

NORMAL SURFACE Q-THEORY

JEFFREY L. TOLLEFSON

We describe an approach to normal surface theory for triangulated 3-manifolds which uses only the quadrilateral disk types (Q-disks) to represent a nontrivial normal surface. Just as with regular normal surface theory, interesting surfaces are among those associated with the vertices of the projective solution space of this new Q-theory.

1. Introduction.

A normal surface F in a triangulated compact 3-manifold M intersects each tetrahedron in elementary disks which are either triangles or quadrilaterals. In this paper we show that only the quadrilaterals are significant in the sense that a normal surface without trivial components is completely determined by its quadrilateral disks alone. From this observation we develop a normal surface Q-theory analogous to Haken's normal surface theory [1], [2]. The normal Q-coordinates \vec{F}_Q of a normal surface F is the vector of nonnegative integers given by the multiplicity of each quadrilateral disk type in F . The vector \vec{F}_Q is characterized by the fact that it is an admissible solution to a system of linear Q-matching equations in which there is just one equation for each interior 1-simplex of the triangulation. A significant feature of this approach is that fewer variables and equations are involved than in regular normal surface theory thus simplifying computations.

The projectivized solution space \mathcal{P}_Q of the Q-matching equations is a compact, convex, linear cell. Interesting normal surfaces occur among those associated with the vertices of \mathcal{P}_Q and a number of algorithms can be rephrased from this new point of view. For example, if there exists a two-sided, incompressible, ∂ -incompressible surface in a triangulated, compact, irreducible, ∂ -irreducible 3-manifold then there exists one corresponding to a vertex of \mathcal{P}_Q .

2. Quadrilateral Disks in Normal Surfaces.

Throughout, M will denote a compact 3-manifold with a fixed triangulation \mathcal{T} . We will discuss only what we need from normal surface theory in order

to explain normal surface Q-theory. See [5] or [6] for additional background on normal surfaces.

A normal surface F intersects the 2-skeleton transversely and intersects each tetrahedron Δ in a pairwise disjoint collection of triangular and quadrilateral disks, called **elementary disks**, which are properly embedded in Δ . Each edge of an elementary disk is a straight line in a 2-simplex of Δ and spans distinct 1-simplexes.

A *normal isotopy* is an isotopy which is invariant on each simplex of \mathcal{T} . The normal isotopy class of an elementary disk in a tetrahedron Δ is called a *disk type* and the normal isotopy class of a spanning arc in a 2-simplex of Δ is called an *arc type*. In each tetrahedron Δ there are seven disk types, four of which consist of triangles (*T-disks*) and three consisting of quadrilaterals (*Q-disks*). We label T-disk types and arc-types by the vertex which they cut off. For example, if a is a vertex of the tetrahedron Δ then a T-disk cutting off the vertex a will be referred to as an a -disk and an edge of such a T-disk will be called an a -arc.

Consider a pair of elementary disks E_1, E_2 in a tetrahedron Δ which intersect *transversely*, that is, each component of $E_1 \cap E_2$ is an arc properly embedded in Δ that spans the interiors of distinct 2-simplex faces of Δ . A component α of $E_1 \cap E_2$ 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. If each intersection curve of $F \cap G$ is *regular*, in the sense that it is a union of regular arcs, then there is a unique normal surface $F + G$ constructed using standard cut-and-past operations along the regular curves of intersection. $F + G$ is a uniquely determined normal surface made up of exactly the same collection of elementary disks, up to normal isotopy of the disks, as in $F \cup G$.

The cell decomposition which a normal surface F inherits from the triangulation \mathcal{T} consists of triangular and quadrilateral disks and is used in normal surface theory to algebraically represent the surface F as a solution of a linear system of matching equations in which variables correspond to elementary disk types of \mathcal{T} . Fix an ordering $d_1^Q, \dots, d_{3t}^Q, d_1^T, \dots, d_{4t}^T$ of the disk types in \mathcal{T} , where the Q-disk types d_i^Q are listed first, followed by the T-disk types d_j^T . A $7t$ -tuple $\vec{F} = (x_1, \dots, x_{3t}, y_1, \dots, y_{4t})$, called the *normal coordinates* of F , is assigned to a normal surface F by letting x_i denote the number of elementary Q-disks in F of type d_i^Q and y_j denote the number of elementary T-disks in F of type d_j^T . A normal surface F is uniquely deter-

mined, up to normal isotopy, by the vector \vec{F} . The *normal Q-coordinates* of F are $\vec{F}_Q = (x_1, \dots, x_{3t})$ and we will show that the remaining $4t$ variables (y_1, \dots, y_{4t}) are redundant when F has no trivial components. A connected normal surface containing only elementary T-disks is a 2-sphere or disk frontier of a regular neighborhood of a vertex of \mathcal{T} and will be referred to as a *trivial surface*.

