

## PL MINIMAL SURFACES IN 3-MANIFOLDS

WILLIAM JACO & J. HYAM RUBINSTEIN

### Introduction

In [12], we studied the application of least weight normal surfaces to the equivariant decomposition theorems of 3-manifolds. The basic idea in [12] was to investigate ways in which such surfaces can intersect, and then to find new surfaces by cutting and pasting. Here we refine the notion of least weight normal surfaces to obtain piece-wise linear (PL) minimal surfaces in 3-manifolds. The main technique for analyzing PL minimal surfaces is by examining PL area under small variations of the surfaces. This yields new proofs of recent applications of least area (minimal) surfaces to the topology of 3-manifolds in [23] and [14], without going through the difficult existence theory for classical analytic minimal surfaces (cf., e.g., [22]).

Also we are able to show these PL minimal surfaces have most of the basic properties of analytic minimal surfaces. Some examples are as follows. We define *PL mean curvature*  $H$  for arbitrary normal surfaces and  $H \equiv 0$  is necessary and sufficient for a normal surface to be PL minimal. In fact, there is a basic first variation argument for PL minimal surfaces, as in the analytic case. PL minimal surfaces satisfy a maximum principle and one can perform the crucial Meeks-Yau exchange and roundoff trick (see [16]–[20]). The notion of a convex boundary for a triangulated 3-manifold can be defined and so barrier arguments are possible (see, e.g., [18]). If two PL minimal surfaces intersect at a point of “tangency,” then either the local picture is like a generalized saddle, as for analytic minimal surfaces (see [1]), or the surfaces locally coincide where they meet the 2-skeleton.

Finally the important results of Freedman-Hass-Scott [5] on the fundamental topological properties of least area surfaces go through in the PL case. Let

---

Received February 18, 1986 and, in revised form, November 10, 1986. The research of the first author was supported in part by National Science Foundation Grant DMS-8602536, a grant from the Mathematical Sciences Research Institute, and the Oklahoma State University Research fund. The research of the second author was supported in part by a grant from the Mathematical Sciences Research Institute.

us call a map  $f: (F, \partial F) \rightarrow (M, \partial M)$  of a compact connected surface  $F$ , distinct from a 2-sphere or disk, to a 3-manifold  $M$ ,  $\pi_1$ -injective if  $f_*: \pi_1(F, \partial F) \rightarrow \pi_1(M, \partial M)$  and  $f_\#: \pi_1(F) \rightarrow \pi_1(M)$  are both one-to-one. In the homotopy class of a  $\pi_1$ -injective map there are always PL least area surfaces if  $M$  is  $P^2$ -irreducible (cf. Theorem 4). The techniques of [5] show that finite coverings of two-sided  $\pi_1$ -injective PL least area surfaces are PL least area and prove that such surfaces have the least number of double curves in their homotopy classes. Also two such surfaces have the smallest number of intersection loops in their homotopy classes. Somewhat surprisingly, we will give analogous “local” results for surfaces which are merely PL minimal and not necessarily  $\pi_1$ -injective or two-sided, by working in normal homotopy classes.

A serious obstacle to the normal surface method developed in [12] is that the technique seems to work well only if one is pulling embedded surfaces completely apart. In fact, if two embedded  $\pi_1$ -injective normal least weight surfaces can be homotoped apart, then a complete cut and past (exchange) along all their curves of intersection can be performed, yielding a new pair of normal least weight embedded surfaces isotopic to the original surfaces (see Proposition 3.7 of [12]). However if there is no homotopy which yields disjoint surfaces, as in e.g. the Seifert fiber space case (see [23] and [14]), this method appears to fail. Note also that the Meeks-Yau exchange and roundoff trick cannot be directly applied to normal surfaces. The problem is that there is a finite ambiguity in the choice of least weight normal surface, for example in the homotopy class of a  $\pi_1$ -injective surface.

PL minimal surfaces are defined relative to a nice Riemannian metric on the 2-skeleton  $\mathcal{T}^{(2)}$  of a fixed triangulation  $\mathcal{T}$  of a particular 3-manifold  $M$ . Then given any normal surface  $F$  in  $M$ , we define the *PL area* of  $F$  as the pair  $(\omega, l)$  lexicographically ordered, where  $\omega$  is the weight of  $F$ , i.e. the number of points where  $F$  crosses the 1-skeleton  $\mathcal{T}^{(1)}$  (see [12] and §1) and  $l$  is the total length of the arcs in  $F \cap \mathcal{T}^{(2)}$ .  $F$  is called *PL minimal* if  $l$  is stationary relative to small variations of  $F$  and  $F$  is a *PL least area* surface if the pair  $(\omega, l)$  is smallest in the homotopy class of  $F$ , e.g. in the case that  $F$  is a  $\pi_1$ -injective surface. Note that to be precise we should specify a map  $f: F \rightarrow M$  and use  $f(F) \cap \mathcal{T}^{(2)}$  instead of  $F \cap \mathcal{T}^{(2)}$ ; however we will often suppress  $f$  and not distinguish between  $F$  and  $f(F)$  if no confusion is likely. It immediately follows from the definitions that if  $F$  is PL minimal or PL least area then  $F$  meets any 2-simplex in  $\mathcal{T}^{(2)}$  in a collection of *geodesic* arcs.

