [(F - (B \cup D_1)) \times \{1\}] \cup A' \cup A \cup [(B \cup D_2) \times \{0\}]$$

and

$$Y = [(F - (B \cup D_2)) \times \{0\}] \cup A' \cup A \cup [(B \cup D_1) \times \{1\}].$$



(a) F = X+Y (b) 2F = X+Y (c) 2F = X+Y

FIG. 4.4 F is not a vertex surface

Then $A \cup A'$ is a proper exchange system for the sum $2F = X + Y$. We may assume, after a normal isotopy of X along A, that $X \cap F = \partial D_1$ and argue as in Example 4.4 that if X and F were normal isotopic then it would follow that D_1 and D_2 are normal isotopic along A. This shows that F is not a vertex.

Now suppose that A meets both sides of F. Let $A' = \overline{A - A \cap (F \times I)}$ and suppose that

$$\partial A' = \partial D_1 \times \{0\} \cup \partial D_2 \times \{1\}.$$

Form the normal surfaces

$$X = [(F - D_1) \times \{0\}] \cup A' \cup [D_2 \times \{1\}]$$

and

$$Y = [(F - D_2) \times \{1\}] \cup A' \cup [D_1 \times \{0\}].$$

Then A' is a proper exchange system for the sum $2F = X + Y$. We may assume, after a normal isotopy of X along $A' \cup (\partial A' \times I)$, that $X \cap F = \partial D_1 \cup \partial D_2$ and argue as above that if X and F were normal isotopic then it would follow that D_1 and D_2 are normal isotopic along A.

We now turn to the proof in the other direction. We assume that F is not a vertex surface and show that there exists an exchange annulus A such that $A \cap F = \partial A$ and the disks with disjoint interiors in F bounded by ∂A are not normal isotopic along A. Since F is not a vertex surface, some multiple of F can be written as the regular sum of normal surfaces which are not multiples of F. Let $\mathscr{A}$ be a proper exchange system for such a sum.

Suppose there exists a moebius band component A of $\mathscr{A}$. Let $\alpha' = \partial A$ where $\alpha' = A \cap F$ and let D_1, D_2 denote the two disks in F bounded by α'. If D_1 and D_2 are not normal isotopic along A then we can find an annulus exchange surface B in the boundary of a regular neighborhood of A, as in Example 4.4, with the disjoint disks D_1', D_2' bounded by ∂B not normal isotopic along B.

So assume that D_1 and D_2 are normal isotopic along A. Then the projective planes $P_1 = D_1 \cup A$ and $P_2 = D_2 \cup A$ are also normal isotopic along A and we can write $F = P_1 + P_2 = 2P$ where $P = P_i$. In this case we show that F can be expressed as a nontrivial sum involving only two-sided intersection curves.

The assumption that F is not a vertex surface, when $F = 2P$ for some one-sided projective plane P, means that we can write $nP = X + Y$ where neither X nor Y are normal isotopic to a multiple of P. We may assume that the number of intersection curves in $X \cap Y$ is minimal relative to all such possible choices of X and Y. Observe that we may further assume X and Y are both connected. For example, suppose $nP = X + Y$ and Y is the disjoint union of Y' and Y''. If $X \cap Y' = \emptyset$ then Y' is normal isotopic to a multiple of P and can be canceled off. If both $X \cap Y' \neq \emptyset$ and $X \cap Y'' \neq \emptyset$ then we can form $W = X + Y'$ and we have $nP = W + Y''$. If $Y'' = kP$ then $(n - k)P = W = X + Y'$. If Y'' is not a multiple of P then we use $nP = W + Y''$. In either case, we have a contradiction to the minimality of the number of intersection curves in $X \cap Y$. Thus, without loss of generality, we may assume that X and Y are connected.

It follows from Euler characteristic considerations that $n \leq 4$. The Euler characteristic is also helpful in analyzing the possible cases. If $n = 4$ then $2F = 4P = X + Y$, where X and Y are two-spheres both distinct from $F = 2P$. If $n = 3$ then one summand, say X, is a two-sphere not equal to $F = 2P$. Hence $3F = 6P = X + (X + 2Y)$ where $X \neq F$. If $n = 1$ or 2 then one summand, say X, must be a two-sphere or a projective plane. If X is a two-sphere then we have $F = 2P = X + Y$ where $X \neq F$. If X is a projective plane then we have $2F = 2X + 2Y$, where $2X$ is a two-sphere distinct from $F = 2P$. Notice that X is a one-sided projective plane since it is contained in the orientable regular neighborhood of F. In all cases there are only two-sided intersection curves between the summands.

We have established that there exist normal surfaces X and Y which are not multiples of F such that $nF = X + Y$ and all intersection curves in $X \cap Y$ are two-sided. We assume that the number of intersection curves in $X \cap Y$ is minimal relative to all possible choices of X and Y in which neither X nor Y is normal isotopic to F and all intersection curves in $X \cap Y$ are two-sided. It follows as before that we may assume X and Y are connected. We let $\mathscr{A}$ denote the proper exchange system for the sum $nF = X + Y$. Our goal is to show that there exists a component A of $\mathscr{A}$ which has the following property: If $A \cap F = \partial A$ then the disks with disjoint interiors in F bounded by ∂A are not normal isotopic along A. If $A \cap F \neq \partial A = A \cap (F \cup F')$, where F' is a copy of F, then there exists an extension A' of A across the product region bounded by $F \cup F'$ such that $A' \cap F = \partial A'$

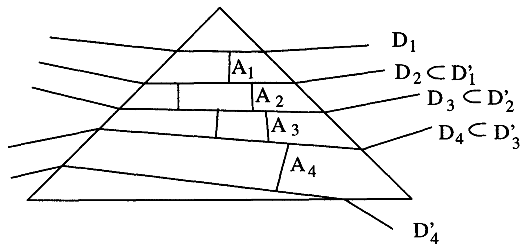


FIG. 4.5 The intersection of a sequence of disk patches with a 2-simplex σ

and the disks with disjoint interiors in F bounded by $\partial A'$ are not normal isotopic along A.

Let $D \subset F$ be a disk patch in nF and let α' denote the trace curve ∂D. Let A denote the annulus in $\mathscr{A}$ such that $\alpha' \subset \partial A$. Let F' be the component of nF containing $\alpha'' = \partial A - \alpha'$ and let $D' \subset F'$ denote the disk bounded by α'' and adjacent to D along A.

Case (1). Suppose that $D \subset D' \subset F$. It follows from Lemma 4.5 that $wt(D) < wt(D')$. Hence neither D and D' nor $F - D$ and $F' - D'$ are normal isotopic along A.

Case (2). Suppose that $D \cap D' = \emptyset$ and neither D and D' nor $F - D$ and $F' - D'$ are normal isotopic along A. If $F = F'$ then there is nothing more to show and so we assume that F and F' are distinct components of nF. In this case A cannot be contained in the product region between F and F', for otherwise D would be normal isotopic to D'. Since F' is a copy of F and A does not cross the product region between F and F', it follows that this product region lies on the side of F opposite that of A. The annulus A can be extended from α'' to a surface A' such that $A' \cap F = \partial A'$ and $A' \cap F' = \alpha''$. Now A' cannot be a moebius band since this would mean that the surface $A' - A$, which spans F and F', is also a moebius band. Hence A' is an annulus and $\partial A'$ bounds a pair of disjoint disks in F. These disks cannot be normal isotopic along A' for otherwise we would have either D' or $F' - D'$ sandwiched in between, thus forcing either D and D' or $F - D$ and $F' - D'$ to be normal isotopic along A and contradicting our assumption for this case.

Case (3). In view of Cases (1) and (2), we may assume that if D is a disk patch then $D \cap D' = \emptyset$ and either D is normal isotopic along A to D' or $F - D$ is normal isotopic along A to $F' - D'$. Since $wt(F) = wt(F)'$, it follows that $wt(D) = wt(D')$ in either case. Our first objective is to show that a disk patch D can be chosen such that D' is also a disk patch.

We begin by choosing a disk patch D_1 that has the least weight among all disk patches for the sum $nF = X + Y$. Consider the disk D'_1 adjacent to D_1 along a component A_1 of $\mathscr{A}$, where $\partial A_1 = \partial D_1 \cup \partial D_2$. If D'_1 is not a patch then we will construct a sequence $\{D_i \subset D'_{i-1}, D'_i, A_i\}$ such that A_i is a component of $\mathscr{A}$ with $\partial A_i = \alpha'_i \cup \alpha''_i$, D_i is a disk patch with $\alpha'_i = \partial D_i$, D'_i is the disk bounded by α''_i and adjacent to D_i along A_i, and $wt(D_i) = wt(D'_i) = wt(D_1)$ (one possible configuration is shown in Figure 4.5). We will show that a disk patch D_n must eventually be reached such that D'_n is also a disk patch, as desired.

Suppose the sequence $D_1, \ldots, D_i$ of least weight disk patches has already been constructed. Let D'_i denote the disk in nF adjacent to D_i with A_i spanning the boundary curves $\alpha'_i = \partial D_i$ and $\alpha''_i = \partial D'_i$. As before, $wt(D'_i) = wt(D_i) = wt(D_1)$. Then either D'_i is the desired disk patch or there exists an annulus A_{i+1} in $\mathscr{A}$ with α'_{i+1} bounding a least weight disk patch $D_{i+1} \subset D'_i$. Since there are only finitely many disk patches, eventually the sequence must either terminate with the desired pair of disk patches or else cycle. We show that it cannot cycle.

Assume the sequence cycles. After relabeling, we may assume that $D_n = D_1$ for some n. We view the sequence in reverse order in a 2-simplex σ for which there exists an arc component d_0 of $D_n \cap \sigma$ with one endpoint in $\partial\sigma$ and the other in $A_n \cap \sigma$. Because $wt(D'_i - D_{i+1}) = 0$, we have $(D'_i - D_{i+1}) \cap \partial\sigma = \emptyset$ and hence the possible configurations in σ are limited. Define d'_1 to be the component of $\sigma \cap D'_{n-1}$ containing d_0 and adjacent in σ to a component of $D_{n-1} \cap \sigma$, denoted by d_1. Continuing in this way, let d_i denote the component of $D_{n-i} \cap \sigma$ adjacent in σ to d'_i and let d'_{i+1} denote the component of $D'_{n-i-1} \cap \sigma$ such that $d_i \subset d'_{i+1}$. When we reach $d_{n-1} \subset D_1 \cap \sigma = D_n \cap \sigma$ then we find a component d'_n of $D_n \cap \sigma$ and this construction of the d_i, d'_i begins to cycle. But this can be shown to be impossible by using an argument similar to that in the proof of Lemma 4.5.

Therefore there exists a pair of disk patches D, D' adjacent along an annulus component A of $\mathscr{A}$. It follows from our assumption on the minimality of the number of intersection curves in $X \cap Y$ that D is not normal isotopic along A to D'. If $F = F'$ then we are done, so suppose that F and F' are distinct components of nF. Since $A \cap nF = \partial A = A \cap (F \cup F')$ and D is not normal isotopic to D', the product region between F and F' is disjoint from A. Hence A can be extended to an annulus A' such that $A' \cap F = \partial A'$. If the two disks D, D'' in F with disjoint interiors bounded by $\partial A'$ were normal isotopic along A' then D and D' would be normal isotopic along A since D'' lies between D and D''. This completes the proof. □

Let A be an orientable exchange surface for a normal surface Σ such that there exist disk patches D_1, D_2 adjacent along A bounded by ∂A and having disjoint interiors. A *bad disk* relative to D_1 is a disk C in a 2-simplex σ such that (i) $\partial C = C \cap (A \cup \Sigma)$ and is the union of four arcs $\gamma_1, \gamma_2, \alpha_1, \alpha_2$ with pairwise disjoint interiors, (ii) $C \cap A = \alpha_1 \cup \alpha_2$, (iii) $C \cap D_1 = \gamma_1$ and $C \cap D_2 = \gamma_2$, (iv) $(\gamma_1 \cup \gamma_2) \cap A = \partial\alpha_1 \cup \partial\alpha_2$, and (iv) $\gamma_2 \subset D_1$.

LEMMA 4.7. *Let A be an annulus or disk exchange surface for a normal surface Σ in which either each component is a 2-sphere or each component is a disk. Suppose that there exist disks D_1, $D_2 \subset \Sigma$ with disjoint interiors, bounded by $fr(A)$, adjacent along A, and not normal isotopic along A. If* $\mathrm{wt}(D_1)$ *is minimal relative to all possible choices of such an exchange surface A then there does not exist a bad disk relative to D_1.*

Proof. If there exists a bad disk C relative to D_1 then we can perform the following cut-and-paste operation on A. Cut A along the arcs $A \cap C$ and paste copies of C on both sides of σ. Use an isotopy that leaves $\Sigma \cup \mathcal{T}^{(1)}$ invariant and removes any newly created spanning disks which meet only one 2-simplex. More precisely, there is an embedded product $N_C = C \times [-1, 1]$ such that $N_C \cap \mathcal{T}^{(1)} = \emptyset$, $C = C \times \{0\}$, $(\gamma_1 \cup \gamma_2) \times [-1, 1] = N_C \cap \Sigma$, and $(\alpha_1 \cup \alpha_2) \times [-1, 1] = N_C \cap A$. We may assume that N_C is chosen such that the number of components of $N_C \cap \mathcal{T}^{(2)}$ is as large as possible. Each such component of $N_C \cap \mathcal{T}^{(2)}$ is a bad disk relative to D_1. Because $\gamma_1 \subset D_1$, it follows that this cut-and-paste operation on A produces a surface $(A - A \cap N_C) \cup (C \times \{-1, 1\})$ having two components, say A' and A'', which are either both annuli or one is an annulus and the other a disk.

Because our choice for N_C contains a maximal number of bad disks relative to A, it follows that both A' and A'' are exchange surfaces. To see this, consider a tetrahedron Δ and a component L of $\Delta \cap A'$ containing the disk(s) $C \times \{i\}$ which are pasted on to form A'. If L meets only one 2-face of Δ, say σ' then $(L \cup A') \cap \sigma'$ contains the boundary of a bad disk C' in σ' and N_C could be enlarged to include C'. Thus each such component L must span two distinct 2-simplexes.