The 3t-tuples of non-negative integers $\vec{x} = (x_1, \dots, x_{3t})$ corresponding to normal surfaces are characterized in normal surface Q-theory by two constraints. The first is that for each tetrahedron, at most one x_i associated with that tetrahedron can be nonzero since Q-disks of different types cannot be disjointly embedded in a tetrahedron. The second constraint concerns the matching of edges of elementary disks around the interior 1-simplexes of \mathcal{T} . To describe this second constraint we consider the cell decomposition of a normal surface F given by its elementary disks. Choose an oriented interior 1-simplex e_k of \mathcal{T} and let v be a point of $e_k \cap F$. A 2-cell neighborhood U of v in F is made up of the union of the elementary disks in F containing v . Let E_1 be a Q-disk in the tetrahedron Δ with a corner at v and suppose the edges of E_1 containing v consist of an a -arc and a b -arc, where a, b are vertices of Δ . If we rotate around v in a fixed direction starting in E_1 and moving from the b -arc to the a -arc of E_1 , notice that after leaving E_1 we meet only a -disks until we reach another Q-disk, which we traverse going from an a -arc to a b -arc. The remainder of the circuit around v is through b -disks unless we encounter additional pairs of Q-disks taking us from b -arcs to a -arcs and then back again to b -arcs. This is illustrated in Figure 1.

The first step in forming the Q-matching equations is to assign to each interior 1-simplex of \mathcal{T} an orientation and a positive direction of rotation about it. In a tetrahedron Δ , only two of the three types of Q-disks meet a given 1-simplex. When we rotate in the positive direction about a 1-simplex with vertices a and b , observe that on one of the Q-disk types we traverse from the a -arc to the b -arc and on the other Q-disk type we traverse from the b -arc to the a -arc.

Consider a tetrahedron Δ of \mathcal{T} with vertices $\{a, b, c, d\}$ and take a 1-simplex $e_k = [ab]$, where the positive orientation is from a to b . Assume the positive rotation around $[ab]$ induces an orientation of the 1-simplex $[cd]$ from c to d as shown in Figure 2. Let d_i^Q and d_j^Q denote the two Q-disk types in Δ represented by elementary Q-disks meeting the edge $[ab]$. Each elementary disk of type d_i^Q and d_j^Q has an a -arc in one 2-simplex face of Δ containing $[ab]$ and a b -arc in the other 2-simplex face containing $[ab]$. If the Q-disk type d_i^Q is represented by a 4-sided disk σ in Δ disjoint from $[b, c]$ then we define the *sense* of σ (relative to the oriented 1-simplex e_k) to be

$\epsilon_{k,i} = +1$. That is, $\epsilon_{k,i} = +1$ if a positive rotation about e_k carries the face of Δ containing the a -arc of σ through Δ to the face of Δ containing the b -arc of σ . For the other Q-disk type d_j^Q , represented by a 4-sided disk disjoint from $[a, c]$ -disk, we set $\epsilon_{k,j} = -1$. For all Q-disk types d_l^Q not meeting e_k we set $\epsilon_{k,l} = 0$.

Consider the collection of elementary Q-disks in a normal surface F which have a corner at v and observe that the number having a positive sense must be equal to the number with a negative sense. This is the constraint that characterizes vectors corresponding to normal surfaces. Thus, if F is a normal surface then $\vec{F}_Q = (x_1, \dots, x_{3t})$ satisfies the following linear system of equations, one equation for each interior 1-simplex e_k of \mathcal{T} .

$$\begin{aligned}
 & \text{Q-matching equations} \\
 & \left\{ \sum_{i=1}^{3t} \epsilon_{k,i} x_i = 0 \right\}_k \\
 & 0 \leq x_i, \quad 1 \leq i \leq 3t.
 \end{aligned}$$

3. Quadrilateral Disks are Sufficient.

A non-negative integral solution (x_1, \dots, x_{3t}) of the Q-matching equations is *admissible* if it has the property that for each tetrahedron Δ of \mathcal{T} at most one of the three variables x_i corresponding to Q-disk types in Δ is nonzero. The frontiers of regular neighborhoods of the vertices of \mathcal{T} are referred to as *trivial surfaces*. We show that an admissible solution to the system of Q-matching equations uniquely determines a normal surface without trivial components. Recall that there is one Q-matching equation associated to each interior 1-simplex of the triangulation.

Theorem 1. *Let M be a compact 3-manifold with a fixed triangulation. If F is a normal surface in M then the Q-coordinates \vec{F}_Q give an admissible solution to the Q-matching equations. Moreover, if \vec{z} is a nonzero admissible solution to the Q-matching equations then there exists a unique normal surface F in M with no trivial components such that $\vec{F}_Q = \vec{z}$.*

Proof. It follows from the [previous](#) section that the Q-coordinates of a normal surface F satisfy the matching Q-equations. So suppose that \vec{z} is an admissible solution to the Q-matching equations. We will construct a normal surface F with $\vec{F}_Q = \vec{z}$.

Case 1. M is a closed 3-manifold.

We first glue together a collection of Q-disks corresponding to \vec{z} and then fill in the holes with T-disks. To picture the construction, consider a normal surface G intersecting e_k . The union of the set of tetrahedra containing e_k as an edge is a 3-ball B . Observe that the types of the elementary disks meeting e_k can be characterized by how they intersect the triangulated 2-sphere boundary of B . A component D of $G \cap B$ is a disk which can be viewed as the join $v \star \partial D$ where $v = D \cap e_k$. The disk D is the union of elementary disks meeting e_k at the point v . The disk types in D can be recognized by the arc types comprising $D \cap \partial B$.