The concept of PL least area in a sense picks out the *best* least weight normal surface in a homotopy class. Note that Gulliver-Scott [6] have found examples of least area singular  $\pi_1$ -injective tori in Seifert fiber spaces with  $S^2$  as orbit surface and three exceptional fibers, which have triple points but are

homotopic to tori with only double curves. Presumably similar examples can be found for PL least area surfaces.

In §1 we give the definitions of a suitable Riemannian metric on  $\mathcal{F}^{(2)}$ , PL mean curvature, PL area and PL minimal surfaces. We also include a very brief discussion of normal surfaces (see [12], [11], and [7] for more details).

In §2 the existence of PL minimal and PL least area surfaces is established. We show that PL minimal surfaces are uniquely specified in normal homotopy classes and discuss first and second variation of length  $l$  and energy for normal surfaces. We observe that one can use energy  $E$  instead of arc length  $l$ , as in the theory of geodesics (see, e.g., [21]).  $E$  has the advantage of being a *convex* function and so possesses a *unique minimum*. A physical model for minimizing energy, which was suggested to us by Bill Thurston, is outlined. Finally the other basic properties of PL minimal surfaces (maximum principle, exchange and round off, nontransversality, barriers) are given.

In §3 the results of Scott [23], Meeks-Scott [14] and [5] are reviewed. Brief descriptions of proofs using PL least area surfaces are given. We also indicate how proofs of the equivariant decomposition theorems of Meeks-Yau [16]–[20] and Meeks-Simon-Yau [15] can be done, using PL least area surfaces rather than least weight normal surfaces, as in [12] (see also [3], [4]). For the latter result, an *isotopy* minimization proposition for spheres not bounding fake cells is needed (see §3). In fact we show generally that PL minimal surfaces are the simplest surfaces in normal homotopy classes, in the sense of [5], i.e. have smallest numbers of double curves. Consequently a PL minimal surface normally homotopic to an embedded surface must be embedded or a double cover of an embedding.

We also show in the Appendix that the metric on the 2-skeleton  $\mathcal{F}^{(2)}$  can be perturbed so that any finite collection of PL minimal surfaces are in general position relative to each other and the triangulation. This is useful for arguments like [5] counting numbers of double curves and intersection curves.

We have included only one new result involving PL minimal or least area surfaces, as compared to classical analytic minimal and least area surfaces. However we show here and in [12] that normal surfaces and PL minimal surfaces suffice to give most of the recent applications to 3-manifold theory, without deep existence theory from partial differential equations and geometric measure theory. (See Hass-Scott [8] for a new approach to existence questions, using topology and Morrey's solution of Plateau's problem in Riemannian 3-manifolds.) We are hopeful that the very algorithmic, explicitly computable nature of normal surfaces may lead to additional uses of PL minimal surfaces. One interesting problem is that it should be possible to subdivide a triangulation arbitrarily finely and obtain sequences of PL minimal surfaces which

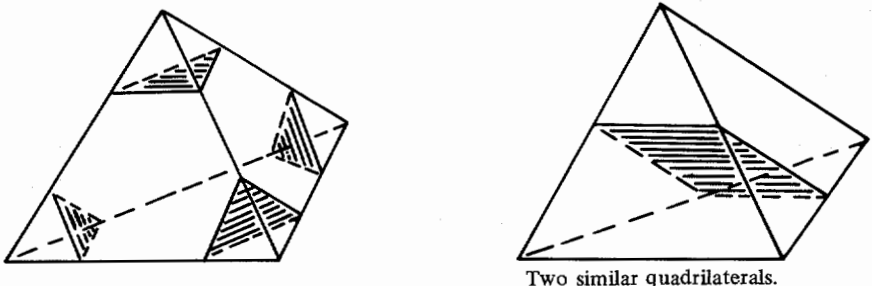
converge to classical analytic minimal surfaces. This would be analogous to some of the early work on the Dirichlet principle (see, e.g., Hilbert's approach [10]).

The idea of putting a Riemannian metric on the 2-skeleton  $\mathcal{T}^{(2)}$  grew out of the analysis in [12] of intersections of least weight normal surfaces, especially the method of viewing the picture as spanning arcs crossing in 2-simplices of  $\mathcal{T}^{(2)}$ .

### 1. Normal and PL minimal surfaces

We will use  $M$  to denote a smooth 3-manifold with a fixed triangulation  $\mathcal{T}$  and  $F$  to denote a compact connected surface throughout.  $\partial X$  will denote the boundary of  $X$ . An immersion  $f: F \rightarrow M$  will be called a *two-sided* surface if it has a trivial normal bundle. We call  $M$   *$P^2$ -irreducible* if there are no two-sided projective planes embedded in  $M$  and if any embedded 2-sphere bounds a 3-cell in  $M$ .  $\mathcal{T}^{(i)}$  will denote the  $i$ -skeleton of the triangulation.