Let D'_1, D'_2 denote the disks in D_1 bounded by $fr(A')$ and let D''_1, D''_2 denote the disks in D_2 bounded by $fr(A'')$. Since $D_1 = D'_1 \cup D''_1 \cup (\gamma_1 \times [-1, 1])$, both D'_1 and D''_1 have smaller weight than D_1. If there exists a normal isotopy along A' between D'_1 and D'_2 and a normal isotopy along A'' between D''_1 and D''_2 then one can easily construct a normal isotopy along $A \cup C$ between D_1 and D_2. Thus, for at least one of A' or A'', say A', the associated disks D'_1, D'_2 are not normal isotopic along A'. But this contradicts the assumption that the weight of D_1 is minimal. □

We next establish some properties related to an exchange surface A for a normal surface Σ when either (1) A is an annulus and Σ a disjoint union of normal two-spheres or (2) A is a disk and Σ a disjoint union of properly embedded normal disks. First we set some notation. Let F_1 and F_2 denote the components of Σ containing $fr(A)$ and let $D_1 \subset F_1$ be a disk patch relative to A. Assume that the disk $D_2 \subset F_2$ adjacent to D_1 along A is not normal isotopic along A to D_1. Let $fr(D_i) = \alpha_i$ and $E_i = \overline{F_i - D_i}$.

If $F_1 \neq F_2$ we form the normal surfaces

$$X_1 = E_1 \cup A \cup D_2 \quad \text{and} \quad X_2 = (\Sigma - (F_1 \cup F_2)) \cup (E_2 \cup A \cup D_1).$$

Then $\Sigma = X_1 + X_2$ and we let $\rho\colon \Sigma \to X_1 \cup X_2$ denote the usual identification map with $\rho^{-1}(X_1 \cap X_2) = A$ (after adjusting X_1, X_2 by a normal isotopy near A). We can also form the normal surface $\Sigma' = (\Sigma - F_2) \cup (E_2 \cup A \cup D_1')$, obtained from Σ by replacing D_2 in F_2 with D_1', a copy of D_1. Let Σ'' denote the normal surface obtained from Σ by replacing all copies of D_2 in Σ with copies of D_1.

If $F_1 = F_2$ and $D_1 \cap D_2 = \emptyset$ then we can form F_1' from F_1 by replacing D_2 by a copy D_1' of D_1 and we can form F_1'' from F_1 by replacing D_1 by a copy D_2' of D_2. Then $2F_1 = F_1' + F_1''$. Hence we can write $2\Sigma = X_1 + X_2$ where $X_1 = (\Sigma - F_1) \cup F_1'$ and $X_2 = (\Sigma - F_1) \cup F_1''$. Here we have the identification $\rho\colon 2\Sigma \to X_1 \cup X_2$. If we let Σ and Σ^* denote the two copies of Σ in 2Σ then we may assume the construction is done such that the exchange system $\rho^{-1}(\Sigma \cap \Sigma^*)$ consists of two annuli (or two disks): A with $fr(A) \subset \Sigma$ and A' spanning Σ and Σ^*. As in the previous case, replacing D_2 with a copy D_1' of D_1 we obtain the normal surface $\Sigma' = (\Sigma - F_1) + F_1' = X_1$ and we obtain Σ'' by replacing all copies of D_2 in Σ with copies of D_1.

LEMMA 4.8. *Let Σ be a disjoint union either of normal two-spheres or of properly embedded, normal disks. Suppose A is an exchange surface for Σ such that $fr(A)$ bounds disk patches D_1 and D_2 which are adjacent along A but not normal isotopic along A. If Σ consists of disks then assume A is an exchange disk. Let F_1, F_2 denote the components of Σ containing D_1 and D_2, respectively. Assume* $\mathrm{wt}(D_1)$ *is minimal relative to all possible choices of the exchange surface A for Σ where A spans the same surfaces F_1, F_2. Then*:

(1) *A is an annulus or a disk.*

(2) *D_1 does not lie on a face-fold along $\rho(A)$ and hence both $E_1 = \overline{F_1 - D_1}$ and $E_2 = \overline{F_2 - D_2}$ lie on a face-fold along $\rho(A)$.*

(3) *If $D_1 \subset D_2$ then* $\mathrm{wt}(D_1) < \mathrm{wt}(D_2)$.

(4) *Suppose $D_1 \cap D_2 = \emptyset$.*

  (a) *If E_1 is not normal isotopic along A to E_2 then* $\mathrm{wt}(D_1) < \mathrm{wt}(E_i)$ *for $i = 1, 2$*

  (b) *If* $\mathrm{wt}(D_1) = \mathrm{wt}(D_2)$ *then for each 2-simplex σ, no component of $A \cap \sigma$ has end-points in two elementary arcs of $\Sigma \cap \sigma$ of the same type in σ.*

  (c) *If* $\mathrm{wt}(D_1) = \mathrm{wt}(D_2)$ *and $\sigma(\Sigma'') = \sigma(\Sigma)$ then there exists an extension of A to a 0-weight annulus or disk A' such that $fr(A')$ bounds D_1 and a disk D_3 in Σ'' adjacent to D_1 along A' and each component of $(A' \cap \Sigma) - fr(A')$ bounds a copy of D_2 in Σ.*

Proof. (1) Assume that A is a moebius band and let F be the component of Σ containing ∂A. Let σ be a 2-simplex meeting A. There exists an innermost face-fold for $\rho(F) \cap \sigma$ in σ on which both D_1 and D_2 lie. Let $N(A)$ be a solid torus regular neighborhood of A such that $N(A) \cap \Sigma$ is an annulus. Then there is an annulus $A' \subset \partial N(A)$ that is an exchange surface for Σ with $\partial A' = \partial(N(A) \cap F)$. The disjoint disks D'_1, D'_2 in F bounded by $\partial A'$ are contained in D_1, D_2, respectively, and are clearly not normal isotopic along A'. Moreover, there is an innermost face-fold in σ relative to A' on which they both lie.

Push A' across the 1-simplex γ of σ on which this face-fold lies to a new annulus $\hat{A}$ where $\partial\hat{A}$ bounds disks $D'_1 \subset D_1$ and $D'_2 \subset D_2$. If $\hat{A}$ is not an exchange annulus then we have the disk $B = \hat{A} - \hat{A} \cap A'$ contained in one tetrahedron such that one face $\hat{\sigma}$ contains $\partial B - (D_1 \cup D_2)$. This process can be repeated as long as the annulus is not an exchange surface as shown in Figure 4.6.

At each stage, the adjacent disks bounded by the boundary curves of the annulus are not normal isotopic along the annulus since D_1 and D_2 are not normal isotopic along A. Thus, this process must terminate with an exchange annulus A'' and disks $D''_1 \subset D_1$ and $D''_2 \subset D_2$. However $\mathrm{wt}(D''_1) < \mathrm{wt}(D_1)$, contradicting our choice of A.

(2) Let σ be a 2-simplex meeting A. There is an innermost face-fold for $\rho(F) \cap \sigma$ relative to A in σ. If D_1, and hence D_2, were to lie along the face-fold then the weight of D_1 could be reduced as in (1) by pushing A across the 1-simplex on which the face-fold lies to obtain a new spanning surface A', contradicting the minimality of $\mathrm{wt}(D_1)$. Therefore, the patches along this face-fold must be contained in $E_1 \cup E_2$.

(3) This follows from Lemma 4.5.

(4) Assume that $D_1 \cap D_2 = \emptyset$. For (a), suppose that $\mathrm{wt}(E_1) \leq \mathrm{wt}(D_1)$. Either E_1 is itself a disk patch or contains D_2. Since the disks E_1 and E_2 lie along a face-fold, we can reduce the weight of E_1 by pushing A across the 1-simplex on which the face-fold lies as in (1). But this would produce a new

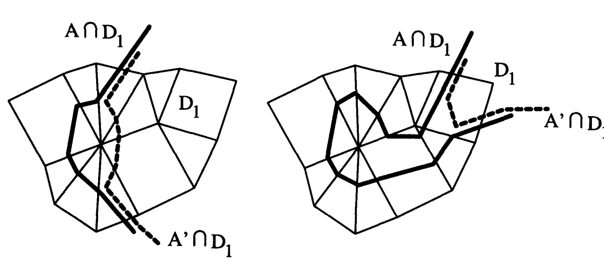


FIG. 4.6 View in F_1 of the push of an exchange annulus A across an edge containing a fold to a new exchange annulus A'

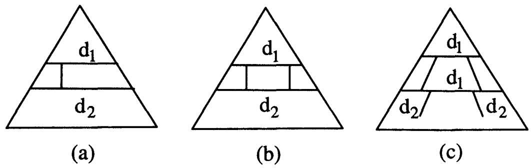


FIG. 4.7 Components d_1 of $D_1 \cap \sigma$ and d_2 of $D_2 \cap \sigma$ contained in arcs of different types.

disk patch (the image under this push of either E_1 or D_2) of weight less than that of D_1, which is impossible. Thus $\text{wt}(E_1) > \text{wt}(D_1)$. Similarly, one shows $\text{wt}(E_2) > \text{wt}(D_1)$.

For (b), assume that $\text{wt}(D_1) = \text{wt}(D_2)$ and consider an arbitrary 2-simplex σ meeting A. Assume there exist components d_1, d_2 of $D_1 \cap \sigma$, $D_2 \cap \sigma$, respectively, which are adjacent along an arc of $A \cap \sigma$ and contained in elementary arcs λ_1, λ_2 of the same type in σ. Since both disk patches are least weight neither can lie on a face-fold. It follows that at least one of the arcs d_1 or d_2 does not meet $fr(\sigma)$. But any component of $A \cap \sigma$ meeting d_1 must lie between λ_1 and λ_2 since $fr(D_1)$ meets just one component of $fr(A)$. There are only two possibilities, either $d_1 \cup d_2$ lies on a bad disk relative to D_1 (Figure 4.7(b)) or else $fr(A) \cap \mathring{d}_2 \neq \emptyset$. If γ is a component of $A \cap \sigma$ meeting $\mathring{d}_2$ then γ cannot lie between λ_1 and λ_2 since $\mathring{d}_1 \cap A = \emptyset$. Hence $\gamma \cap \mathring{d}_2 \subset fr(D_1)$ (as illustrated in Figure 4.7(c)), which implies that $D_1 \subset D_2$ and contradicts our assumption that $D_1 \cap D_2 = \emptyset$.

For (c), assume that $\text{wt}(D_1) = \text{wt}(D_2)$ and $\sigma(\Sigma'') = \sigma(\Sigma)$. We already know that (i) components of $A \cap \sigma$ can only span elementary arcs of $\Sigma \cap \sigma$ having different types, (ii) there does not exist a bad disk relative to D_1, and (iii) D_1 does not lie on a face-fold. It is easy to display all possible configurations in an arbitrary 2-simplex σ meeting A and these are shown in Figure 4.8.

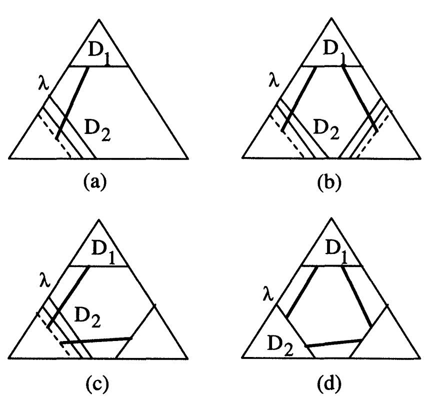


FIG. 4.8 Components of $D_1 \cap \sigma$ and $D_2 \cap \sigma$ adjacent along $A \cap \sigma$

We can immediately rule out the situation shown in Figure 4.8(d) since neither D_1 nor D_2 is adjacent to itself. Let λ denote an elementary arc of $\Sigma \cap \sigma$ containing a component of $D_2 \cap \sigma$ meeting $A \cap \sigma$. In the remaining cases, observe that each time we replace a copy of D_2 by a copy of D_1, one elementary arc of $\Sigma \cap \sigma$ of the same type as λ is eliminated. Since $\sigma(\Sigma) = \sigma(\Sigma'')$, each disk type occurring in Σ must also occur in Σ''. It follows that there remains an elementary arc λ' of $\Sigma'' \cap \sigma$ which is of the same type as λ. Thus there is an extension of the exchange surface A to a 0-weight annulus or disk A' with $fr(A') = fr(D_1) \cup fr(D_3)$, where D_3 is a disk in Σ'' adjacent to D_1 along A'. □

5. A complete system of 2-spheres at the vertices

Let M be a non-irreducible closed 3-manifold with a given triangulation $\mathcal{T}$. Kneser [K] proved that every closed 3-manifold admits a reduction to irreducible 3-manifolds in the following sense. There exists a finite collection $\Sigma = \{F_1, \dots, F_n\}$ of pairwise disjoint 2-spheres in M, called a *complete system of 2-spheres*, such that each component of the 3-manifold $\widehat{(M - \Sigma)}$, which is obtained from $\overline{M - \Sigma}$ by capping off the boundary 2-spheres with 3-balls, is irreducible. We say that Σ is a *minimal complete system* if none of the components of $\widehat{(M - \Sigma)}$ are 3-spheres. In [JR] it is shown that there exists a complete system such that each 2-sphere is a normal surface. In this section we show the existence of a complete system of 2-spheres among the vertex surfaces. More precisely, we prove that there exists a minimal complete system Σ of normal 2-spheres such that the unique face in $\mathcal{P}_{\mathcal{T}}$ carrying Σ is an $(n-1)$-dimensional simplex with vertex set Σ.