We begin by choosing a collection \mathcal{E} of elementary Q-disks corresponding to the given solution \vec{z} . Take a 1-simplex $e_k = [ab]$ and choose an orientation, say from a to b . Let \mathcal{E}_k denote the set of pairs $\{(E, v)\}$ such that $E \in \mathcal{E}$ and v is a corner of a Q-disk E meeting e_k . The first step is to describe how to glue together the corners v of the pairs (E, v) of \mathcal{E}_k along e_k . View the 2-sphere ∂B as the union $(a \star lk([ab])) \cup (b \star lk([ab]))$ as shown in Figure 3. Let $\{\alpha_i\}_{i=1}^p$ denote the arcs in ∂B coming from the Q-disks for which $\epsilon_{k,i} = +1$ and let $\{\beta_j\}_{j=1}^q$ denote those having $\epsilon_{k,j} = -1$. According to the Q-matching equation associated with the 1-simplex e_k , the number of Q-disks from a to b is equal to the number from b to a .

There is a unique set of a -type arcs in $a \star lk([ab])$ and a unique set of b -type arcs in $b \star lk([ab])$ whose union with the set $\cup\{\{\alpha_i\} \cup \{\beta_j\}\}$ produces a normal system of pairwise disjoint simple closed curves in ∂B (after sliding the ends of the α_i and β_j together where required). To construct this set, begin in $a \star lk([ab])$ by locating and joining an outermost pair of arcs $\{\alpha_i \cap (a \star lk([ab])), \beta_j \cap (a \star lk([ab]))\}$. By *outermost* we mean that $\alpha_i \cap (a \star lk([ab]))$ can be joined to $\beta_j \cap (a \star lk([ab]))$, by either a -type arcs or by sliding ends together, to form a properly embedded arc η in $a \star lk([ab])$ that splits off an outermost disk. After joining a pair of outermost arcs, locate and join another outermost pair from among the remaining collection of arcs that have not yet been joined, and so on. Similarly, there is a unique way to add b -type arcs in $b \star lk([ab])$ to obtain a collection of properly embedded arcs in $b \star lk([ab])$. The union of these two sets of arcs gives us the desired family $\{A_\lambda\}$ of disjoint simple closed curves in ∂B . Each curve A_λ winds around e_k exactly once and intersects the circle $lk([ab])$ along corners of elementary Q-disks in \mathcal{E}_k . Divide the Q-disks of \mathcal{E}_k into equivalence classes \mathcal{V}_λ corresponding to the simple closed curves A_λ . The corners $\{\sigma \cap e_k \mid \sigma \in \mathcal{V}_\lambda\}$ are glued together to form a common corner $v_\lambda \in e_k$. If two Q-disks in \mathcal{V}_λ each have an edge containing v_λ in the same 2-simplex then glue the two edges together also. This identification of corners and edges of Q-disks will be referred to as the *Q-corner gluing rule*.

After making all the identifications called for by the Q-corner gluing rule for the disks in \mathcal{E}_k , the surface is completed around e_k by adding elementary T-disks. For each index λ there exists a unique collection of elementary T-disks in B of types a and b that meet e_k and which complete the disk $\{v_\lambda\} \star A_\lambda$ when added to the Q-disks in \mathcal{V}_λ . Let \mathcal{E}'_k denote the collection of all T-disks in B added at this time and observe that $\mathcal{E}_k \cup \mathcal{E}'_k$ is a pairwise disjoint collection of 2-cells of the form $\{v_\lambda\} \star A_\lambda$.

This local Q-corner gluing rule is well-defined globally. Suppose we have two distinct applications of the Q-corner gluing rule that apply to a pair of Q-disks E and E' . Then E and E' lie in adjacent tetrahedra Δ, Δ' , respectively, and intersect the 2-simplex $\sigma = \Delta \cap \Delta'$ in normal isotopic arcs as shown in Figure 4. Suppose σ has vertices $\{a, b, c\}$ and let $e_k = [ab]$. Assume that the Q-corner gluing rule corresponding to the edge e_k stipulates that a corner from each of E and E' are identified with v_λ on e_k . Under the Q-corner gluing rule for e_k , the edges $E \cap \sigma$ and $E' \cap \sigma$ are also glued together to form a common edge with endpoints $\{u_\lambda, v_\lambda\}$. Let e_l denote the 1-simplex $[ac]$ containing u_λ . The gluing rule for e_l also applies to E and E' . Let $\mathcal{E}_{k,l}$ denote the subset of \mathcal{E}_k consisting of those disks meeting $\overset{\circ}{\sigma}$. The collection $\mathcal{E}_{k,l}$ contains only Q-disks of the two disk types represented by E or E' . Because of the parallelity of the Q-disks, the identification instructions associated with the 1-simplex e_l applied to the pairs of Q-disks in $\mathcal{E}_{k,l}$ agrees with that for e_k .

Let F_1 denote the union of the Q-disks $\cup \mathcal{E}_k$ with the corners identified via the Q-corner gluing rule. We need more notation in order to describe how to fill in F_1 with T-disks to form F . Fix a vertex a and let e_k be a 1-simplex with endpoints a, b . We have the 3-ball C_a obtained by taking the union of the tetrahedra containing a and another 3-ball $B_k \subset C_a$ obtained by taking the union of the tetrahedra containing e_k . There is a simplicial projection $p: C_a - a \rightarrow S_a$ onto the 2-sphere $S_a = \partial C_a$.