We now give a very brief description of normal surfaces. In §1 of [12] a more detailed review is included and for a full exposition, see [11]. Let  $\tau$  be a 3-simplex of  $\mathcal{T}$ . There are seven special disks properly embedded in  $\tau$ , called *disk types*, which are shown in Figure 1. A *surface*  $F$  in  $M$  will always mean a proper immersion  $f: (F, \partial F) \rightarrow (M, \partial M)$ , where possibly  $\partial F$  and  $\partial M$  are empty. A homotopy of  $f$  will always mean a homotopy of maps  $(F, \partial F) \rightarrow (M, \partial M)$ . We will usually not distinguish in notation between  $F$  and  $f(F)$ .  $F$  will be called a *normal surface* if  $F$  meets transversely each 3-simplex of  $\mathcal{T}$  in a finite collection of (possibly intersecting) disk types. Note that in [12], only embedded normal surfaces were considered, but it is straightforward to work with singular normal surfaces.



Disk types for normal surfaces in a 3-simplex.

FIGURE 1

The *weight*  $\omega$  of a normal surface  $F$  is defined as the number of points in  $F \cap \mathcal{T}^{(1)}$ , counted with multiplicity. So if  $n$  disk types of  $F$  in some 3-simplex all meet a particular edge in the same point, we count this point  $n$  times. We will sometimes refer to the *sheets* of  $F$  at a point  $x$ . Precisely, given an immersion  $f: F \rightarrow M$  and  $x \in f(F)$ , then  $f^{-1}(x) = x_1 \cup \dots \cup x_n$  and we can choose small disjoint disks  $D_i \subset F$ ,  $1 \leq i \leq n$ , with  $x_i \in D_i$  and  $f|_{D_i}$  an embedding. Then the  $f(D_i)$  are the sheets of  $F$  at  $x$ . The number of sheets is the *multiplicity* of  $x$ .

**Remark 1.** A normal surface can be thought of as a surface which meets  $\mathcal{T}^{(1)}$  in a minimal number of points amongst nearby surfaces. (See §2 and also [13], [7], [9] for the basic result of Kneser.) One can make an analogy between minimal surfaces and normal surfaces by “blowing up” a Riemannian metric on  $M$  along  $\mathcal{T}^{(1)}$ , i.e. concentrate all the area of a surface  $F$  near  $\mathcal{T}^{(1)}$  by multiplying the metric by a very large bump function in a small neighborhood of  $\mathcal{T}^{(1)}$ . Then the weight  $\omega$  of  $F$  is a good measure of the area of  $F$ .

A *normal homotopy* is defined to be a smooth map  $\xi: F \times [0, 1] \rightarrow M$  so that for each fixed  $t \in [0, 1]$ , the surface  $F_t$  given by  $\xi|_{F \times \{t\}}$  is a normal surface.  $\xi$  is a *normal isotopy* if in addition, each  $F_t$  is embedded.

**Remark 2.** A normal homotopy can be viewed as a compatible collection of isotopies of each of the normal disk types of the initial normal surface  $F_0$  in the 3-simplices of  $\mathcal{T}$ .

The *normal homotopy class*  $\mathcal{N}(f)$  of a normal surface  $f: F \rightarrow M$  is defined as the set of all normal surfaces  $g: F \rightarrow M$  which are normally homotopic to  $f$ .

We now define a simple Riemannian metric on  $\mathcal{T}^{(2)}$ . For each 2-simplex  $\sigma$  in  $\mathcal{T}^{(2)}$  we identify  $\sigma$  with an ideal hyperbolic triangle in the hyperbolic plane  $\mathbf{H}^2$  (see Figure 2). This gives an induced metric on  $\sigma$ . Finally we glue these hyperbolic 2-simplices together by isometries between their common edges. (There is some freedom of choice here which does not affect matters.) For later

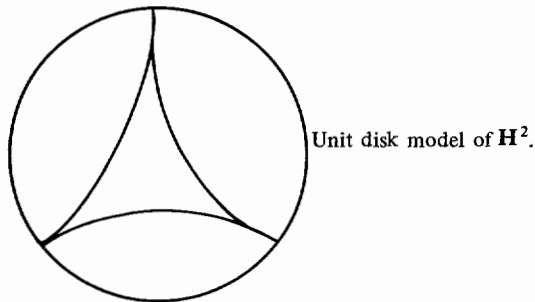


FIGURE 2

applications, we will sometimes be given a group  $G$  of simplicial homeomorphisms of  $M$ , for which  $\text{Fix } \eta = \{x \in M: \eta x = x\}$  is a subcomplex of  $\mathcal{T}$  for each  $\eta \in G$ . In this case we choose the identifications and gluings so that  $G$  acts isometrically on  $\mathcal{T}^{(2)}$ . (For example, we can choose a metric as above on the 2-complex  $\mathcal{T}^{(2)}/G$  and lift it back to  $\mathcal{T}^{(2)}$ .) In the sequel, we will denote by  $\nabla$  covariant differentiation relative to the hyperbolic metric on the 2-simplices.

Given a normal surface  $F$  in  $M$ , define its *length*  $l$  to be  $l(F \cap \mathcal{T}^{(2)})$ , i.e. the total length of all the arcs in the boundaries of the disk types in which  $F$  meets the 3-simplices of  $\mathcal{T}$ . We will refer to these arcs as the arcs of  $F \cap \mathcal{T}^{(2)}$ . Since  $F$  is compact and misses the vertices of  $\mathcal{T}^{(0)}$ ,  $l$  is a finite sum of finite lengths, so is finite. We define the *PL area* of  $F$  to be the pair  $(\omega, l)$ , lexicographically ordered. (This is analogous to area if we first “blow up” a Riemannian metric along  $\mathcal{T}^{(1)}$  and then by a smaller factor along  $\mathcal{T}^{(2)}$ .)