Let $\Sigma = \{F_1, \dots, F_n\}$ be a pairwise disjoint collection of 2-spheres in M. We say that a 2-sphere $F \subset M - \Sigma$ is *dependent* on Σ if F bounds a 3-cell in $\widehat{(M - \Sigma)}$. The collection $\Sigma = \{F_1, \dots, F_n\}$ is an *independent* set if no 2-sphere $F_i \in \Sigma$ is dependent on $\Sigma - \{F_i\}$. Thus a minimal complete system is a maximal independent set of pairwise disjoint 2-spheres. If D and E are disks such that $D \cap E = \partial D = \partial E$ then we write $D \sim E$ if $D \cup E$ is a 2-sphere bounding a 3-cell in $\widehat{(M - \Sigma)}$. Suppose Σ is an independent set and D is a disk such that $D \cap \Sigma = D \cap F_1 = \partial D$ and ∂D splits F_1 into two disks E' and E''. Then $D \sim E'$ if and only if the 2-sphere $D \cup E'$ is dependent on $\{F_2, \dots, F_n\}$. It follows that if $D \sim E'$ then $\{E'' \cup D, F_2, \dots, F_n\}$ is an independent set of 2-spheres. In other words, we can modify F_1 by replacing E' with D and not affect the independence of the set Σ (see Figure 5.1).

Consider a system of pairwise disjoint, normal, independent 2-spheres $\Sigma = \{F_1, F_2, \dots, F_n\}$ and let Σ_i denote the subcollection $\{F_1, F_2, \dots, F_i\}$. We say that the system Σ is *efficient* if the following properties are satisfied:

(a) Each F_i is a vertex surface.

(b) Suppose A is an exchange annulus for Σ such that $A \cap \Sigma = \partial A = \alpha_i \cup \alpha_j$, $\alpha_i = A \cap F_i$ and $\alpha_j = A \cap F_j$, where F_i and F_j are distinct compo-

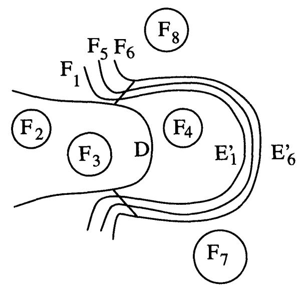


FIG. 5.1 An independent set $\{F_1, \ldots, F_6\}$ of 2-spheres in M

nents of Σ. Let D_i, E_i denote the disks in F_i bounded by α_i and let D_j, E_j denote the disks in F_j bounded by α_j, where notation is chosen such that D_i is adjacent to D_j. Then either D_i is normal isotopic to D_j along A or E_i is normal isotopic to E_j along A.

(c) For each Σ_{i-1}, the 2-sphere F_i has the property that $(\text{wt}(F_i), \sigma(\Sigma_i), \sigma(F_i))$ is minimal relative to all possible choices of 2-sphere vertex surfaces F_i for which Σ_i satisfies (b), where the triples $(wt(F_i), \sigma(\Sigma_i), \sigma(F_i))$ are ordered lexicographically.

LEMMA 5.1. *Let $\Sigma_n = \{F_1, F_2, \ldots, F_n\}$ be an independent system of pairwise disjoint 2-spheres in M defined inductively as follows: If there exists a 2-sphere in $M - \Sigma_i$ that is independent of Σ_i then let F_{i+1} be a normal 2-sphere in $M - \Sigma_i$ such that F_{i+1} is independent of Σ_i and $(\text{wt}(F_{i+1}), \sigma(\Sigma_i \cup \{F_{i+1}\}), \sigma(F_{i+1}))$ is minimal among all such 2-spheres. Then Σ_n is an efficientsystem.*

Proof. The proof is by induction on n. Assume that we already have an independent set of pairwise disjoint 2-spheres Σ_n constructed in the prescribed manner and that Σ_n is efficient.Let F be a 2-sphere in $M - \Sigma_n$ which is independent of Σ_n and chosen such that $(\text{wt}(F), \sigma(\Sigma_n \cup \{F\}), \sigma(F))$ is minimal. We show that $\Sigma = \Sigma_n \cup \{F\}$ satisfies conditions (a) and (b). Keep in mind that $\text{wt}(F) \geq \text{wt}(F_i)$ for $i = 1, \ldots, n$.

We first show that condition (b) holds for Σ. Suppose, to the contrary, that there exists an exchange annulus A for Σ for which condition (b) fails. Let $A \cap F = \alpha$ and $A \cap F_i = \alpha_i$, where $F \neq F_i$. There exists adjacent disks $D \subset F$ and $D_i \subset F_i$ bounded by ∂A such that D is not normal isotopic to D_i and $E = \overline{F - D}$ is not normal isotopic to $E_i = \overline{F_i - D_i}$. We may assume that A and D are chosen such that $\text{wt}(D)$ is minimal relative to all possible choices for A and D. It follows from Lemma 4.8 that $\text{wt}(D) < \text{wt}(E)$.

Suppose that $\text{wt}(D_i) < \text{wt}(D)$. Then $\text{wt}(D_i) < \text{wt}(E)$. The 2-sphere $D \cup A \cup D_i$ has smaller weight than F and can be moved into normal form, if it is not already in normal form, without increasing its weight. It follows from the minimality of $\text{wt}(F)$ that $D \cup A \cup D_i$ is dependent on Σ_n and thus $D \sim D_i$

$\cup A$. If we take $F' = E \cup A \cup D_i$ then F' is a 2-sphere in $M - \Sigma_n$ independent of Σ_n and such that $\mathrm{wt}(F') < \mathrm{wt}(F)$, contradicting the minimality of $\mathrm{wt}(F)$.

Suppose that $\mathrm{wt}(D) < \mathrm{wt}(D_i)$. Then $\mathrm{wt}(E_i) < \mathrm{wt}(E)$ and hence $\mathrm{wt}(E_i \cup A \cup D) < \mathrm{wt}(F_i) \leq \mathrm{wt}(F)$. By the minimality of $\mathrm{wt}(F_i)$, the 2-sphere $E_i \cup A \cup D$ is dependent on Σ_{i-1}. Since E and D_i lie on opposite sides of $D \cup A \cup E_i$, one of these disks must lie in a 3-cell component of $(M - \Sigma_{i-1})^\wedge$. This is impossible since otherwise it would follow that either F_i or F is dependent on Σ_{i-1}.

Suppose that $\mathrm{wt}(D) = \mathrm{wt}(D_i)$. Again $\mathrm{wt}(D \cup A \cup D_i) < \mathrm{wt}(F)$ and as before $D \sim D_i \cup A$. Since D is not normal isotopic along A to D_i, each copy of D in Σ must lie on the side of D opposite that of D_i. We have a product $D \times I \subset M$ containing all copies of D in Σ such that $D \times \partial I \subset \Sigma$. If $F_j \cap (D \times I) \neq \emptyset$ for some j, then there exists a nearest disk component $D'_j \subset F_j$ of $\Sigma \cap (D \times I)$ to which we can extend A to A' such that $\partial A' = A' \cap \Sigma_n = \partial D_i \cup D'_j$. Then A' is an exchange annulus for Σ_n bounding the adjacent disks D_i and D'_j. Now D_i and D'_j cannot be normal isotopic along A' since D lies between them. Similarly, E lying between E_i and $\overline{F_j - D'_j}$ prevents them from being normal isotopic along A. This gives a contradiction to either (a) or (b) and so we have $\Sigma_n \cap D \times I = \emptyset$. In particular, there are no copies of D in Σ_n. We can form the 2-sphere F'' from F by replacing each copy of D with a copy of D_i. Now F'' is independent of Σ_n and $\mathrm{wt}(F'') = \mathrm{wt}(F)$. By our choice of F we must have $\sigma(\Sigma_n \cup F'') = \sigma(\Sigma)$. By Lemma 4.8, the exchange annulus A can be extended to a 0-weight annulus A' which, in this case, is an exchange annulus for Σ_n. Thus $\partial A'$ bounds disks D_i and $D'_k \subset F_k$ adjacent along A'. But this gives us a contradiction as before since (a) or (b) implies either D_i and D'_k are normal isotopic along A' or their complementary disks are. This completes the proof that condition (b) holds.

Now, in order to show F is a vertex surface, we assume that it is not and reach a contradiction. There exists an exchange annulus A spanning F such that disjoint disks C, D and in F bounded by ∂A are not normal isotopic. Among all such instances, we assume that A and the labeling have been chosen such that $wt(D)$ is minimal.

Case (1). $A \cap \Sigma_n = \emptyset$.

It follows from Lemma 4.8 (2) that $\mathrm{wt}(\overline{F - (C \cup D)}) > 0$ and hence $\mathrm{wt}(C \cup A \cup D) < \mathrm{wt}(F)$. If C is not adjacent to D then the 2-sphere $F' = C \cup A \cup D$ is nonseparating in $M - \Sigma_n$ and $\mathrm{wt}(F') < \mathrm{wt}(F)$, contradicting our choice of F. Thus C is adjacent to D. The 2-sphere $C \cup A \cup D$ is inessential in $(M - \Sigma_n)^\wedge$ since it's weight is less than that of F and so $C \cup A \sim D$. The 2-sphere $F' = (F - D) \cup C'$, where C' is a copy of C with $\partial C' = \partial D$, is independent of Σ_n. Thus we must have $\mathrm{wt}(C) = \mathrm{wt}(D)$ for otherwise F' would be a 2-sphere having less weight than that of F.

Let $C \times I, D \times I$ be products in M containing all copies of C, D, respectively, in Σ and such that $(C \cup D) \times \partial I \subset \Sigma$. Suppose that $\Sigma_n \cap C \times I \neq \emptyset$ and $\Sigma_n \cap D \times I \neq \emptyset$. Then we can extend A to an exchange annulus A' for Σ_n with $\partial A'$ bounding the disks $D_j \subset F_j$ and $D'_k \subset F_k$ adjacent along A' and adjacent to D. Since C and D are not normal isotopic along A', it follows from conditions (a) and (b) that $\overline{F_j - D_j}$ and $\overline{F_k - D'_k}$ are disjoint disks which are normal isotopic along A'. But this gives us an I-bundle $\overline{F_j - D_j} \times I$ in which the annulus $\overline{F - (C \cup D)}$ is embedded transverse to the fibers, which is impossible.

Since $\mathrm{wt}(C) = \mathrm{wt}(D)$ and from what we have just shown, we may assume, without loss of generality, that $\Sigma_n \cap D \times I = \emptyset$. This allows us to form the 2-sphere $F'' \subset M - \Sigma$ from F by replacing every copy of D in F by copies of C. Now $\mathrm{wt}(F'') = \mathrm{wt}(F)$ and hence $\sigma(\Sigma_n \cup \{F''\}) = \sigma(\Sigma)$. Using Lemma 4.8 again, it follows that there exists an extension of A across $\partial D \times I$ to a 0-weight annulus A' containing the boundary of a disk D_j in some $F_j \in \Sigma_n$ such that D_j is adjacent (but not normal isotopic) to the last copy $D' = D \times \partial I - D$ of D along an exchange annulus $B \subset A'$ for Σ. Consider the other end of A. We can show that there is a component F_k to which A' can be extended across $\partial C \times I$ to obtain an exchange annulus A'' for Σ_n with $\partial A''$ bounding disks $D_j \subset F_j$ and $D'_k \subset F_k$ which are adjacent along A'' and both adjacent to C and D. If $\Sigma_n \cap C \times I \neq \emptyset$ this is immediate, and if this fails we can apply the same argument used earlier to find D_j. Since C and D are not normal isotopic along A it follows that D_j and D'_k are not normal isotopic along A''. But because Σ_n satisfies conditions (a) and (b), we must have that $\overline{F_j - D_j}$ and $\overline{F_k - D'_k}$ are disjoint disks normal isotopic along A''. As before, this gives us an I-bundle $\overline{F_j - D_j} \times I$ in which the annulus $\overline{F - (C \cup D)}$ is embedded transverse to the fibers, which is impossible. Thus we are led to a contradiction in all situations when $A \cap \Sigma_n = \emptyset$.

Case (2). $A \cap \Sigma_n \neq \emptyset$.
We set the following notation which is illustrated in Figure 5.2. Let

$$A \cap \Sigma = \alpha_1 \cup \alpha_2 \cup \cdots \cup \alpha_{m+1}$$

where $\partial A = A \cap F = \alpha_1 \cup \alpha_{m+1}$ and A is the union of annuli A_i with mutually disjoint interiors and $\partial A_i = \alpha_i \cup \alpha_{i+1}$.

Let $D_1 = D \subset F$ and assume notation is chosen such that $\partial D_1 = \alpha_1$. We inductively define D_{i+1} to be the disk in Σ bounded by α_{i+1} adjacent to the already labeled disk D_i. If $D_i \subset F_k$ then we define $E_i = \overline{F_k - D_i}$. With this notation the disk $C \subset F$ is denoted by either D_{m+1} or E_{m+1}. By analyzing what can happen along consecutive exchange surfaces A_i and A_{i+1}, we will show that either D_i, D_{i+1}, D_{i+2} are normal isotopic disks or E_i, E_{i+1}, E_{i+2} are normal isotopic disks. Using this, along with the facts that no two distinct

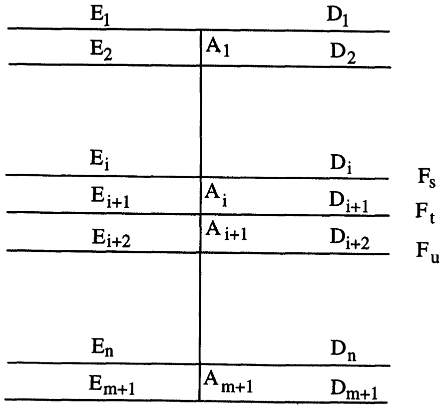


FIG. 5.2 Intersection of Σ with an exchange annulus $A = \cup A_i$

F_i''s are normal isotopic and $D_{m+1} \subset E_1$, it is easy to show that D_i is normal isotopic to D_{i+1} for $1 \leq i \leq m$. This is our desired contradiction.

Each 2-sphere in Σ_n is a vertex surface and we have already established that condition (b) holds for Σ. Thus, for each i we have a normal isotopy along A_i between either D_i and D_{i+1} or E_i and E_{i+1}. We make two more observations that will be used in this proof. By Lemma 4.8, E_1 and E_{m+1} lie on a face-fold along A in some 2-simplex σ. This forces all the E_i to lie along a face-fold in σ. From this it follows that the patch P_i contained in E_i with a boundary curve α_i has non-zero weight.