Each type of elementary disk in C_a projects into S_a in a distinct way. Consider an elementary disk τ in C_a with a corner on e_k . If τ is an a -type T-disk then $p(\tau)$ is a 2-simplex in S_a . If τ is a b -type T-disk then $p(\tau)$ is a triangle cut out of a 2-simplex in S_a by a normal arc. If τ is a Q-disk then $p(\tau)$ is a quadrilateral cut out of a 2-simplex in S_a by a normal arc. These projections are illustrated in Figure 5.

Let \hat{F}_1 denote the union of $F_1 \cap C_a$ with all the T-disks in $\cup \mathcal{E}'_k$ which are contained in C_a and are not of the a -type. Fix a component W of \hat{F}_1 . The following properties of the projection $p(W) \subset S_a$ can be established by considering the restriction of p to a typical 3-ball B_k in C_a . A component α of $W \cap S_a$ is a simple closed curve. Let Γ_α denote the union of all the Q-disks and (non- a -type) T-disks in W which have an edge in α . Then $p(\Gamma_\alpha) \subset S_a$

can be viewed as an annulus with identifications in one boundary component along pairs of vertices and pairs of 1-simplices. Thus one boundary curve is α and the second boundary is, after the identifications, a graph α' contained in the 1-skeleton of S_a . Hence $p(\Gamma_\alpha) - \alpha'$ is an half-open annulus.

Now consider a second component β of $\hat{F}_1 \cap S_a$. The two possible ways that $p(\Gamma_\alpha)$ can intersect $p(\Gamma_\beta)$ are illustrated in Figures 6 and 7. Suppose $p(\Gamma_\alpha)$ and $p(\Gamma_\beta)$ intersect along α' and β' such that the corresponding half-open annuli lie on opposite sides as illustrated in Figure 6. Then all intersections between $p(\Gamma_\alpha)$ and $p(\Gamma_\beta)$ are also of this same type. This follows from the fact that the simple closed curve α bounds a disk in the 2-sphere S_a and thus β lies on only one side of α . In this case $p(W)$ is a disk-with-holes which may have isolated pairs of vertices in its boundary identified. On the other hand, suppose that $p(\Gamma_\alpha)$ and $p(\Gamma_\beta)$ intersect along α and β such that the corresponding half-open annuli lie on the same side, as illustrated in Figure 7. Then again all intersections between $p(\Gamma_\alpha)$ and $p(\Gamma_\beta)$ are of this same type. It follows that there are no Q-disk corner identifications between Γ_α and Γ_β and hence Γ_β belongs to a component of $F_1 \cap C_a$ distinct from W .

There are two kinds of boundary components of $p(W)$: (i) Components of $W \cap S_a$ made up of normal arcs and (ii) graphs in the 1-skeleton of S_a .

Consider a component X of $S_a - p(W)$ that has its boundary ∂X contained in the 1-skeleton of S_a . X is an open disk whose closure is the union of 2-simplices of S_a . Each of these 2-simplices is the image of an a -type triangular disk under the projection p . We add to W a disk D_X which is a union of a -type triangles such that ∂D_X matches up to the appropriate boundary arcs of W and $p(D_X) = X$. Do this for each such component X of $S_a - p(W)$ for which ∂X is contained in the 1-skeleton. Let W_a denote the resulting collection of disks composed of a -type triangles which fill in the boundary curves of W made up of a -type arcs.

Repeat this construction for each component W of $\hat{F}_1 \cap C_a$ and for each vertex a in the triangulation. At the end the added disks W_a may intersect each other nontrivially. We may assume that they intersect transversely and in simple closed curves. Since these intersection curves involve only triangular disks, they can be removed by making regular exchanges. Discard any trivial 2-sphere components resulting from these regular exchanges to obtain the normal surface F with $\vec{F}_Q = \vec{z}$.

The uniqueness of F follows from the construction. Alternatively, suppose that F and G are normal surfaces without trivial components such that $\vec{F}_Q = \vec{G}_Q$. Then there exist trivial surfaces Σ_1, Σ_2 such that $\vec{F} + \vec{\Sigma}_1 = \vec{G} + \vec{\Sigma}_2$. Hence $F + \Sigma_1 = G + \Sigma_2$. Since the trivial surfaces Σ_1, Σ_2 can be taken disjoint

from F and G it follows that $F = G$.

Case 2. M has nonempty boundary.

Let \vec{z} be a nonzero admissible solution to the Q-matching equations for M . Consider $2M$, the double of M , obtained by gluing a mirror image M' of M to ∂M , and the corresponding double of \vec{z} , namely $\vec{w} = (\vec{z}, \vec{z})$. It is clear that \vec{w} satisfies each Q-matching equation associated to an interior 1-simplex in M or in M' . The Q-matching equations for $2M$ associated to boundary 1-simplices of M are also satisfied by \vec{w} . The mirror image of a Q-disk in M with a corner on a boundary 1-simplex e is a Q-disk in M' having the opposite sense relative to e . Thus, by Case 1, there exists a unique normal surface G in $2M$ without trivial components such that $\vec{G}_Q = \vec{w}$. $F = G \cap M$ is the desired normal surface. \square