A normal surface  $F$  in  $M$  is called *PL minimal* if its length  $l$  is stationary with respect to small variations of  $F$ . More precisely, let  $\rho: F \times (\delta, \delta) \rightarrow M$  be a normal homotopy and let  $F_s$  denote the normal surface given by  $\rho|_{F \times \{s\}}$  with  $F_0 = F$ . Let  $l(s) = l(F_s)$ . Then  $F$  is PL minimal if the derivative  $l'(0) = 0$  for any variation  $F_s$  of  $F$ . (Note that any sufficiently small smooth variation of  $F$  will automatically be a normal homotopy, by transversality of  $F$  and  $\mathcal{T}^{(0)} \cup \mathcal{T}^{(1)} \cup \mathcal{T}^{(2)}$ .)

A normal surface  $f: F \rightarrow M$  is called *PL least area* if the PL area of  $f$  is equal to the infimum of all PL areas of normal surfaces  $g: F \rightarrow M$  in the homotopy class of  $f$ . This will be most useful in the cases when  $f$  is  $\pi_1$ -injective or  $F$  is a disk or 2-sphere and  $f$  is an *essential* map, i.e. either  $f: (D, \partial D) \rightarrow (M, \partial M)$  represents a nontrivial element of  $\pi_2(M, \partial M)$  or  $f: S^2 \rightarrow M$  is nontrivial in  $\pi_2(M)$  or  $f: S^2 \rightarrow M$  is an embedding and  $f(S^2)$  bounds a fake 3-ball in  $M$  and does not bound a 3-ball. (A fake 3-ball is a compact contractible 3-manifold not equal to  $B^3$ .)

Clearly length and weight can be defined generally for immersed surfaces  $F$  which are transverse to  $\mathcal{T}^{(2)}$ , not just for normal surfaces. For in this case,  $F \cap \mathcal{T}^{(0)} = \emptyset$ ,  $F$  meets  $\mathcal{T}^{(1)}$  transversely in a finite number  $\omega$  of points (with multiplicity), and  $F \cap \mathcal{T}^{(2)}$  is a finite union of arcs with a finite total length  $l$ . However, if we used such general surfaces, PL minimal would not necessarily imply normal and PL least area would give normality only if we assume  $M$  is  $P^2$ -irreducible and either  $F$  is  $\pi_1$ -injective or  $F$  is an essential disk or 2-sphere which has smallest PL-area amongst all such disks or 2-spheres (cf. §2).

**Remark 3.** It is clear that if  $F$  is PL minimal or PL least area, then any arc  $\alpha$  of  $F \cap \mathcal{T}^{(2)}$  is a (hyperbolic) geodesic segment running between two sides

of an ideal hyperbolic 2-simplex. We will see in §2 that PL minimality for normal surfaces  $F$  is equivalent to this together with a condition on the angles at which the sheets of  $F$  meet the edges in  $\mathcal{F}^{(1)}$ .

The energy  $E$  of a normal surface  $F$  is defined to be the sum of the energies of the arcs  $\alpha$  of  $F \cap \mathcal{F}^{(2)}$ . As in [21], we will suppose always that  $\alpha$  is parametrized proportional to arc length, i.e. the tangent vector field  $\alpha'$  has length 1. Then the energy of  $\alpha$  is given by  $E[\alpha] = l(\alpha)^2$ . Note that in §2 we show that if  $F_s$  is a variation of  $F$  then  $E(F_s) = E(s)$  is a convex function of  $s$ . This will be important in showing that PL minimal surfaces are uniquely defined in normal homotopy classes. We would like to thank Bill Thurston for suggesting to us the consideration of energy.

Finally let  $f: F \rightarrow M$  be any normal surface. The mean curvature field  $H$  of  $f$  is defined as follows and is a discontinuous piece-wise smooth vector field along  $f|f^{-1}(\mathcal{F}^{(2)})$  (cf. [2, p. 3]). Let  $\alpha$  be an arc of  $f(F) \cap \mathcal{F}^{(2)}$ , let  $x \in \text{int } \alpha$  and let  $y \in \beta$ , where  $\beta$  is the arc in  $F$  with  $f(\beta) = \alpha$  and  $y$  satisfies  $f(y) = x$ . We define  $H(y) = \nabla_T T(x)$ , where  $\nabla$  is hyperbolic covariant differentiation and  $T$  is the tangent vector field  $\alpha'$  to  $\alpha$ . Note that since  $\alpha$  is parametrized with  $|T| = 1$ ,  $|H(y)|$  is the geodesic curvature of  $\alpha$  at  $x$ . On the other hand if  $y \in f^{-1}(e)$ , where  $e$  is an edge of  $\mathcal{F}^{(1)}$ , then let  $V$  be a unit vector tangent to  $e$  at  $x = f(y)$ . Clearly, there is a uniquely specified sheet of  $f(F)$  at  $x$  coming from a small disk containing  $y$ . Let  $\alpha_1, \dots, \alpha_k$  be the arcs of  $f(F) \in \mathcal{F}^{(2)}$  which have  $x$  as an endpoint and have neighborhoods of  $x$  in this sheet (see Figure 3). We parametrize the  $\alpha_i$  so that  $x = \alpha_i(0)$ ,  $1 \leq i \leq k$ ,