We are now ready to show that either D_i, D_{i+1}, D_{i+2} or E_i, E_{i+1}, E_{i+2} are disks normal isotopic along A_i and A_{i+1}. We set notation so that $D_i \subset F_s$,

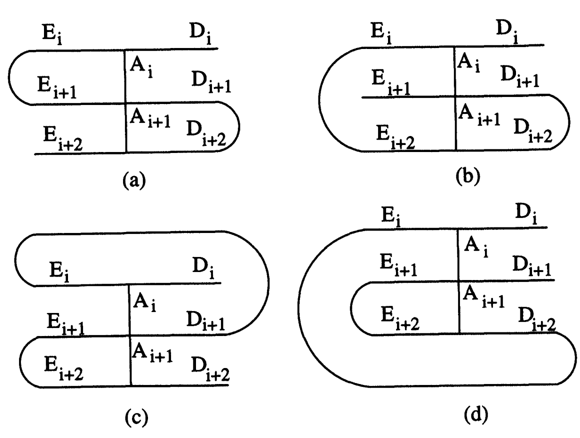


FIG. 5.3 Possible configurations along consecutive exchange annuli A', A'' when $F_s = F_t = F_u$

$D_{i+1} \subset F_t$ and $D_{i+2} \subset F_u$ and consider the various possibilities for the 2-spheres F_s, F_t, F_u in Σ.

(i) Suppose $F_s = F_t = F_u$. We claim this cannot occur. For both $j = i$ and $j = i + 1$, either the pair E_j, E_{j+1} or the pair D_j, D_{j+1} are disjoint disks which are normal isotopic along A_j. Since $\mathrm{wt}(E_j) = \mathrm{wt}(F_s) - \mathrm{wt}(D_j)$, it follows in any case that $\mathrm{wt}(D_i) = \mathrm{wt}(D_{i+1}) = \mathrm{wt}(D_{i+2})$. We may assume that D_i is a component of $F_s - (\alpha_i \cup \alpha_{i+1})$ as the argument for the case when E_i is a component is the same. Under this assumption there are four possible configurations as shown in Figure 5.3. Observe that in each case we have $D_j \subset D_k$ for some pair of the three disks $\{D_i, D_{i+1}, D_{i+2}\}$. We have pointed out that the patch P_j in E_j along α_j has nonzero weight. Since $P_j \subset D_k - D_j$ and $\mathrm{wt}(D_k - D_j) = 0$, we have a contradiction.

(ii) Suppose $F_s = F_t \neq F_u$. We show that either D_i, D_{i+1}, D_{i+2} or E_i, E_{i+1}, E_{i+2} are normal isotopic disks. If D_i and E_{i+1} were disk components of $F_s - \partial A_i$ then we could replace F_s by the nonseparating 2-sphere $D_i \cup A_i \cup E_{i+1}$. Since $P_i \subset F_s - (D_i \cup E_{i+1})$, we have $\mathrm{wt}(F_s - (D_i \cup E_{i+1})) \neq 0$ and the new 2-sphere $D_i \cup A_i \cup E_{i+1}$ would have strictly less weight than that of F_s, which is impossible. Similarly, we cannot have D_{i+1} and E_i as the disk components of $F_s - \partial A_i$. First assume that D_i and D_{i+1} are the disk components of $F_s - \partial A_i$. We must have that D_i is normal isotopic to D_{i+1} along A_i. For if E_i were normal isotopic to E_{i+1} along A_i then the subdisks D_i and D_{i+1} would also be normal isotopic along A_i. If E_{i+1} is normal isotopic to E_{i+2} consider the 2-sphere

$$F' = D_{i+1} \cup A_{i+2} \cup D_{i+2}.$$

It follows that $\mathrm{wt}(F') = \mathrm{wt}(D_{i+1}) + \mathrm{wt}(D_{i+2}) = \mathrm{wt}(D_i) + \mathrm{wt}(D_{i+2}) < \mathrm{wt}(F_u)$ since $F_u = D_{i+2} + E_{i+2}$, E_{i+2} contains D_{i-1} which is normal isotopic to D_i, and $\mathrm{wt}(E_{i+2} - (D_{i-1} \cup D_{i+2})) > 0$. Therefore F' must be dependent on Σ_{u-1} and hence it follows that F_s and F_u are dependent relative to Σ_{u-1}, a contradiction. Thus we have D_{i+1} is normal isotopic to D_{i+2} as desired. For the case when E_i, E_{i+1} are the disk components of $F_s - \partial A_i$ it follows by a similar argument that the disks E_i, E_{i+1}, E_{i+2} are normal isotopic.

(iii) Suppose $F_s \neq F_t = F_u$. The same argument as in (ii) leads to the same conclusion.

(iv) Suppose $F_s \neq F_t \neq F_u \neq F_s$. First assume that D_i is normal isotopic to D_{i+1}. We want to show that D_{i+1} is normal isotopic to D_{i+2}. Suppose that this is not the case and hence E_{i+1} is normal isotopic to E_{i+2}. Consider the 2-sphere $F' = E_i \cup A_i \cup E_{i+1}$. Since D_i is normal isotopic to D_{i+1}, F' can replace F_s in Σ. This, together with the fact that F' lies on a face-fold implies that $\mathrm{wt}(F') > F_s$. Thus $\mathrm{wt}(E_{i+1}) > \mathrm{wt}(D_i) = \mathrm{wt}(D_{i+1})$. On the other hand, the 2-sphere $F'' = D_{i+1} \cup A_{i+1} \cup D_{i+2}$ can replace F_u in Σ and hence $\mathrm{wt}(F'') \geq F_u$. From this we obtain $\mathrm{wt}(D_{i+1}) \geq \mathrm{wt}(E_{i+2}) = \mathrm{wt}(E_{i+1})$, which is a contradiction. A symmetric argument shows that if E_i is normal isotopic to E_{i+1} then E_{i+1} is normal isotopic to E_{i+2}. □

THEOREM 5.2. *Suppose M is a non-irreducible closed 3-manifold with a given triangulation $\mathcal{T}$. Then there exists an efficient minimal, complete system of 2-spheres $\Sigma = \{F_1, \dots, F_n\}$ for M such that the unique face $\mathcal{C}(\Sigma)$ of $\mathcal{P}_T$ carrying Σ coincides with the $(n-1)$-dimensional simplex having vertex set Σ.*

Proof. Let $\Sigma = \{F_1, \dots, F_n\}$ denote the efficient, minimal, complete system of vertex 2-spheres obtained by using Lemma 5.1. Let $[\Sigma]$ denote the convex subset of $\mathcal{P}_T$ spanned by the vertex 2-spheres in Σ. We want to show that Σ is affinelyindependent and that $[\Sigma] = \mathcal{C}(\Sigma)$. It sufficesto show that whenever we have $X + Y = \sum_{i \le k} n_i F_i$ then X and Y are each a disjoint union of F_i's or one-sided projective planes P_i with $2P_i = F_i$, $i \le k$. So suppose that $X + Y = \sum_{i \le k} n_i F_i$.

If there exists a one-sided intersection curve α in $X \cap Y$ then there exists a solid torus neighborhood V of α such that $(X \cup Y) \cap V$ is a pair of moebius bands intersecting in a 1-sided curve. We can form the sum $2X + 2Y = \sum_{i \le k} 2n_i F_i$ with $2X \cap V$ an annulus and $(2X \cup 2Y) \cap V$ a pair of annuli intersecting in two boundary parallel curves. If we show that $2X$ consists of copies of the F_i's then it follows that X consists of copies of the F_i's and P_i''s. Thus, without loss of generality, we may assume that all intersection curves are two-sided. We may also assume that the number of intersection curves in $X \cap Y$ is minimal relative to normal isotopy of X and Y. We suppose that $X \cap Y \neq \emptyset$ and reach a contradiction.

Let $D' \subset F_i'$ denote a least weight disk patch in $\sum_{i \le k} n_i F_i$, with respect to the sum $X + Y$, where F_i' is a copy of F_i and boundary $\partial D' = \alpha'$ is a trace curve corresponding to the intersection curve α. Let $D'' \subset F_j''$ denote the adjacent disk with boundary α'', where F_j'' is a copy of F_j. There is an exchange annulus A with $A \cap \Sigma n_i F_i = \partial A$.

We first observe that $\mathrm{wt}(D') = \mathrm{wt}(D'')$. Since Σ is efficient,either D' is normal isotopic to D'' or $F_i' - D'$ is normal isotopic to $F_j'' - D''$. If D' is normal isotopic to D'' then clearly $\mathrm{wt}(D') = \mathrm{wt}(D'')$. Suppose that $F_i' - D'$ is normal isotopic to $F_j'' - D''$. If $i = j$ then D' is again normal isotopic to D''. If $i \neq j$ then $D' \cup A \cup D''$ is a 2-sphere independent of $\Sigma_k - \{F_j\}$ and hence $\mathrm{wt}(D' \cup A \cup D'' \ge \mathrm{wt}(F_j)$. Since $\mathrm{wt}(D') \le \mathrm{wt}(F_i' - D') = \mathrm{wt}(F_j'' - D'')$ it follows that

$$\mathrm{wt}(D' \cup A \cup D'') \le \mathrm{wt}(F_j'').$$

Thus

$$\mathrm{wt}(D' \cup A \cup D'') = \mathrm{wt}(F_j'') \text{ and } \mathrm{wt}(D') = \mathrm{wt}(F_j'' - D'') \ge \mathrm{wt}(D'').$$

But D is a least weight disk patch and so we have $\mathrm{wt}(D'') = \mathrm{wt}(D')$. Thus, in all cases, the weights of D' and D'' are equal.

In the next lemma we show that there exists a least weight disk patch $D' \subset F_i'$ as above with the additional properties that the disk $D'' \subset F_j''$ adjacent along the exchange annulus A is also a least weight disk patch and the 2-sphere $D' \cup A \cup D''$ lies on a fold. By the previous analysis, either D' is normal isotopic to D'' or the 2-sphere $D' \cup A \cup D''$ has weight equal to that of F_j'' and is independent of $\Sigma_k - \{F_j''\}$. The latter is impossible since an isotopy removing the fold would create a 2-sphere of less weight than that of F_j''. But if D' and D'' are adjacent disk patches that are normal isotopic along A, then the number of components of $X \cap Y$ can be reduced by a normal isotopy. This shows that we had $X \cap Y = \emptyset$ to begin with. □

For future convenience, we broaden the context for next lemma by allowing F to be a properly embedded disk in M, as well as a 2-sphere. Remember that a disk patch D for a sum $X + Y$ is a disk which is a not only a patch but also has the property that ∂D meets ∂M in at most an arc.

LEMMA 5.3. *Let $nF = X + Y$ be a sum such that* (i) *all intersection curves are two-sided and* (ii) *every component A of the proper exchange system $\mathcal{A}$ relative to this sum, where $fr(A) = \alpha' \cup \alpha''$, has the property that if α' is the frontier of a least weight disk patch D' then α'' is the frontier of a disk D'' in nF adjacent to D' such that* $\mathrm{wt}(D'') = \mathrm{wt}(D')$. *Then, if there exists a disk patch relative to $\mathcal{A}$, there exists a pair of least weight disk patches E' and E'' adjacent along a component B of $\mathcal{A}$. Suppose additionally* (iii) *for the disks in* (ii), *if F_i', F_j'' are the components of nF containing D', D'', respectively, then* $\mathrm{wt}(D') = \mathrm{wt}(F_i' - D')$ *and* $\mathrm{wt}(D'') = \mathrm{wt}(F_j'' - D'')$. *Then $E' \cup B \cup E''$ lies on a fold.*

Proof. We let $\{F_1, \ldots, F_n\}$ denote the pairwise disjoint copies of F in nF. Each component of the proper exchange system $\mathcal{A}$ is an annulus or a disk. Let D_1' be a least weight disk patch relative to $\mathcal{A}$ and let A_1 denote the component of $\mathcal{A}$ containing $fr(D_1')$. Assume $D_1' \subset F_1$. Let $D_1'' \subset nF$ be the disk with $\partial D_1'' \subset \partial A_1$ and adjacent to D_1' along A_1. By assumption, $\mathrm{wt}(D_1') = \mathrm{wt}(D_1'')$. If D_1'' is not a disk patch then we will construct a sequence of least weight disk patches leading to a pair of adjacent disk patches.

Suppose we have already found the sequence $D_1', D_2' \subset D_1'', \ldots, D_i' \subset D_{i-1}''$ where each D_j' is a least weight disk patch adjacent along the exchange surface A_j to the disk D_j'. Let D_i'' be the disk in nF with $fr(D_i'') \subset A_i$ and which is adjacent to D_i' along A_i. By hypothesis, $\mathrm{wt}(D_i') = \mathrm{wt}(D_i'')$. If D_i'' is itself not a disk patch then there exists a disk patch $D_{i+1}' \subset D_i''$. Since $\mathrm{wt}(D_{i+1}') \leq \mathrm{wt}(D_i'') = \mathrm{wt}(D_i') = \mathrm{wt}(D_1')$, it follows that D_{i+1}' is also a least weight disk patch and hence $\mathrm{wt}(D_i'' - D_{i+1}') = 0$. This construction either terminates with a pair of adjacent least weight disk patches D_p', D_p'' or else it cycles. But the same kind of argument used in Case (3) of the proof of Theorem 4.1 shows that it does not cycle.