4. Q-vertex Surfaces.

The normal (Q-normal) coordinates of a normal surface F are solutions to the linear system of matching (Q-matching) equations associated with the triangulation. The projection ρ of the solution space onto the unit sphere $\sum |x_i| = 1$ is a compact, convex, linear cell called the *projective solution space* \mathcal{P} (Q-projective solution space \mathcal{P}_Q). Normal surfaces correspond, up to multiples, to the admissible rational solutions in \mathcal{P} . The connected, two-sided, normal surfaces corresponding to the vertices of \mathcal{P} are called *vertex surfaces* and such nontrivial normal surfaces corresponding to the vertices of \mathcal{P}_Q are called *Q-vertex surfaces*. It is an elementary exercise in linear algebra to construct the list of vertex (Q-vertex) surfaces. These surfaces can be characterized geometrically in the following manner. Let F be a connected, two-sided normal surface. F is a vertex surface if and only if multiples of F are the only two-sided normal surfaces X, Y such that $nF = X + Y$ for some integer n . Similarly, F is a Q-vertex surface if and only if whenever X, Y are two-sided normal surfaces such that $nF + \Sigma = X + Y$, for some integer n and family of trivial surfaces Σ , then each component of X and Y is normal isotopic to either F or a component of Σ .

We say that a sum $F = X + Y$ is in *reduced form* if the number of intersection curves in $X \cap Y$ is minimal relative to all normal surfaces X', Y' isotopic to X, Y , respectively, such that $F = X' + Y'$. The *weight* of a normal surface F (or disk D embedded in F) is the number of intersection points of F (of D) with the 1-skeleton of the triangulation.

Theorem 2. *Let M be a triangulated, compact, irreducible, ∂ -irreducible 3-manifold. If there exists a two-sided, incompressible, ∂ -incompressible surface in M then there exists one that is a Q-vertex surface.*

Lemma 3. *Let F be a connected, two-sided, incompressible, ∂ -incompressible vertex surface in M . Assume that F is a vertex surface and is also least weight in its isotopy class. If X, Y are two-sided normal surfaces such that $n\vec{F}_Q = \vec{X}_Q + \vec{Y}_Q$, for some positive integer n , then some components of X and Y are isotopic to F and the remaining components are disks and 2-spheres.*

Proof. By Theorem 1, $nF + \Sigma = X + Y$ where Σ is a collection of disjoint trivial normal 2-spheres and disks. There is no loss of generality in considering a normal sum $nF + \Sigma = X' + Y'$ in reduced form where X', Y' are isotopic to X, Y , respectively. In particular, since M is irreducible and ∂ -irreducible we may assume that there does not exist an adjacent pair of disk patches (see [7] for example).

First suppose that each pair of trace curves belongs to either nF or Σ . In this case we can write X' and Y' each as the disjoint union of X'_1, X'_2 and Y'_1, Y'_2 , respectively, such that $nF = X'_1 + Y'_1$ and $\Sigma = X'_2 + Y'_2$. Since F and each component of Σ are vertex surfaces it follows that each component of X'_1 and Y'_1 is normal isotopic to a multiple of F and each component of X'_2 and Y'_2 is normal isotopic to a component of Σ .

It remains to consider a sum $nF + \Sigma = X' + Y'$ in reduced form for which there exists a pair of trace curves (β', β'') where $\beta' \subset nF$ and $\beta'' \subset \Sigma$. There exists an associated 0-weight annulus with frontier $\beta' \cup \beta''$ that spans nF and Σ . Our conclusion will follow if we show that this leads to a contradiction. Since the given trace curves are inessential, β' is the frontier of a disk in nF and we can find a pair of trace curves $\{\alpha', \alpha''\}$ such that α' is the frontier of a disk patch D in nF . The second trace curve α'' is the frontier of a disk E which may lie in either in either nF or Σ .

Case 1. For every pair of trace curves $\{\alpha', \alpha''\}$ for which α' is the frontier of a disk patch in nF , the second trace curve α'' is also subset of nF .

Let $\{\alpha'_1, \alpha''_1\}$ be pair of trace curves such that α'_1 is the frontier of a least weight disk patch D_1 in nF and $\alpha''_1 \subset nF$. Then α''_1 is the frontier of a disk E_1 in nF which, because of the irreducibility and ∂ -irreducibility of M , must be adjacent to D_1 in the sense that a regular exchange along the corresponding intersection curve α_1 interchanges pieces of D_1 and E_1 near α_1 . For otherwise one could construct a 2-sphere separating nF by joining together disks parallel to D_1 and E_1 . Observe that $wt(E_1) = wt(D_1)$ since replacing E_1 by a copy of D_1 produces a normal surface isotopic to F and F is already of least weight. The disk E_1 cannot contain two distinct disk patches because this would violate the minimal weight of D_1 among disk patches. Hence the set of disks in E_1 bounded by trace curves must be nested and all with the same weight as that of D_1 .

E_1 itself cannot be a disk patch (adjacent to D_1) since the sum $X' + Y'$ is in reduced form. Thus E_1 must contain a disk patch D_2 which in turn is adjacent to a disk E_2 , also in nF . Repeating this process, one eventually encounters a sequence without repetitions of the form (after relabeling if necessary) $D_{n+1} = D_1 \subset E_n, D_2 \subset E_1, \dots, D_{n+1} \subset E_n$ (this is similar to the construction in Lemma 2.1 of [4]). Perform regular exchanges along all the intersection curves in $X' \cap Y'$ except those corresponding to the trace curves $fr(D_i)$ and discard any components of Σ appearing after summing. What is left is a compressible normal torus or Klein bottle T , consisting of the annuli $E_i - \overset{\circ}{D}_{i+1}$ glued together, and a normal surface G isotopic to nF by an isotopy pushing each E_i to D_i . This gives the normal sum $nF = T + G$, which contradicts that F is a vertex surface.