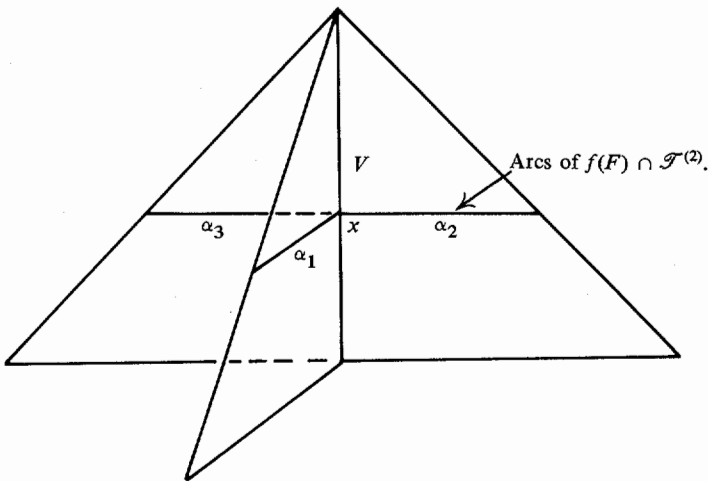


FIGURE 3

and define  $H(y) = \sum_{i=1}^k \langle T_i, V \rangle V$ , where  $T_i = \alpha'_i(0)$ . Note that  $H(y)$  is independent of the orientation of  $V$ . We will see in §2 that  $F$  is PL minimal if and only if  $H \equiv 0$  and  $H$  can be viewed in some sense as the negative gradient derivative of PL area (really of length since weight is constant for small variations). Notice that  $H(y) = 0$  for  $y \in f^{-1}(e)$ ,  $e$  an edge in  $\mathcal{T}^{(1)}$ , exactly when the sum of the cosines of the angles between the arcs  $\alpha_i$ ,  $1 \leq i \leq k$ , and the edge  $e$  is zero.

## 2. Properties of PL minimal and PL least area surfaces

A *linking 2-sphere* in a triangulation  $\mathcal{T}$  of a 3-manifold is a normal 2-sphere given by the boundary of a small 3-ball neighborhood of a vertex in  $\mathcal{T}^{(0)}$ . A multiple of a linking 2-sphere  $S$  is any normal surface obtained by gluing together all the triangles in some finite number of copies of  $S$ . By projection onto  $S$ , it can be seen such a surface is a branched cover of  $S$ .

The basic existence result for PL minimal surfaces is as follows.

**Theorem 1.** *Let  $f: F \rightarrow M$  be a normal surface which is not a multiple of a linking 2-sphere. Then there exists a PL minimal surface in the normal homotopy class  $\mathcal{N}(f)$  of  $f$ .*

*Proof.* Let  $\alpha_1, \alpha_2, \dots, \alpha_m$  denote all the arcs of  $g(F) \cap \mathcal{T}^{(2)}$ , where  $g \in \mathcal{N}(f)$ . Since a normal homotopy does not change disk types of a surface in the 3-simplices, we see that  $m$  is constant over  $\mathcal{N}(f)$ . Moreover, we can write  $\alpha_i = \alpha_i(g)$ ,  $1 \leq i \leq m$ , noting that the arc types of the  $\alpha_i$  (i.e. which edges of  $\mathcal{T}^{(1)}$  contain the endpoints of the  $\alpha_i$ ) are also unaltered as  $g$  varies over  $\mathcal{N}(f)$ . We wish to find  $g_0 \in \mathcal{N}(f)$  for which

$$l(g_0) = \sum_{i=1}^m l(\alpha_i(g_0)) = \inf\{l(g) : g \in \mathcal{N}(f)\}.$$

Then  $g_0$  will be the desired PL minimal surface, since of course length  $l$  is stationary for variations of such a surface. Let  $l_0$  denote  $\inf\{l(g) : g \in \mathcal{N}(f)\}$ .

Observe that it suffices to consider only *special normal surfaces*  $g$  for which all the arcs  $\alpha_i(g)$  are geodesics. In fact, if  $g$  is any normal surface in  $\mathcal{N}(f)$ , we can replace each arc  $\alpha_i(g)$  by the unique geodesic arc  $\gamma_i$  with the same endpoints as  $\alpha_i$ . Then clearly there is a unique special normal surface  $\bar{g}$  in  $\mathcal{N}(f)$  with  $\alpha_i(\bar{g}) = \gamma_i$ ,  $1 \leq i \leq m$ , and  $l(\bar{g}) \leq l(g)$ . Equality holds only if  $g$  is already special. Let  $\mathcal{S}(f)$  denote the subset of special normal surfaces in  $\mathcal{N}(f)$ .

A special normal surface  $g \in \mathcal{S}(f)$  is completely determined by the endpoints of the arcs  $\alpha_1(g), \dots, \alpha_m(g)$ . We assume the  $\alpha_i(g)$  have an orientation independent of  $g$ , and denote by  $x_{2i-1}$  and  $x_{2i}$  the initial and final endpoint



of  $\alpha_i(g)$  respectively. It is immediate that for  $g \in \mathcal{S}(f)$ ,  $l(g) = l(x_1, \dots, x_{2m})$  is a smooth function of the variables  $x_i$ , where each  $x_i$  can be viewed as a real parameter, after identifying all the edges in  $\mathcal{T}^{(1)}$  with unbounded hyperbolic geodesics (see Figure 1).