Now assume that condition (iii) holds and the least weight disk patches $D' \subset F_i$, $D'' \subset F_j$ constructed above do not lie on a fold along A. Choose a 2-simplex σ for which an arc component d_0 of $D' \cap \sigma$ has one endpoint in an edge γ of $\partial\sigma$ and the other endpoint in $A \cap \sigma$. Let e denote the component of $D' \cap \sigma$ containing d_0. Since D' and D'' do not lie on a face-fold in σ adjoining γ, it follows that there exists a disk patch contained in $\overline{F_i - D'}$ such that $D' \cap e \cap \gamma \neq \emptyset$. D_1' is also a least weight disk patch because $\mathrm{wt}(D_1') \leq \mathrm{wt}(F_i - D') = \mathrm{wt}(D')$. Repeat the earlier construction, but this time begin using this choice for D_1', and observe that $D' \cup A \cup D''$ acts as a barrier which forces all the disks D_i', D_i'' constructed to meet the edge γ. The construction will end with adjacent least weight disk patches D_p', D_p'' both meeting the edge γ. □

6. Boundary compression disks and injective surfaces

Let M be a compact irreducible 3-manifold. A collection $\{D_1, \dots, D_n\}$ of pairwise disjoint, properly embedded, essential compression disks in M is called a *complete system of disks* for M if each boundary component of the 3-manifold obtained by splitting M along $\cup_{i=1}^{n} D_i$ is incompressible. In this section we prove that there always exists a complete system of essential compression disks occurring as vertex surfaces. We also extend Theorem 1.1, the key result in [JO], by proving that if F is a least weight, incompressible, ∂-incompressible, two-sided normal surface in a compact, irreducible, ∂-irreducible 3-manifold M, then all summands of nF are also incompressible and ∂-incompressible. This provides the necessary essential annuli and tori vertex surfaces which, along with the essential compression disk vertex surfaces, allow us to give algorithms for deciding if a 3-manifold is a product $F \times I$, if two normal surfaces in M are parallel, if a 3-manifold is a Seifert fiber space, and an algorithm for splitting an irreducible 3-manifold along essential annuli and tori into its characteristic Seifert submanifold and simple 3-manifolds. We also use the existence of essential compression disk vertex surfaces to improve on Haken's algorithm (see [JO]) to decide if a surface is injective.

If F is a two-sided surface properly embedded in a 3-manifold M, we let $\sigma_F(M)$ denote the 3-manifold obtained by splitting M along F. We will usually refer to a disk D properly embedded in M such that ∂D does not bound a disk in ∂M as an *essential compression disk*. However, in the context of a product $M = F \times [-1, 1]$, where F is a compact surface with nonempty boundary, we impose an additional restriction on D. Let

$$\partial^- M = (F \times \{-1\}) \cup (\partial F \times [-1, 1-\varepsilon])$$

and

$$\partial^+ M = (F \times \{1\}) \cup (\partial F \times [1-\varepsilon, 1])$$

for some small $\varepsilon > 0$. In this context we say that a disk D properly embedded in M is an *essential compression disk* if $\partial^- M \cap D$ is an essential arc in $\partial^- M$ and $\partial^+ M \cap D$ is an essential arc in $\partial^+ M$.

THEOREM 6.1. *Let $M = F \times [-1, 1]$, where F is a compact surface with nonempty boundary. Assume that $\partial F \times \{1\}$ is contained in the 1-skeleton of the given triangulation $\mathcal{T}$. Then there exists a system $\Sigma = \{D_1, \dots, D_n\}$ of pairwise disjoint, properly embedded, normal, essential compression disks such that each D_i is a vertex surface and $\sigma_\Sigma(M)$ is a 3-cell.*

THEOREM 6.2. *Let M be a compact, irreducible 3-manifold with a compressible boundary. There exists a complete system $\Sigma = \{D_1, \dots, D_n\}$ of normal, essential, compression disks such that each disk is a vertex surface.*

Since the proofs of these two theorems are similar and follow closely the proof of Theorem 5.2, we will give only an outline for the proof of Theorem 6.1. We need some addition definitions parallel to those used in Section 5. Let $\Sigma = \{G_1, \dots, G_n\}$ be a pairwise disjoint collection of essential compression disks in M. If G is a properly embedded essential compression disk in M such that $G \subset M - \Sigma$, we say that G is *dependent* on Σ if G is the frontier of a 3-cell in $\sigma_\Sigma(M)$. The system $\Sigma = \{G_1, \dots, G_n\}$ is *independent* if no disk $G_i \in \Sigma$ is dependent on $\Sigma - \{G_i\}$. We say that Σ is a *minimal complete system* if every component of $\sigma_\Sigma(M)$ is a 3-cell and no proper subcollection of Σ achieves such a decomposition into 3-cells. Thus a minimal decomposition system is a maximal independent set of pairwise disjoint compression disks. Consider a system of pairwise disjoint normal compression disks $\Sigma = \{G_1, G_2, \dots, G_n\}$ and let Σ_i denote the subcollection $\{G_1, G_2, \dots, G_i\}$. We say that the system Σ is *efficient* if the following properties are satisfied:

(a) Each G_i is a vertex surface.

(b_1) Suppose A is an exchange annulus for Σ such that $A \cap \Sigma = \partial A = \alpha_i \cup \alpha_j$, $\alpha_i = A \cap G_i$, and $\alpha_j = A \cap G_j$ where, G_i and G_j are distinct components of Σ. Let D_i, D_j denote the disks in G_i, G_j bounded by α_i, α_j, respectively. Then D_i is normal isotopic to D_j along A.

(b_2) Suppose A is an exchange disk for Σ such that $A \cap \Sigma = fr(A) = \alpha_i \cup \alpha_j$, $\alpha_i = A \cap G_i$ and $\alpha_j = A \cap G_j$, where G_i and G_j are distinct components of Σ. Let D_i, E_i denote the disks in G_i bounded by α_i and let D_j, E_j denote the disks in G_j bounded by α_j, where notation is chosen such that D_i is adjacent to D_j along A. Then there is a normal isotopy along A between either D_i and D_j or E_i and E_j.

(c) For each i, the disk G_i has the property that $(wt(G_i), \sigma(\Sigma_i), \sigma(G_i))$ is minimal relative to all possible choices of compression disk vertex surfaces G_i for which Σ_i satisfies (b_1) and (b_2).

LEMMA 6.3. *Let $\Sigma_n = \{G_1, G_2, \dots, G_n\}$ be an independent system of pairwise disjoint essential compression disks in M defined inductively as follows: If there exists a compression disk in $M - \Sigma_i$ that is independent of Σ_i then let G_{i+1} be a normal compression disk in $M - \Sigma_i$ such that G_{i+1} is independent of Σ_i and* $(\text{wt}(G_{i+1}), \sigma(\Sigma_i \cup \{G_{i+1}\}), \sigma(G_{i+1}))$ *is minimal among all such compression disks. Then Σ_n is an efficientsystem.*

Outline of proof. The proof mimics that of Lemma 5.1, using the characterization of disk vertex surfaces of Theorem 4.3 in place of Theorem 4.1. However, several aspects of the argument are simplified since M is irreducible and because of the following observations. Any disk properly embedded in M whose boundary is disjoint from $\gamma = \partial^- M \cap \partial^+ M$ is necessarily the frontier of a 3-cell in M. Since all exchange disks have zero weight, they are disjoint from the $\partial F \times \{1\} \subset \mathcal{T}^{(1)}$. Thus, if ε is chosen small, we may assume that any exchange disk is disjoint from γ. Moreover, any properly embedded disk must intersect γ in an even number of points.

In the inductive step, an essential compression disk G is chosen such that $G \subset M - \Sigma_n$, G is independent of Σ_n, and $(\text{wt}(G), \sigma(\Sigma_n \cup \{G\}), \sigma(G))$ is minimal. In particular, we have that $\text{wt}(G) \geq \text{wt}(G_i)$, $i = 1, \dots, n$. It remains to show that $\Sigma = \Sigma_n \cup \{G\}$ satisfies conditions (a), (b_1) and (b_2). One first shows that condition (b_1) is satisfied by following the argument in the proof of Lemma 5.1. Since the cutting and pasting is along simple closed curves and does not affect the boundaries of the compression disks, there is little to add to that argument. For condition (b_2), the general argument is the same but one must pay attention to the way the disks constructed intersect $\partial^- M$ and $\partial^+ M$.

To show that the new compression disk G is a vertex surface, we must consider the two cases in the characterization of disk vertex surfaces from Theorem 4.3.

Case (a). A is an exchange annulus. The argument in this case is very close to the corresponding part of the proof of Lemma 5.1 with the simplification that only the one component D is a disk and the other component C is an annulus.

Case (b). A is an exchange disk. The argument follows the outlines of the proof of Lemma 5.1 using intersection arcs instead of simple close curves. Each time a newly constructed disk is claimed to be a compression disk, the various possibilities for its boundary meeting γ must be checked. Aside from this detail, the argument is the same as that for 2-spheres. □

COROLLARY 6.4. *Let M be a compact, irreducible 3-manifold and suppose D is a normal surface that is an essential compression disk for ∂M. Assume that* $(\text{wt}(D), \sigma(D))$ *is minimal among all such essential compression disks in M. Then D is a vertex surface.*

This last result is applied in Section 8 to obtain an elementary algorithm to determine whether or not a knot K is the unknot.

Our next goal is to prepare the way for a series of related algorithms, based entirely on normal vertex surfaces, which will allow us to recognize two-sided incompressible, ∂-incompressible surfaces as well as products, regions of parallelity, and Seifert fiber spaces. The cornerstone of these algorithms is the following extension of Theorem 1.1 to 3-manifolds with boundary.

THEOREM 6.5. *Let M be an irreducible, ∂-irreducible 3-manifold. Suppose F is a least weight normal surface properly embedded in M such that F is not a disk and $nF = F_1 + F_2$. If F is two-sided, incompressible and ∂-incompressible then F_1 and F_2 are each incompressible, ∂-incompressible and not a disk.*

For the proof of the incompressibility of F_i we will closely follow the proof of Theorem 2.2 in [JO], adapting it to normal surfaces relative to a triangulation of M and using weight instead of complexity for the measure on our normal surfaces. The argument for the ∂-incompressibility of F_i proceeds in the same spirit. Without loss of generality, we may assume that the proper exchange system $\mathscr{A}$ contains no moebius bands. For if so, then we may just as well consider the sum $2nF = 2F_1 + 2F_2$ which can be arranged to have no one-sided intersection curves (see Example 4.4). This is a local construction in a solid torus regular neighborhood of each component of $F_1 \cap F_2$. Since $2F_i$ is the boundary of a regular neighborhood of F_i, showing that $2F_i$ is incompressible and ∂-compressible implies that F_i is also. We may also assume that the sum $F_1 + F_2$ is in *reduced form*. By this we mean that $F_1 + F_2$ cannot be written as a sum $F_1' + F_2'$ where F_i' is a normal surface isotopic to F_i in M $(i = 1, 2)$ and $F_1' \cap F_2'$ has fewer components than $F_1 \cap F_2$.

The first step is to prove that each patch is incompressible and ∂-compressible.

LEMMA 6.6. *Let M be an irreducible, ∂-irreducible 3-manifold. Suppose F is a least weight, incompressible, ∂-incompressible, two-sided normal surface properly embedded in M and F is not a disk. If $nF = F_1 + F_2$ is in reduced form and each intersection curve in $F_1 \cap F_2$ is two-sided then each patch of $F_1 + F_2$ is incompressible, ∂-incompressible and there are no disk patches.*

Proof. Suppose the sum $nF = F_1 + F_2$ is in reduced form and let ρ: $nF \to F_1 \cup F_2$ be the usual identification map. Once we prove there do not exist any disk patches, it easily follows (see Lemma 1.1 of [JO]) that each path is incompressible and ∂-incompressible. The presence of patches which are disks meeting ∂M in more than one component does not contradict the conclusion of this lemma since they are not disk patches.

We suppose there exists a disk patch and choose a least weight disk patch D_1. Let notation be chosen such that $\rho(D_1) \subset F_1$ and $fr(D_1) = \alpha_1'$. Let $A_1 = \rho^{-1}(\rho(\alpha_1'))$ be an exchange surface in the proper exchange system for the sum $F_1 + F_2$. The corresponding trace curve $\alpha_1'' = fr(A_1) - \alpha_1'$ is the frontier of a unique second disk $D_1' \subset nF$. It is not hard to check that the disks D_1, D_1' must be adjacent along A_1. For suppose D_1, D_1' are not adjacent along A_1. If $D_1 \cap D_1' = \emptyset$ then we can construct a new 2-sphere or compression disk $D_1 \cup A_1 \cup D_1'$ which would form the frontier of a 3-cell into which a component of nF could be isotoped. This is impossible. If $D_1' \subset D_1$ then the surface $(F - D_1') \cup (D_1 \cup A)$ obtained by replacing the disk D_1' by the disk $A \cup D_1$ is isotopic to F but has less weight, again an impossibility.

If D_1' is also a disk patch then $\rho(D_1)$ and $\rho(D_1')$ are parallel disks which are switched when a regular exchange is made along α. This contradicts the assumption that $F_1 + F_2$ is in reduced form. If D_1' is not a disk patch, we observe that Lemma 5.3 can be applied to construct a sequence of least weight disk patches leading to a pair of adjacent disk patches and the same contradiction. To see that this lemma applies, we first consider a least weight disk patch D_i with $\partial D_i = \alpha_i'$ and let D_i' denote the disk in nF with frontier α_i''. As above, these two disks must be adjacent. If $\mathrm{wt}(D_i) < \mathrm{wt}(D_i')$ then a normal surface isotopic to F and of smaller weight could be constructed by replacing D_i' with a copy of D_i. Therefore $\mathrm{wt}(D_i) = \mathrm{wt}(D_i')$ and thus Lemma 5.3 applies. It follows that no patch can be a disk patch. □

Proof of Theorem 6.5. Suppose we have $nF = F_1 + F_2$ in reduced form. As we have already observed, we may assume that all intersection curves in $F_1 \cap F_2$ are two-sided. To show that F_1 and F_2 are incompressible one can use Lemma 6.6 and follow the proof of Theorem 2.1 in [JO]. Thus, we will assume the incompressibility of F_1 and F_2 and show ∂-incompressibility.