Case 2. There exists a pair $\{\alpha', \alpha''\}$ of trace curves such that α' is the frontier of a disk patch D in nF and $\alpha'' \subset \Sigma$.

α'' is the frontier of a disk $E \subset \Sigma$ which can be chosen to be adjacent to D . Let A denote the 0-weight annulus associated with the pair of trace curves. The 2-sphere $D \cup E \cup A$ bounds a 3-cell B whose interior is disjoint from nF but may contain components of Σ . Since the sum is in reduced form, E cannot be a disk patch and there exists an innermost trace curve α'_1 which is the frontier of a disk patch D_1 in the E . The matching trace curve α''_1 must lie in a component of Σ which is contained either in B or in the 3-cell C bounded by the component of Σ containing E .

Suppose that $\alpha''_1 \subset C$. Note that $\Sigma \cap C$ is a nested family S_1, S_2, \dots, S_{k+1} of trivial surfaces normal isotopic to $fr(C)$, where S_{i+1} is inside the 3-cell bounded by S_i . Now $D_1 \subset S_1$ and α''_1 is the frontier of a disk $E_1 \subset S_2$ such that E_1 is adjacent to D_1 . Again E_1 cannot be a disk patch and so there exists an innermost trace curve $\alpha'_2 \subset E_1$ which is the frontier of a disk patch $D_2 \subset E_1 \subset S_2$. Since we cannot have a pair of trace curves inside C which are the frontier of adjacent disk patches, this sequence must continue with disk patches $D_i \subset S_i$ with adjacent disks $E_i \subset S_{i+1}$. Eventually we reach the innermost component of Σ and find a pair of adjacent disk patches D_k, E_k . But this is impossible since the sum is in reduced form.

Hence $\alpha''_1 \subset B$. We find ourselves in the same situation as before. There exists a sequence of disk patches lying on a nested sequence of trivial normal surfaces which must eventually end with a pair of adjacent disk patches inside B . This again gives a contradiction to the sum being in reduced form and the proof is complete. \square

Proof of Theorem. Choose any least weight, incompressible, ∂ -incompressible surface and let \mathcal{C} denote the face of \mathcal{P} carrying it in its interior. Let F be any two-sided normal surface F corresponding to a vertex of \mathcal{C} . It follows

from [4] and [6] that F is incompressible and ∂ -incompressible. Furthermore, F is also least weight in its isotopy class [8]. We show that there exists a Q-vertex surface isotopic to F .

If F is not a Q-vertex surface then let \mathcal{C}_Q denote the Q-face carrying \vec{F}_Q in its interior. Given a vertex of \mathcal{C}_Q we can choose rational points \vec{X}_Q and \vec{Y}_Q such that the projection $\rho(\vec{X}_Q)$ is a the given vertex and $n\vec{F}_Q = \vec{X}_Q + \vec{Y}_Q$ for some positive integer n . Hence, by Theorem 1 there exist normal surfaces X and Y such that $nF + \Sigma = X + Y$ where Σ is a collection of disjoint trivial normal 2-spheres and disks. We may assume that X and Y are both two-sided surfaces and that all intersection curves are also two-sided by replacing, if necessary, nF by $2nF$.

By Lemma 3, each component of X and Y is either a disk, a 2-sphere, or isotopic to F . If the Q-vertex surface X can be chosen such that it is isotopic to F we are done. Otherwise, we have the situation in which all the Q-vertex surfaces associated to the vertices of \mathcal{C}_Q are either disks or 2-spheres. We will show that this is impossible.

We can write $nF + \Sigma = S_1 + S_2 + \dots + S_k$ where the S_j are Q-vertex surfaces associated with the face \mathcal{C}_Q . Assume that all the S_j are disks or 2-spheres. We can apply Lemma 3 to the sum $nF + \Sigma = (S_1 + \dots + S_i) + (S_{i+1} + \dots + S_k)$. Let i be the largest integer such that all components of $X = S_1 + \dots + S_{i-1}$ are disks and 2-spheres. Then $1 < i \leq k$ and S_i is a disk or 2-sphere such that $X + S_i$ has a component G isotopic to F . Thus G can be viewed as being built by gluing together components of $(X \cup S_i) - (X \cap S_i)$, which are all disks or disks-with-holes. Stated differently, G is decomposed by trace curves, which are all inessential, into planar pieces coming from the components of $(X \cup S_i) - (X \cap S_i)$. But such trace curves can only split off disks from G and such a decomposition is clearly impossible. Therefore some Q-vertex S_j must be isotopic to F . \square