Let  $g_n \in \mathcal{S}(f)$  be a sequence for which  $l(g_n) \searrow l_0$ . We call  $g_n$  a *minimizing sequence* for  $\mathcal{N}(f)$ , by analogy with the classical analytic minimal surface case (see, e.g., [15]). There are two situations to consider. Firstly, suppose all the endpoints  $x_i^n$  of the arcs of  $g_n$  remain bounded, i.e. stay in bounded regions of the hyperbolic geodesic edges in  $\mathcal{T}^{(1)}$  (see Figure 4). Then we can find a subsequence of  $g_n$ , again denoted by  $g_n$ , for which  $x_i^n$  converges to  $x_i^*$  for all  $1 \leq i \leq 2m$ . Clearly  $g_n$  converges to a uniquely specified special normal surface  $g_0 \in \mathcal{S}(f)$  with endpoints of the arcs of  $g_0$  given by the  $x_i^*$ ,  $1 \leq i \leq 2m$ . Also it follows that  $l(g_0) = \lim l(g_n) = l_0$  and so  $g_0$  is a PL minimal surface in  $\mathcal{N}(f)$ , as required.

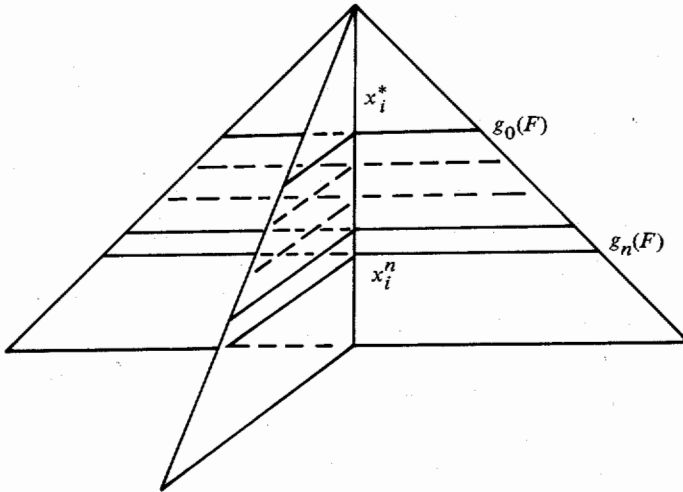


FIGURE 4

On the other hand, suppose for some  $i$  and for some subsequence of  $g_n$ , again denoted by  $g_n$ , it happens that say  $x_{2i}^n \rightarrow \infty$  in some edge  $e$  of  $\mathcal{T}^{(1)}$ , i.e.  $x_{2i}^n$  converges to an endpoint (vertex)  $v$  of  $e$ . Since  $l(g_n) \searrow l_0$ , it follows that  $l(\alpha_i(g_n))$  remains bounded and so the other endpoint  $x_{2i-1}^n$  of  $\alpha_i(g_n)$  must also converge to  $v$  (see Figure 5). This argument is obviously symmetric in  $x_{2i}^n$  and  $x_{2i-1}^n$ . So *both* ends of  $\alpha_i(g_n)$  converge to  $v$ . Similarly if  $\alpha_j$  is another arc sharing a common endpoint  $x$  with  $\alpha_i$  and in the same sheet as  $\alpha_i$  at  $x$ , then both endpoints of  $\alpha_j^n$ , i.e.  $x_{2j-1}^n$  and  $x_{2j}^n$ , must converge to  $v$  (see Figure 5). We conclude that all the endpoints  $x_i^n$ ,  $1 \leq i \leq 2m$ , converge to the same

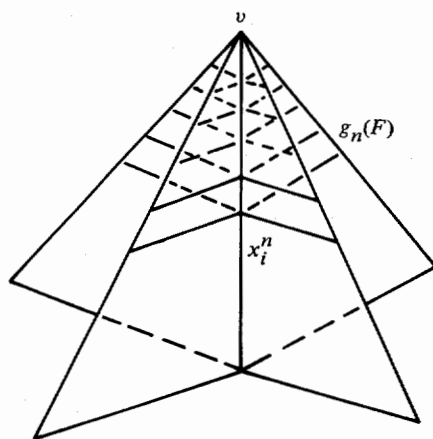


FIGURE 5

vertex  $v$  of  $\mathcal{T}^{(0)}$ . Hence the normal homotopy class  $\mathcal{N}(f)$  must contain a multiple of the linking 2-sphere of  $v$  only, contrary to hypothesis. This shows that every minimizing sequence has convergent subsequences.

Next we want to establish the uniqueness of the PL minimal surface bound in Theorem 1. To do this, we first discuss first (and second) variation of length and energy and their relationship with mean curvature for normal surfaces.

Let  $\rho: F \times (-\delta, \delta) \rightarrow M$  be a variation (normal homotopy) through normal surfaces  $F_s$ ,  $s \in (-\delta, \delta)$ , with initial normal surface  $F_0$  given by the map  $f: F \rightarrow M$ , where  $f(y) = \rho(y, 0)$  for all  $y \in F$ . We compute the derivative  $l'(0)$ , where  $l(s) = l(F_s)$ , by the first variation formula for arc length of curves, following Cheeger-Ebin (see [2, p. 5]), except that we will parametrize arcs so that tangent vectors have length 1.