Suppose there exists an essential ∂-compression disk for F_1. Among all such essential ∂-compression disks we choose D to be transverse to F_2 and such that $F_2 \cap D$ has a minimal number of components. Let $\partial D = \mu \cup \nu$, where μ denotes the arc $D \cap F_1$ and $\nu = D \cap \partial M$. We must have $F_2 \cap D \neq \emptyset$ for otherwise we would be able to find a disk patch of $F_1 + F_2$. Observe that $F_2 \cap D$ has no simple closed curve components since F_2 is incompressible and such components could be removed by an isotopy of D. In a similar fashion, observe that no component of $F_2 \cap D$ can have both endpoints in $\nu \subset \partial M$ since such an arc innermost on D would be the frontier of a ∂-compression disk for F and removable by an isotopy of D. Thus each component of $F_2 \cap D$ is a spanning arc of D with at most one end point in ν.

We refer to the closure of a component of $D - (D \cap F_2)$ as a *region* in D. Let α be a component of $D \cap F_2$ and let x_0 denote an end point of α in μ. Then x_0 lies on a regular intersection curve in $F_1 \cap F_2$. There are two

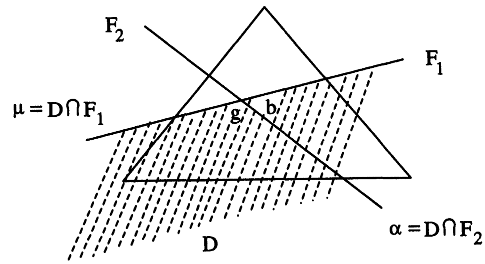


FIG. 6.1 Corner labels

regions in D abutting α with only one meeting adjacent patches along the intersection curve containing x_0. Following [JO], we assign the label b to the corner at x_0 of the region meeting adjacent patches and the label g to the corner at x_0 of the neighboring region meeting nonadjacent patches (see Figure 6.1). Observe that if Δ is a region with only g labels at corners then Δ yields a compression or a ∂-compression disk for nF.

CLAIM. *There exists a region Δ in D such that either Δ contains no b corners or Δ contains only one b corner and $\Delta \cap \partial M = \emptyset$.*

First suppose that for some component α with both endpoints in μ there is a g label at one of the corners abutting both α and the arc τ in μ which has endpoints $\alpha \cap \mu$. The disk D' in D bounded by $\alpha \cup \tau$ contains a region with at most one b corner. To see this, let n be the number of spanning arcs of $F_2 \cap D'$ in D'. Observe that these arcs cut D' into $n + 1$ regions with a total of either $2n$ or $2n + 1$ corners of type b. Thus at least one region in D' must have less than two b corners.

Now suppose that every component α_i of $D \cap F_2$ with both endpoints in μ has only b labels at the corners adjacent to both α_i and the arc in μ having endpoints $\alpha_i \cap \mu$. Let D' denote the closure of the component of $D - \cup \alpha_i$ containing ν. Let n denote the number of components of $D \cap F_2$ with one end point in μ and one in ν. The spanning arcs of this type cut the disk D' into $n + 1$ regions with exactly n corners labeled b. Any component α_i with both endpoints in μ contributes only g corners to regions contained in D'. Clearly there is a region with only g corners and this establishes the claim.

Consider a region with no b corners. Δ corresponds to a compression or a ∂-compression disk for nF, which we also denote by Δ. It follows that there exists a disk $\Delta' \subset nF$ such that $fr(\Delta') = \Delta \cap nF$ and $\Delta \cup \Delta'$ is the frontier of a 3-cell. The trace curves intersect Δ' in spanning arcs that split Δ' into regions. Let E denote an outermost one of these regions such that the frontier of E in Δ' consists of a single trace curve β. Let $\alpha = E \cap \Delta$, a component of $D \cap F_2$. We have three possibilities to consider.

(i) If $\rho(E) \subset F_2$ and $\partial\beta \subset \Delta \cap nF$ then $\partial\alpha = \partial\beta$. Let disks D', D'' denote the two disks into which α splits D, where $D' \cap \partial M = D \cap \partial M$. Since F_1 is already known to be incompressible, it follows that $D'' \cup E$ is an inessential compression disk for F_1 and $D_1 = D' \cup E$ is an essential ∂-compression disk for F_1.

(ii) If $\rho(E) \subset F_2$ and $\partial\beta \not\subset \Delta \cap nF$, let D', D'' denote the two disks into which D is split by α. Then either $E \cup D'$ or $E \cup D''$ is an essential compression disk D_1 for F_1 which can be isotoped so as to intersect F_2 in fewer components than in $F_2 \cap D$.

(iii) If $\rho(E) \subset F_1$ then $\partial\alpha = \partial\beta \subset \partial D$ and α can simply be pushed across E past β to obtain an isotopy of D reducing the number of components in $F_2 \cap D$. In all three cases we have a contradiction to our choice of D, which was chosen to meet F_2 in as few components as possible.

If there does not exist a region with only g corners then there exists a region Δ with exactly one b corner and which is disjoint from $\nu = D \cap \partial M$. The argument in [JO] shows that there exists an exchange annulus or disk A corresponding to the intersection curve through the b corner such that $A \cap F$ bounds a parallel annulus or disk $A' \subset F$, respectively. An isotopic surface of strictly smaller weight can be constructed from F by taking $(F - A') \cup A$. This contradicts the hypothesis that F is a least weight normal surface and completes the proof of the theorem. □

COROLLARY 6.7. *Let $M = F \times I$ where F is a closed surface that is not a 2-sphere or projective plane. Then there exists an essential two-sided annulus A which is a vertex surface and ∂A meets both $F \times \{0\}$ and $F \times \{1\}$.*

Proof. Among all two-sided essential annuli having a boundary component in both $F \times \{0\}$ and $F \times \{1\}$, let A be one with the least weight. If A is not already a vertex then we can write

$$nA = V_1 + \cdots + V_k,$$

where each V_i is a two-sided vertex surface. By Theorem 6.5, each V_i is incompressible and ∂-incompressible. If some V_i is a moebius band then consider the lift V_i^* of V_i to the orientable double-cover of $F \times I$. By [W_3], the annulus V_i^* must be ∂-parallel and it follows that V_i is ∂-parallel. This is impossible since V_i is ∂-incompressible and hence none of the V_i can be moebius bands.

We show that among the V_i there is a two-sided essential annulus spanning the two boundary components of M. By computing the Euler characteristics of nA and $V_1 + \cdots + V_k$ we obtain

$$\sum_{i=1}^{k} b_i = 2k - \sum_{i=1}^{k} x_i,$$

where b_i is the number of boundary components of V_i and x_i is the twice the genus of V_i if V_i is orientable and x_i is the number of crosscaps in V_i if V_i is non-orientable. Since none of the V_i can be spheres, disks, projective planes, or moebius bands, either $x_i \geq 2$ or $x_i = 0$ and $b_i \geq 2$.

Since $\sum_{i=1}^{k} b_i > 0$, we cannot have all the $x_i \geq 2$. Suppose that we have chosen notation such that $x_i \geq 2$ for $i \leq j$ and $x_i = 0$ for $i > j$. Then we have

$$\sum_{i=j+1}^{k} b_i \leq 2k - \sum_{i=1}^{k} x_i \leq 2(k-j).$$

If some $b_i > 2$ for $j < i \leq k$, then there must be some V_t with $b_t < 2$ for $j < t \leq k$, which cannot occur. Thus $b_i = 2$ and $x_i = 0$ for $i = j+1, \dots, k$ and hence each corresponding V_i is a two-sided incompressible, ∂-compressible annulus. Such an annulus in $F \times I$ cannot have both boundary components in the same boundary component of $F \times I$ without being boundary parallel. □

If we assume that M is orientable then we can prove a stronger result.

COROLLARY 6.8. *Let A be a normal, two-sided, essential annulus or torus in the orientable, compact, irreducible, ∂-irreducible 3-manifold M. If A is least weight in its isotopy class then each vertex surface in the face $\mathscr{C}(A)$ is either an essential annulus or an essential torus.*

Proof. Assume that A is an essential annulus or torus which is least weight relative to its isotopy class. If A is not already a vertex surface, let V_1 be a two-sided vertex surface in the face $\mathscr{C}(A)$ and write $nA = V_1 + V_2 + \cdots + V_k$, where the V_i are all two-sided vertex surfaces in $\mathscr{C}(A)$. By Theorem 6.5, each V_i is incompressible and ∂-incompressible. As in the previous lemma, by computing the Euler characteristics of nA and $V_1 + \dots + V_k$, we obtain $\sum_{i=1}^{k} b_i = 2k - 2\sum_{i=1}^{k} g_i$, where b_i is the number of boundary components of V_i and g_i is the genus of V_i (V_i is now orientable).

One can use an induction argument to prove the following claim: Assume g_i, b_i are n pairs of nonnegative integers such that (a) $b_i \geq 2$ whenever $g_i = 0$ and (b) $\sum_{i=1}^{n} b_i \leq 2n - 2\sum_{i=1}^{n} g_i$. Then for each $i = 1, \dots, n$, either $b_i = 0$, $g_i = 1$ or $b_i = 2$, $g_i = 0$. For the inductive step, observe that if

$$\sum_{i=1}^{n+1} b_i \leq 2(n+1) - 2\sum_{i=1}^{n+1} g_i$$

then either some $g_j = 0$ or for all j we have $g_j = 1$ and $b_j = 0$. If $g_j = 0$ for some j then $b_j \geq 2$ and we can omit the j-th and apply the induction hypothesis.

It follows that the chosen vertex surface V_i is either an essential annulus or an essential torus. □

7. Splitting a 3-manifold into irreducible submanifolds

Let M be a closed irreducible 3-manifold with a fixed triangulation $\mathcal{T}$. We describe an algorithm to decompose M into irreducible 3-manifolds. Since this is achieved without a solution to the 3-sphere recognition problem, many of the irreducible 3-manifolds obtained will be 3-spheres.

ALGORITHM 7.1. *For the decomposition of M into irreducible 3-manifolds.*

Procedure. Let $\Sigma_1 = \{S_1, \dots, S_n\}$ denote the set of all normal vertex surfaces that are 2-spheres. Set $F_{1,1} = S_1$. Construct from $\{F_{1,1}, S_2\}$ a finite collection of pairwise disjoint 2-spheres $\{F_{2,1}, \dots, F_{2,k(2)}\}$ by cutting S_2 along the boundaries of innermost disks in $F_{1,1}$ and capping the resulting disk pieces with copies of the innermost disks. This process is continued until we have the desired disjoint collection $\{F_{2,1}, \dots, F_{2,k(2)}\}$. We let

$$\Sigma_2 = \{F_{2,1}, \dots, F_{2,k(2)}\} \cup \{S_3, \dots, S_n\}.$$

Suppose we have constructed the collection

$$\Sigma_i = \{F_{i,1}, \dots, F_{i,k(i)}\} \cup \{S_{i+1}, \dots, S_n\}$$

where $\{F_{i,1}, \dots, F_{i,k(i)}\}$ is a pairwise disjoint collection of 2-spheres constructed by this inductive procedure from $\{S_1, \dots, S_i\}$. We proceed to modify the next vertex 2-sphere S_{i+1} by cutting and pasting along disks in $\{F_{i,1}, \dots, F_{i,k(i)}\}$ spanning S_{i+1}. The cutting and pasting is always done along a spanning disk that is innermost among those along which the operation has yet to be performed. We let $\{F_{i+1,1}, \dots, F_{i+1,k(i+1)}\}$ denote the collection of pairwise disjoint 2-spheres obtained by taking the union of $\{F_{i,1}, \dots, F_{i,k(i)}\}$ together with the 2-spheres obtained by our cut and paste modifications to S_{i+1}. We let

$$\Sigma_{i+1} = \{F_{i+1,1}, \dots, F_{i+1,k(i+1)}\} \cup \{S_{i+2}, \dots, S_n\}.$$

This process eventually leads us to a pairwise disjoint collection Σ_n of 2-spheres constructed from $\{S_1, \dots, S_n\}$. □

THEOREM 7.2. *Σ_n decomposes M into irreducible 3-manifolds.*

Proof. It follows from Theorem 5.2 that there exists a minimal complete system $\{X_1, \dots, X_r\}$ of 2-spheres for M in the collection Σ_1. We separate the collection of modified 2-spheres $\{F_{i,1}, \dots, F_{i,k(i)}\}$ into two sets: let $\mathscr{A}_i$ denote those arising from cutting and pasting on members of $\{X_1, \dots, X_r\}$ and let $\mathscr{B}_i$ denote the others. Thus we have $\Sigma_i = \mathscr{A}_i \cup \mathscr{B}_i \cup \{S_{i+1}, \dots, S_n\}$, where we assume we have chosen notation so that $\mathscr{A}_i$ contains the 2-spheres resulting from modifications made to the set $\{X_1, \dots, X_{j(i)}\}$ and $\{X_{j(i)+1}, \dots, X_n\} \subset \{S_{i+1}, \dots, S_n\}$.

Define $\Lambda(i)$ to be the set of pairwise disjoint 2-spheres we get by taking the union of $\mathscr{A}_i$ together with the 2-spheres obtained from $\{X_{j(i)+1}, \dots, X_n\}$ by the process of cutting and capping with innermost disks from $\mathscr{A}_i$. Thus $\Lambda(1) = \{X_1, \dots, X_n\}$. We claim that for each i, $1 \leq i \leq n$, the system of 2-spheres Λ_i splits M into punctured irreducible 3-manifolds. We assume that this is the case for $i = m$ and show that the collection $\Lambda(m+1)$ also has this property.