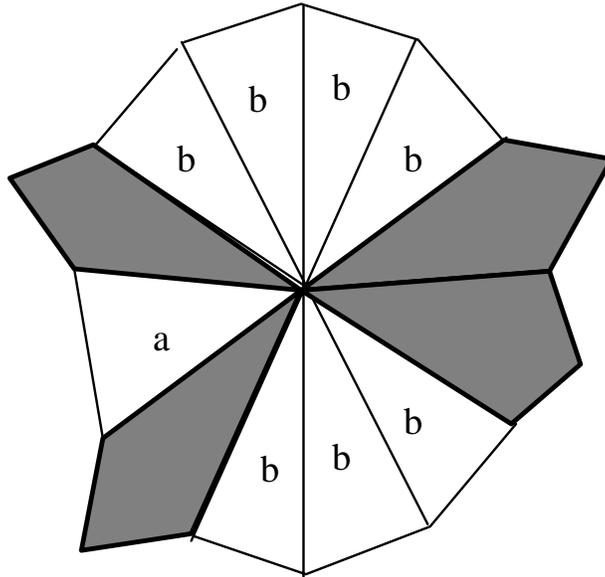


Figure 1. Elementary disk patterns in a normal surface.

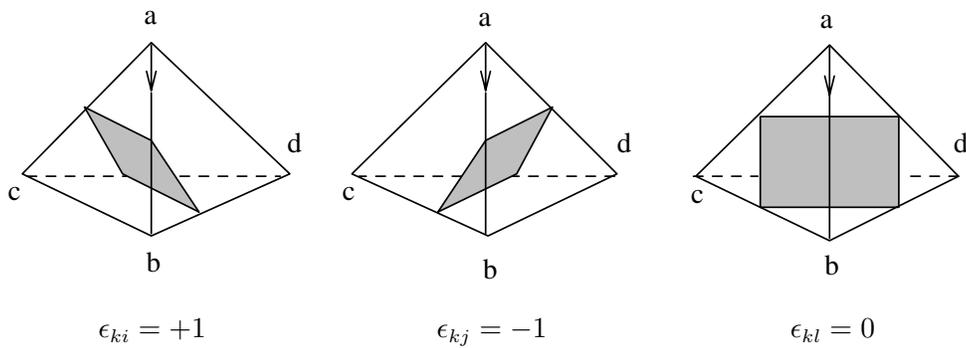


Figure 2. Types of Q-disks relative to the oriented 1-simplex $e_k = [ab]$.

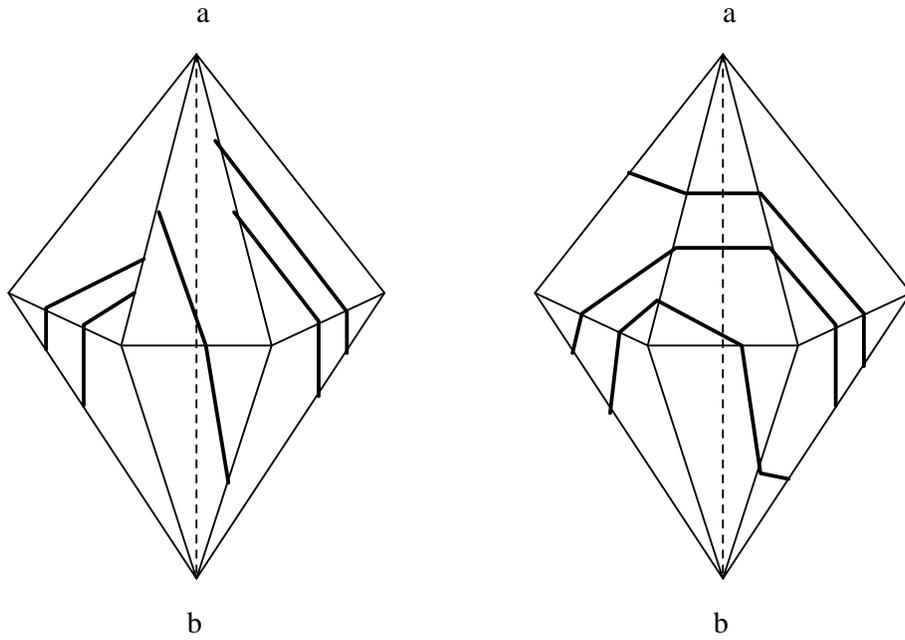


Figure 3. Arcs in the 2-sphere B before and after gluing.

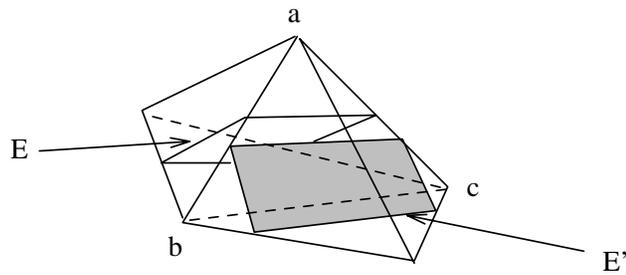


Figure 4. Quadrilaterals subject to two gluing rules.