Denote  $\alpha_1^s, \alpha_2^s, \dots, \alpha_m^s$  the arcs of  $F_s \cap \mathcal{T}^{(2)}$  and in the case that  $s = 0$ , we use the notation  $\alpha_i$  for these arcs, instead of  $\alpha_i^0$ . Let  $T_i$  denote the tangent vector field  $\alpha_i'$  to  $\alpha_i$  and we will suppose that the  $\alpha_i$  are parametrized so that  $|T_i| = 1$ . Then

$$\begin{aligned}
 l(s) &= \sum_{i=1}^m l(\alpha_i^s), \\
 l'(0) &= \sum_{i=1}^m l'(\alpha_i^s) \Big|_{s=0} \\
 &= \sum_{i=1}^m \langle V_i, T_i \rangle \Big|_0^{l_i} - \sum_{i=1}^m \int_0^{l_i} \langle V_i, \nabla_{T_i} T_i \rangle dt,
 \end{aligned}$$

where  $l_i = l(\alpha_i)$ , the parameter  $t$  for  $\alpha_i$  is chosen to run from 0 to  $l_i$ , and  $V_i$  is the variation vector field along  $\alpha_i$  given by

$$V_i = (\alpha_i^s)_* \left( \frac{\partial}{\partial s} \right) \Big|_{s=0}.$$

Let  $y$  be some point of  $f^{-1}(\mathcal{T}^{(1)})$  and assume  $e$  is the edge of  $\mathcal{T}^{(1)}$  with  $x = f(y)$  in  $e$ . Suppose that  $\beta_i$ ,  $1 \leq i \leq k$ , are the arcs of  $f^{-1}(\mathcal{T}^{(2)})$  with  $y$  an endpoint of  $\beta_i$  and let  $\alpha_i = f(\beta_i)$  be the corresponding arcs of  $f(F) \cap \mathcal{T}^{(2)}$ . We note that all the variation vectors  $V_i(x)$ ,  $1 \leq i \leq k$ , are necessarily equal and given some vector tangent to  $e$  at  $x$ . Without essential loss of generality, we will suppose that  $V_i(x) = V(x)$  is a unit vector tangent to  $e$  at  $x$ , for all  $1 \leq i \leq k$ . Then the first sum in the expression for  $l'(0)$  can be rewritten as

$$\sum_{i=1}^m \left\langle V_i, T_i \right\rangle \Big|_0^{l_i} = \sum_{y \in f^{-1}(\mathcal{T}^{(1)})} \langle H(y), V(f(y)) \rangle,$$

where  $x = f(y)$ . The second sum can be viewed as an integral over the interiors of the paths  $\beta_i$  in  $f^{-1}(\mathcal{T}^{(2)})$  of the form  $- \int \langle V, H \rangle dt$ , where  $V$  is the variation vector field and  $H$  is the mean curvature field of  $f$ . (Note both are vector fields along  $f|f^{-1}(\mathcal{T}^{(2)})$ .) In conclusion we can loosely write  $l'(0) = - \int \langle V, H \rangle dt$ , where we interpret this as a sum with signs over  $f^{-1}(\mathcal{T}^{(1)})$  and an integral over  $f^{-1}(\mathcal{T}^{(2)} - \mathcal{T}^{(1)})$ .

If we consider the variation of energy  $E(s) = E(F_s)$  instead of length, then clearly  $E'(0) = l'(0)$ , since we have assumed that  $|T_i| = 1$ . Hence  $F$  is PL minimal if and only if  $F$  is a stationary point for *either* length or energy under variations. This is of course also equivalent to the condition that the mean curvature field  $H$  of  $f$  is identically zero, by the above expression for  $l'(0)$ .

Finally we need to examine second variation of length and energy for normal surfaces. As in Cheeger-Ebin (see [2, pp. 20–21]), we find that for a variation  $F_s$  of  $F$ ,

$$l''(0) = \sum_{i=1}^m \left\{ \left\langle \nabla_{V_i} V_i, T_i \right\rangle \Big|_0^{l_i} + \int_0^{l_i} \left( \left| \nabla_{T_i} V_i \right|^2 - \langle R(V_i, T_i) T_i, V_i \rangle - (T_i \langle V_i, T_i \rangle)^2 \right) dt, \right.$$

using the same notation as for first variation, where  $R$  is the curvature tensor for the hyperbolic metric.

Let  $x \in f(F) \cap \mathcal{T}^{(1)}$ , let  $e$  be the edge of  $\mathcal{T}^{(1)}$  containing  $x$ , and let  $\alpha_i$  be any arc of  $f(F) \cap \mathcal{T}^{(2)}$  with  $x$  as an endpoint. Without loss of generality, we may assume the variation vector field  $V_i$  at points of  $e$  is just the tangent

vector field of unit length of the geodesic  $e$ . Hence at  $\mathcal{F}^{(1)}$ ,  $\nabla_{V_i} V_i = 0$ . Also  $\langle R(V_i, T_i)T_i, V_i \rangle = -|V_i \times T_i| \geq 0$ , since the hyperbolic metric has Gaussian curvature  $-1$ . So

$$l''(0) = \sum_{i=1}^m \int_0^{l_i} (|\nabla_{T_i} V_i|^2 + |V_i \times T_i| - (T_i \langle V_i, T_i \rangle)^2) dt.$$

This function has no obvious positivity property, due to the third term.