There is nothing to show if $\Lambda(m) = \Lambda(m+1)$ and this is the case unless $S_{m+1} = X_{j(m+1)}$. Thus, let us assume $S_{m+1} = X_{j(m+1)}$. The system $\Lambda(m+1)$ consists of the 2-spheres in $\mathscr{A}_m$, the set $\mathscr{C}_{m+1}$ resulting from the cutting and capping of $X_{j(m+1)}$ along $\mathscr{A}_m \cup \mathscr{B}_m$, and finally those 2-spheres resulting from the cutting and capping of $\{X_{j(m+1)+1}, \dots, X_n\}$ along $\mathscr{A}_m \cup \mathscr{C}_m$. It follows that $\Lambda(m)$ can be transformed into $\Lambda(m+1)$ by a sequence of elementary cut-and-paste steps, each preserving the desired decomposing properties. Each step in the sequence is one of cutting a 2-sphere S along the boundary of a spanning disk D and capping the resulting disk components of the split S with disjoint copies of D so as to produce a pair $\{S', S''\}$ of disjoint 2-spheres from S. It is clear that if S is a member of a collection of pairwise disjoint 2-spheres Λ that decomposes M into punctured irreducible 3-manifolds then the collection $(\Lambda - \{S\}) \cup \{S', S''\}$ (assuming it is a pairwise disjoint collection) also splits M into punctured irreducible 3-manifolds. This completes the proof that Σ_n decomposes M into punctured irreducible 3-manifolds. □

8. Splitting an irreducible 3-manifold into simple and characteristic submanifolds

Let M be an *orientable*, compact, irreducible, ∂-irreducible, sufficiently large 3-manifold with a triangulation $\mathscr{T}$. It is shown in [JS, Jo] that there exists a canonical system of pairwise disjoint, properly embedded, essential annuli and tori in M which split M into a simple 3-manifold and a characteristic submanifold $V(M)$. The characteristic submanifold is a Seifert fibered space and is unique up to isotopy. In this section we give an algorithm that uses vertex surfaces which are essential annuli and tori to produce this splitting. It is first necessary that we be able to recognize a Seifert fibered space.

ALGORITHM 8.1. *For determining if M is a Seifert fibered space.*

Procedure. If M is a closed Seifert fiber space, it follows from Corollary 6.8 that there exists an essential torus among the vertex surfaces. We can use Algorithm 9.6 to test each vertex torus to determine if any are essential. If an essential torus T is found then split M along T to obtain M_1 and proceed to triangulate M_1. In the case M already has boundary, we may assume that each boundary component is a torus and let $M_1 = M$.

Using Algorithms 9.6 and 9.7, we look for an essential annulus among the vertex surfaces of M_1. We know from Corollary 6.8 that if one exists then one can be found among the vertex surfaces. Assume a vertex surface A_1 that is an essential annulus has been found. Let M_2 be obtained from M_1 by splitting along A_1. Let $\partial^- A_1$ and $\partial^+ A_1$ denote the traces of ∂A_1 in M_2. Test each component of M_2 to see if it is (i) a solid torus, (ii) $S^1 \times S^1 \times I$, or (iii) $M(K)$, a twisted I-bundle over the Klein bottle. We fiber each solid torus component V_1 so that each component of

$$V_1 \cap (\partial^- A_1 \cup \partial^+ A_1)$$

is a fiber. If a component V_1 is either $S^1 \times S^1 \times I$ or $M(K)$ then there exist two possible Seifert fiberings of $M(K)$ and an infinite number of Seifert fiberings for $S^1 \times S^1 \times I$, up to isotopy. We attempt to fiber V_1 such that each component of $V_1 \cap (\partial^- A_1 \cup \partial^+ A_1)$ is a fiber. If it is not possible to fiber all such special components of M_2 in this way then we are done and M is not a Seifert fiber space.

As long as it is possible, we continue a refinement of the above process in which we find at each step an essential annulus A_i among the vertex surfaces of M_i. We only look for essential annuli in components of M_i that we have not previously endowed with a fibering. If an essential annulus A_i is found then we isotope the boundary of A_i, if possible, so it is disjoint from the traces $\{\partial^- A_j \cup \partial^+ A_j\}_j$ of the boundaries of the previous annuli. If this cannot be done then M is not a Seifert fiber space (the component V_i of M_i containing A_i is neither $S^1 \times S^1 \times I$ nor $M(K)$). Thus, we may assume that ∂A_i is disjoint from $\{\partial^- A_j \cup \partial^+ A_j\}_j$ and split M_i along A_i to obtain M_{i+1}. Eventually, this process can no longer be carried out. In particular, if M has t tetrahedra in its triangulation then its closed Haken number $\overline{h}(M)$ is less than or equal to $61t$ [H4]. According to Theorem IV.7 of [Ja], no partial hierarchy such as we are constructing here can have length greater than $\overline{h}(M)$.

The 3-manifold M_1 is fibered if and only if we end up with a disjoint union of Seifert fibered solid tori, $S^1 \times S^1 \times I$'s, and $M(K)$'s. If M_1 is fibered and M is closed then it only remains to decide whether or not the fibering we are working with, or possibly another fibering of M_1, can be matched up when forming M. Thus assume M is closed. If no component of M_1 is a product or a twisted I-bundle over the Klein bottle then M_1 has a unique Seifert fibering and M is a Seifert fiber space if and only if the fibers in M_1 at hand match up along T. If M_1 is a product $S^1 \times S^1 \times I$ and no fibering in M_1 can be matched up in M along T, then M is a Seifert fiber space if and only if the gluing homeomorphism is homotopic to one of the seven periodic ones in the list on page 122 of [He]. If a component of M_1 is a twisted I-bundles over a Klein bottle, then there is a second fibering on this component that we can employ to try for a match along T. If this fails then M is not a Seifert fiber space. □

We now describe a procedure to produce the characteristic Seifert fiber space of M. Let

$$\mathscr{F} = \{F_1, \dots, F_m\}$$

be a canonical system of essential annuli and tori in M that splits M into a characteristic Seifert fiber space $V(M)$ and a simple 3-manifold. We may assume that $\mathscr{F}$ is a normal surface and that $wt(\mathscr{F})$ is minimal relative to the isotopy class of $\mathscr{F}$. It is a consequence of Corollary 6.8 that all vertex surfaces carried by the face $\mathscr{C}(\mathscr{F})$ are essential annuli and tori. Thus, if $\mathscr{V} = \{T_1, \dots, T_n\}$ denotes the collection of all essential annuli and tori vertex surfaces, we clearly have $\mathscr{F}$ contained in a regular neighborhood $N(\mathscr{V})$ of $\cup T_i$.

While the details of the construction of $V(M)$ from $\mathscr{V}$ are somewhat detailed, the idea is rather simple. One takes up one of the essential annuli or tori T_i, after having already used $T_1, \dots, T_{i-1}$ to construct a Seifert fibered submanifold Σ_{i-1} in M. We isotope $N(T_i)$ so that there are no regions of parallelity between $fr(N(T_i))$ and $fr(\Sigma_{i-1})$. However, in this process we only pull them apart along disks and leave them to intersect along essential annuli in the intersection of their frontiers. Then we look at all the pieces consisting of $N(T_i)$, the components of Σ_{i-1}, and the Seifert fiber space components of $Cl(M - (N(T_i) \cup \Sigma_{i-1}))$. We unite those with compatible Seifert fiberings and pull the annuli or tori frontiers apart where the fiberings are not compatible. This pulling apart leaves products which will eventually be simple product components in the complement of the final M. The process is continued until we have used all the surfaces in $\mathscr{V}$.

ALGORITHM 8.2. *The decomposition of M into its characteristic submanifold V(M) and simple 3-manifolds.*

Procedure. If M is a Seifert fiber space we set $\Sigma_1 = M$ and are finished. Thus we may assume that M is not a Seifert fiber space and form the list $\{T_1, \dots, T_n\}$ of all essential normal tori and annuli in M that are vertex surfaces. We may assume the T_i intersect pairwise in a transverse fashion.

We construct a sequence $\Sigma_1, \ldots, \Sigma_n$ of Seifert fiber spaces such that $fr(\Sigma_i)$ is incompressible and ∂-incompressible in M and $\mathscr{F}$ is contained, up to isotopy, in $\Sigma_i \cup N(T_{i+1} \cup \cdots \cup T_n)$. The sequence terminates with the desired characteristic Seifert fiber space $V(M) = \Sigma_n$.

We begin by setting $\Sigma_0 = \emptyset$. Assume that the following construction has been carried out using the vertex surfaces $T_1, \ldots, T_{i-1}$ and we have obtained the Seifert fiber space Σ_{i-1} such that $W_{i-1} = fr(\Sigma_{i-1})$ is incompressible and ∂-incompressible in M and that $\mathscr{F}$ is contained, up to isotopy, in $\Sigma_{i-1} \cup N(T_i \cup \cdots \cup T_n)$.

Step 1. We consider the next vertex annulus or torus T_i from our list and perform the following simplification. Suppose $T_i \cap fr(\Sigma_{i-1})$ contains an inessential component, either an arc or simple closed curve. These are eliminated by the following construction. We can choose an innermost disk component D of $W_{i-1} - (T_i \cap W_{i-1})$ such that $fr(D) \subset (T_i \cap W_{i-1})$ and $fr(D)$ contains an inessential component of $T_i \cap fr(\Sigma_{i-1})$. Let D' denote the disk in T_i with frontier $\overline{D} \cap T_i$. Form $T_i' = (T_i - D') \cup D$ and isotope it off W_{i-1} slightly along D. T_i' is isotopic to T_i and we can continue this process until we obtain a new annulus or torus, which we again denote by T_i, such that $T_i \cap W_{i-1}$ contains no inessential intersection arcs or simple closed curves.

Step 2. We next eliminate regions of parallelity between annuli or tori in $fr(\Sigma_{i-1})$ and corresponding surfaces in $fr(N(T_i))$. We say that a product $F \times [-1, 1] \subset M$ is a region of parallelity between surfaces G^- and G^+ in the following circumstance. Let γ be a union of components of ∂F, possibly empty, and assume that $\gamma \times [-1, 1] \subset \partial M$. Let

$$G^- = F \times \{-1\} \cup (\partial F - \gamma) \times [-1, 0])$$

and

$$G^+ = F \times \{1\} \cup (\partial F - \gamma) \times [0, 1]).$$

Set $G(0) = N(T_i)$, $\Sigma_{i-1}(0) = \Sigma_{i-1}$ and suppose we have already constructed the 3-manifolds $G(j-1), \Sigma_{i-1}(j-1)$ in M. Suppose V_j is an innermost region of parallelity between an annulus or torus in $fr(\Sigma_{i-1}(j-1))$ and one in $fr(G(j-1))$. We remove V_j as follows, depending on how it is situated.

(i) If $V_j \cap G(j) \subset fr(V_j)$ then let

$$G(j) = G(j-1) \cup V_j \quad \text{and} \quad \Sigma_{i-1}(j) = \Sigma_{i-1}(j-1).$$

(ii) If $V_j \subset G(j-1)$ and $V_j \cap \Sigma_{i-1}(j-1) = fr(V_j)$ then let

$$G(j) = G(j-1) \quad \text{and} \quad \Sigma_{i-1}(j) = \Sigma_{i-1}(j-1) \cup V_j.$$

(iii) If $V_j \subset G(j-1) \cap \Sigma_{j-1}$ then let

$$G(j) = Cl\big(G(j-1) - V_j\big) \quad \text{and} \quad \Sigma_{i-1}(j) = \Sigma_{i-1}(j-1).$$

Eventually we obtain $X_0 = G(r)$, which is isotopic to $N(T_i)$, and $\Sigma'_{i-1} = \Sigma_{i-1}(r)$, which is isotopic to Σ_{i-1}, and there do not exist any regions of parallelity between annuli or tori in their frontiers.

Step 3. If no component in Σ'_{i-1} intersects X_0, let $\Sigma^*_i = X_0 \cup \Sigma'_{j-1}$ and proceed to Step 4. Otherwise, form a list $\{Y_1, \dots, Y_q\}$ consisting of the components of Σ'_{i-1} which meet X_0. Suppose that $\mathring{X}_0 \cap \mathring{Y}_j \neq \emptyset$. Let K be the closure of a component of $fr(X_0) \cap \mathring{Y}_j$. Since $fr(X_0) \cap fr(Y_j)$ contains only essential curves, it follows that K must be an injective annulus. Because of our construction, K cannot be a torus. If the fibering of Y_j cannot be deformed so K is a union of fibers then Y_j is $K \times S^1$ and a new fibering can be chosen for Y_j so that K is fibered. This fibering of K can be extended to a Seifert fibering of X_0. Since any other such component of $fr(X_0) \cap \mathring{Y}_j$ would be disjoint from K, it would give rise to compatible fiberings of X_0 and of Y_k. Thus, we may assume that the Seifert fiberings of X_0 and Y_j agree on $\mathring{X}_0 \cap \mathring{Y}_j$.

Let X_1 denote the union of X_0 together with all components of Σ'_{i-1} which are disjoint from X_0. We take up the remaining components Y_j of Σ'_{i-1} one at a time, see how they fit together with the Seifert fiber space X_1, and either pull them apart or combine them into a fibered Seifert fiber space.

Suppose we have already considered $\{Y_1, \dots, Y_{k-1}\}$ and constructed the pairwise disjoint Seifert fiber spaces $X_k, Y'_1, \dots, Y'_{k-1}$ (some of which may be empty sets) such that $\{Y'_1, \dots, Y'_{k-1}, Y_k, \dots, Y_q\}$ is also a pairwise disjoint collection. Consider the next Seifert fiber space Y_k. We will construct the Seifert fiber spaces Q, P, R and S which will be used to form X^*_{k+1} and Y'_k. We let $\{b_1, \dots, b_p\}$ denote the subcollection of 2-dimensional components of $fr(X_k) \cap fr(Y_k)$ for which X_k and Y_k lie on opposite sides. Each b_j is contained in a component B_j of $fr(Y_k)$ and in a component C_j of $fr(X_k)$. Consider collar neighborhoods $B_j \times [0,1]$ of $B_j = B_j \times \{0\}$ in Y_k and $C_j \times [0,1]$ of $C_j = C_j \times \{0\}$ in X_k.

We first consider the case when both X_k and Y_k have unique Seifert fiber structures. We list the possible ways in which these two fiber structures can come together along the annuli and tori b_j.

(i) b_j is a union of fibers in both fiber structures and fibers from each are isotopic in b_j. In this case b_j may be either an annulus or a torus.

(ii) b_j is a union of fibers in both fiber structures but the two fiberings of b_j are not isotopic. Here b_j is a torus.