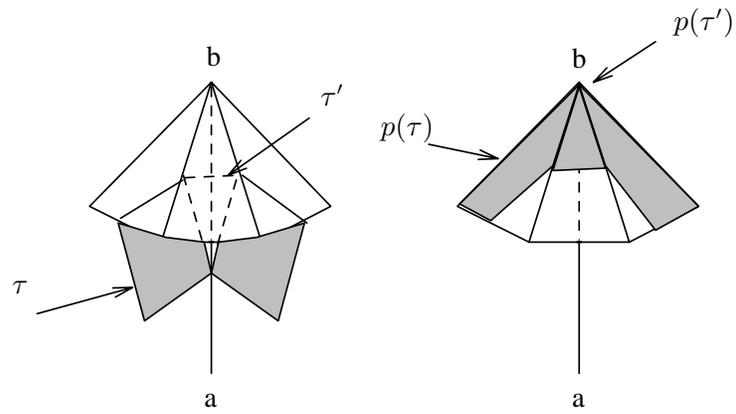


Figure 5. Projections of elementary disks onto S_a

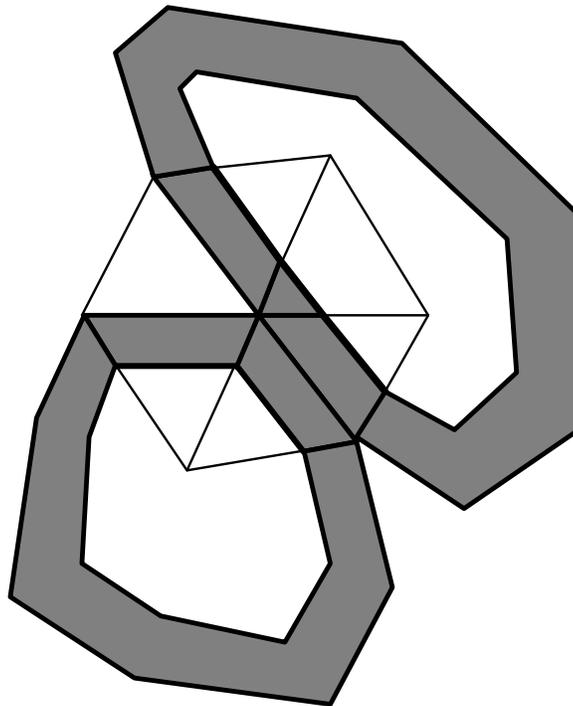


Figure 6. The annuli $p(\Gamma_\alpha) - \alpha'$ and $p(\Gamma_\beta) - \beta'$ lie on opposite sides of the intersecting boundaries.

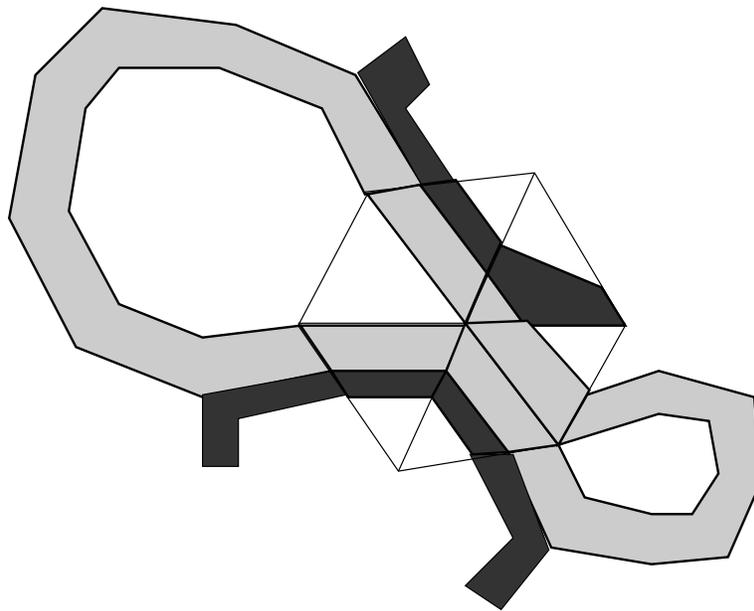


Figure 7. The annuli $p(\Gamma_\alpha) - \alpha'$ and $p(\Gamma_\beta) - \beta'$ lie on the same side of the intersecting boundaries.

References

- [1] W. Haken, *Theorie der normal Flächen*, Acta Math., **105** (1961), 245-375.
- [2] ———, *Ein verfahren aufspaltung einer 3-mannigfaltigkeiten in irreduzible 3-mannigfaltigkeiten*, Math. Z., **76** (1961), 427-467.
- [3] ———, *Über das homöomorphieproblem der 3-mannigfaltigkeiten I*, Math. Zeit., **80** (1962), 89-120.
- [4] W. Jaco and U. Oertel, *An algorithm to decide if a 3-manifold is a Haken manifold*, Topology, **23** (1984), 195-201.
- [5] W. Jaco and J. H. Rubinstein, *PL equivariant surgery and invariant decompositions of 3-manifolds*, Advances in Mathematics, **73** (1989), 149-191.
- [6] W. Jaco and J. Tollefson, *Algorithms for the complete decomposition of a closed 3-manifold*, Illinois J. of Math., **39** (1995), 358-406.
- [7] J. Tollefson, *Innermost disk pairs in least weight normal surfaces*, Topology and Its Applications, **65** (1995), 139-154.

- [8] ———, *Isotopy classes of incompressible surfaces in irreducible 3-manifolds*, Osaka Journal of Mathematics, **32** (1995), 1087-1111.

Received October 28, 1996 and revised March 24, 1997.

UNIVERSITY OF CONNECTICUT
STORRS, CT 06269-3009
E-mail address: tolfeso@math.uconn.edu