(iii) The annulus b_j is a union of fibers from X_k but not from Y_k.

(iv) The annulus b_j is a union of fibers from Y_k but not from X_k.

(v) The annulus b_j is not a union of fibers from either Y_k or X_k.

For $\lambda = i, \dots, v$, let $\Gamma(\lambda)$ denote the set of indices $\{j | b_j \text{ is of type } (\lambda)\}$. Then let

$$P = \left(\bigcup_{j \in \Gamma(ii)} B_j \times [0,1) \right) \cup \left(\bigcup_{j \in \Gamma(iii)} B_j \times [0,1) \right) \cup \left(\bigcup_{j \in \Gamma(v)} B_j \times [0,1) \right)$$

$$Q = \left(\bigcup_{j \in \Gamma(iv)} C_j \times [0,1) \right) \cup \left(\bigcup_{j \in \Gamma(v)} C_j \times [0,1 \right) \Big)$$

$$R = \left(\bigcup_{j \in \Gamma(iii)} B_j \times [0,1/2] \right) \cup \left[\bigcup_{j \in \Gamma(v)} (B_j \times [0,1/2] \cup C_j \times [0,1/2]) \right]$$

$$S = \left(\bigcup_{j \in \Gamma(iv)} B_j \times [0,1/2] \right)$$

Now suppose one or both X_k and Y_k have more than one Seifert fiber structure. Any such Seifert fiber space must be a twisted I-bundle over the Klein bottle or $S_1 \times S_1 \times I$ (we cannot have a solid torus because of our construction). We proceed as in the case when the fiberings are both unique but now we remain flexible as long as possible as to which fibering we use when forming the groupings $\Gamma(\lambda)$.

Whether we have unique Seifert fiber structures or not, we use the same notation for the following Seifert fiber spaces. If $\Gamma(i) \neq \emptyset$ or $\mathring{X}_0 \cap \mathring{Y}_k \neq \emptyset$ then we set

$$X_{k+1}^* = Cl(X_k \cup Y_k - (P \cup Q)) \cup R \cup S \quad \text{and} \quad Y_k' = \emptyset.$$

Otherwise we let

$$X_{k+1}^* = Cl(X_k - Q) \cup R \quad \text{and} \quad Y_k' = Cl(Y_k - P) \cup S.$$

If some essential annulus or torus F_t from $\mathscr{F}$ happens to intersect a component b_j of $fr(X_k) \cap fr(Y_k)$ then either F_t can be isotoped off b_j or else $j \in \Gamma(i)$. In either case, the property of keeping $\mathscr{F}$ inside

$$\Sigma_{i-1}' \cup X_{k+1}^* \cup Y_1' \cup \cdots \cup Y_k' \cup Y_{k+1} \cup \cdots \cup Y_q \cup N(T_{i+1} \cup \cdots \cup T_n),$$

up to isotopy, is maintained.

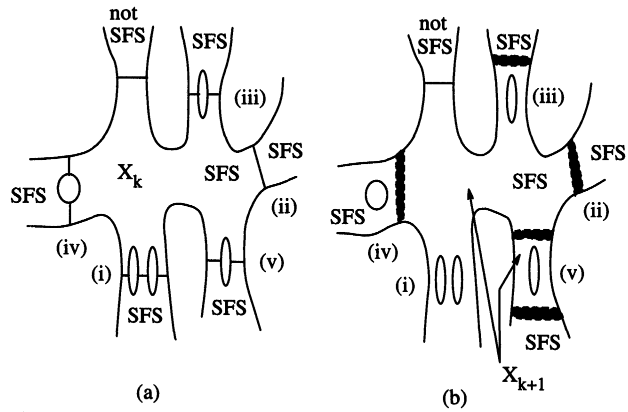


FIG. 8.1 Construction of $V(M)$

After considering each Seifert fiber space Y_j, $j = 1, \dots, q$, we finally construct the Seifert fiber space $\Sigma_i^* = X_{q+1}^* \cup Y_1' \cup \cdots \cup Y_q'$.

Step 4. Suppose Z is a solid torus component of $Cl(M - \Sigma_i^*)$. Since M is irreducible, no fiber in Σ_i^* can bound a meridian disk of the solid torus. Thus the fibering on $fr(Z)$ induced by that of Σ_i^* can be extended to Z. Let Σ_i^{**} denote the Seifert fibered space obtained by taking the union of Σ_i^* with all the solid torus components Z of $Cl(M - \Sigma_i^*)$. Observe that $fr(\Sigma_i^{**})$ now consists of only essential annuli and tori.

Let $\{Z_1, \dots, Z_r\}$ denote the components of $Cl(M - \Sigma_i^{**})$ which are Seifert fiber spaces. Repeat Step 3 using the Seifert fibered spaces $\{Z_1, \dots, Z_r\}$ now in place of the $\{Y_1, \dots, Y_q\}$ and letting $X_0 = \Sigma_i^{**}$. We obtain the Seifert fibered spaces $X_{r+1}, Z_1', \dots, Z_r'$ and let $\Sigma_i = X_{r+1} \cup Z_1' \cup \cdots \cup Z_r'$.

Step 5. We do this for each vertex surface $T_1, \dots, T_n$ in the list and obtain the desired characteristic Seifert fiber space $V(M) = \Sigma_n$. □

This procedure clearly gives us a Seifert fibered submanifold Σ_n in M such that no component of $Cl(M - \Sigma_n)$ is a Seifert fiber space other than $S^1 \times S^1 \times I$. The only question that remains is whether or not there exist essential tori or annuli in a component of $Cl(M - V(M))$. However, we were careful to ensure the existence of a canonical system of annuli and tori $\mathscr{F} = \{F_1, \dots, F_n\}$ for M which is contained in Σ_n. Since splitting M along $\mathscr{F}$ produces only Seifert fibers spaces and simple 3-manifolds, it follows that each component of $Cl(M - V(M))$ is simple.

9. Appendix: Miscellaneous algorithms

We collect together a number of useful algorithms, some of which are needed in §7 and §8. We assume that M is a compact 3-manifold with a given triangulation $\mathscr{T}$.

ALGORITHM 9.1. *For computing the Euler characteristic of a normal surface F in M.*

Procedure. For each edge e_i in $\mathscr{T}$ let t_i denote the number of tetrahedron in $\mathscr{T}$ containing e_i. Set $\varepsilon_{ij} = 1$ if the edge e_i meets a disk of type i and otherwise set $\varepsilon_{ij} = 0$. Suppose that F has normal coordinates $\mathscr{N}_F = (x_1, \dots, x_{7t})$. Let f_3 denote the total number of 3-sided elementary disks in F. Then $\chi(F) = (\frac{1}{2})f_3 - \sigma(F) + wt(F)$ where $\sigma(F) = \Sigma x_i$ and $\mathrm{wt}(F) = \Sigma_{i,j}\varepsilon_{ij}x_j/t_i$. □

ALGORITHM 9.2. *For deciding if a knot K in S^3 is unknotted.*

Procedure. Assume the knot K in $M = S^3$ is given so as to be contained in the 1-skeleton of the triangulation $\mathscr{T}$ of S^3. Let $N(K)$ be a regular neighborhood of K and construct a triangulation $\mathscr{T}'$ of $S^3 - \mathring{N}(K)$. Find the vertex surfaces of $\mathscr{P}_{\mathscr{T}'}$ which are disks. For each disk vertex surface D, determine if ∂D bounds a disk in ∂M by calculating Euler characteristics. The knot K is nontrivial if and only if all the disks D tested are inessential. □

ALGORITHM 9.3. *For deciding if a compact, irreducible 3-manifold M is a handlebody.*

Procedure. Form a list of the compression disks among the vertex surfaces and discard those whose boundary curve bounds a disk in ∂M. Assume we have constructed, from this list, a system $\mathscr{D}_j = \{D_1, \dots, D_j\}$ of pairwise disjoint, independent, essential compression disks. Choose the next unused disk D from the list. We can use cut and paste techniques to find a new disk D' isotopic to D and disjoint from $\mathscr{D}_j$. By considering the boundary curves in ∂M, we can determine whether or not D' is independent of $\mathscr{D}_j$. If it is independent then let $D_{j+1} = D'$ and if not, we discard D'. After we have exhausted the list of compression disk vertex surfaces and constructed the pairwise disjoint system of compression disks $\mathscr{D}_n$, we check to see if the boundary curves in $\mathscr{D}_n$ split ∂M into disks and annuli. If they do then M is a handlebody and otherwise it is not. □

ALGORITHM 9.4. *For deciding if a normal surface F is connected.*

Procedure. Consider the equivalence relation between elementary disks in F generated by the relation obtained by saying that two elementary disks of F meeting the same 2-simplex σ of $\mathcal{T}$ are equivalent if their edges in σ are identified. (Whether or not they are identified is well-determined by ordering the sets of elementary disks in each tetrahedron that have an edge of the given arc type.) Divide the elementary disks of F into equivalence classes. The components of F correspond to the equivalence classes. □

ALGORITHM 9.5. *For deciding if a two normal surfaces F and G intersect.*

Procedure. If F and G are not summable then they must intersect. If they are summable, form the sum $F + G$. Use Algorithm 9.4 to find the components of $F + G$. If the components are normal isotopic to F and G then F and G are disjoint (up to normal isotopy). If the components are anything else, then F and G do intersect and cannot be separated by a normal isotopy. □

ALGORITHM 9.6. *For determining if a surface F in a compact, irreducible 3-manifold M is injective.*

Procedure. Split M along F to obtain the 3-manifold M' and construct a triangulation $\mathcal{T}'$ of M' by subdividing the cell decomposition induced by $\mathcal{T}$. Form the system of normal equations for $\mathcal{T}'$. List the finite set of normal compression disk vertex surfaces. Test each of these compression disks D to see if they are essential by calculating the Euler characteristics of the components of $\partial M' - \partial D$. If none of the compression disks tested are essential then F is an injective surface. □

ALGORITHM 9.7. *To test a compact irreducible, P^2-irreducible 3-manifold M for a product structure $F \times [-1, 1]$.*

Procedure. We may as well assume that ∂M is either connected or has two components. We consider first the case where M has a connected boundary that is divided into two homeomorphic pieces $\partial^- M$ and $\partial^+ M$ intersecting in their common boundary. We follow the steps of Algorithm 9.3 with the additional stipulation that we only consider disk vertex surfaces that are essential compression disks in the context of a product $F \times [-1, 1]$. If we find enough essential compression disks to split M into 3-cells then we had the desired product in the beginning. Otherwise, M was not the product expected.

Now assume that ∂M has two components. We look for an essential annulus A among the vertex surfaces meeting both boundary components.

This will require Algorithm 9.6 to decide if A is injective and the previous case to decide if A is boundary parallel. If we find such an A then we proceed to split M and continue with the test as in the case where ∂M is connected and the two boundary pieces meet along the center curves of the two copies of A. □

ALGORITHM 9.8. *For determining if a closed, irreducible 3-manifold M is sufficientlylarge.*

Procedure. Form the system of normal equations for the triangulation $\mathcal{T}$ of M. List the finite constructable set of normal vertex solutions, discarding 2-spheres. Test each of these surfaces for injectivity using Algorithm 9.6. If none of the vertex surfaces are injective then M is not sufficientlylarge. □

REFERENCES

[H_1] W. HAKEN, *Theorie der Normal Flächen*, Acta Math., vol. 105 (1961), pp. 245–375.

[H_2] ______, *Ein Verfahren Aufspaltung einer 3-Mannigfaltigkeiten in irreduzible* 3-Mannigfaltigkeiten, Math. Zeitschr., vol. 76 (1961), pp. 427–467.

[H_3] ______, *Über das Homöomorphieproblem der 3-Mannigfaltigkeiten I*, Math. Zeitschr., vol. 80 (1962), pp. 89–120.

[H_4] ______, "Some results on surfaces in 3-manifolds" in *Studies in modern topology*, MAA Studies in Mathematics, no. 5, Prentice-Hall, Englewood Cliffs, 1968, pp. 39–98.

[He] J. HEMPEL, *3-Manifolds*, Ann. of Math. Studies, no. 86. Princeton University Press, New Jersey 1976.

[Ja] W. JACO, *Lectures on three-manifold topology*, Regional Conference Series in Mathematics, no. 43., Amer. Math. Soc., Providence, R.I., 1980.

[JO] W. JACO and U. OERTEL, *An algorithm to decide if a 3-manifold is a Haken manifold*, Topology, vol. 23 (1984), pp. 195–201.

[JR] W. JACO and J.H. RUBINSTEIN, *PL equivariant surgery and invariant decompositions of 3-manifolds*, Advances in Mathematics, vol. 73 (1989), pp. 149–191.

[JS] W. JACO and P. SHALEN, *Seifert fibered spaces in 3-manifolds*, Mem. Amer. Math. Soc. 21, No. 220, 1979.

[Jo] K. JOHANNSON, *Homotopy equivalences of 3-manifolds with boundaries*, Lecture Notes in Math., no. 761, Springer Verlag, New York, 1979.

[K] H. KNESER, *Geschlossene Flächen in dreidimensionalen Mannigfaltigkeiten*, Jber. Deutsch. Math.-Verein., vol. 38 (1929), pp. 248–260.

[S] H. SCHUBERT, *Bestimmung der Primfaktorzerlegung von Verkettungen*, Math. Zeitschr., vol. 76 (1961), pp. 116–148.

[W_1] F. WALDHAUSEN, *Recent results on sufficientlylarge 3-manifolds*, Proc. Sympos. Pure Math., vol. 32 (1978), pp. 21–38.

[W_2] ______, *The word problem in fundamental groups of sufficientlylarge irreducible 3-manifolds*, Ann. of Math. (2), vol. 88 (1968), pp. 272–280.

[W_3] ______, *On irreducible 3-manifolds which are sufficientlylarge*, Ann. of Math. (2), vol. 87 (1968), pp. 56–88.

AMERICAN MATHEMATICAL SOCIETY
PROVIDENCE, RHODE ISLAND

UNIVERSITY OF CONNECTICUT
STORRS, CONNECTICUT