On the other hand if we use energy  $E(s)$  and compute  $E''(0)$  by the same method

$$E''(0) = \sum_{i=1}^m \int_0^{l_i} (|\nabla_{T_i} V_i|^2 + |V_i \times T_i|) dt.$$

This is a strictly positive function, since  $V_i$  is assumed not parallel to  $T_i$  everywhere, and so  $E$  is convex. This implies that  $E$  has at most one minimum in  $\mathcal{N}(f)$ , for any normal surface  $f: F \rightarrow M$ . But as we observed previously,  $l'(0) = E'(0)$  so critical points of  $E$  and  $l$  coincide. Combining this with the existence of at least one critical point from Theorem 1, we have completed the proof of the following:

**Theorem 2.** *There is exactly one PL minimal surface in a normal homotopy class  $\mathcal{N}(f)$  for any normal surface  $f: F \rightarrow M$  which is not a multiple of a linking 2-sphere.*

**Corollary 1.** *Suppose  $G$  is a group of simplicial homeomorphisms acting on a triangulated 3-manifold  $M$  so that for each  $\eta$  in  $G$ ,  $\text{Fix } \eta$  is simplicial. Assume  $f: F \rightarrow M$  is a normal surface which is not a multiple of a linking 2-sphere and that the metric on  $\mathcal{F}^{(2)}$  is chosen to be  $G$ -invariant. Finally suppose that for each  $\eta \in G$ ,  $f$  and  $\eta f$  are normally homotopic. Then the (unique) PL minimal surface  $g$  in  $\mathcal{N}(f)$  satisfies  $g(F) = \eta g(F)$  for all  $\eta \in G$ .*

*Proof.* By Theorem 2, since clearly  $g$  and  $\eta g$  are both PL minimal in  $\mathcal{N}(f)$ , we conclude  $g(F) = \eta g(F)$ .

A nice physical model for the procedure of finding the PL minimal surface  $g$  in  $\mathcal{N}(f)$ , where  $f: F \rightarrow M$  is any normal surface, was suggested to us by Bill Thurston. The idea is to minimize energy  $E$  rather than length  $l$ . Think of  $\mathcal{F}^{(1)}$  as a collection of wires and the endpoints  $x_{2i-1}$  and  $x_{2i}$  of the arcs  $\alpha_i$  of  $f(F) \in \mathcal{F}^{(2)}$ ,  $1 \leq i \leq m$ , as beads which are free to slide along the wires. Note that the beads are able to pass each other on the wires. Finally we view the arcs  $\alpha_i$  as rubber bands connecting the beads. Then Theorems 1 and 2 state that the PL minimal surface  $g$  is given by the *unique equilibrium (minimum energy) position* for the configuration.

We now consider analogous properties of PL minimal surfaces to those of analytic minimal surfaces.

**2.1. Maximum principle, nontransversality and barrier.** There are many forms of the maximum principle. We will be mainly interested in its application to nontransversality. Suppose that two (pieces of) normal surfaces  $f_j: F_j \rightarrow M, j = 1, 2$ , meet nontransversely at  $x \in \mathcal{F}^{(2)}$ , where possibly  $f_1 = f_2$ . There are two possibilities; either  $x \in \text{int } \sigma$ , where  $\sigma$  is a 2-simplex of  $\mathcal{F}^{(2)}$ , or  $x \in e$  for  $e$  an edge of  $\mathcal{F}^{(1)}$ . In the first case we assume there are arcs  $\alpha_j$  of  $F_j \cap \sigma, j = 1, 2$ , so that  $\alpha_1$  and  $\alpha_2$  intersect and are tangential at  $x$ . In the second case, we suppose that  $F_1$  and  $F_2$  have a common tangent plane at  $x$ . Also we can obviously make  $f_1$  and  $f_2$  transverse in the complement of  $\mathcal{F}^{(2)}$ , since we are not restricted in our choice for the interiors of the disk types in the 3-simplices.

We will view the maximum principle as giving strong restrictions on the local picture of the surfaces  $F_1$  and  $F_2$  at such a point  $x$ , in the case that  $F_1$  and  $F_2$  are PL minimal. (One can also assume  $F_1$  is PL minimal and that  $F_2$  is two-sided and has a nonzero mean curvature field so that  $F_2$  is a barrier—see later.) In fact, if  $x \in \text{int } \sigma$  then we conclude that  $\alpha_1 = \alpha_2$ , since  $\alpha_1$  and  $\alpha_2$  are then (hyperbolic) geodesics with a common tangent at  $x$ . However this does not imply  $F_1 = F_2$ , as in the case of analytic minimal surfaces. The reason is that although the part of a normal surface  $F$  in  $\mathcal{F}^{(2)}$  in some sense completely determines  $F, F \cap \mathcal{F}^{(2)}$  is still only one-dimensional.

Returning to our case of PL minimal surfaces, if  $x \in e$  then  $F_1$  and  $F_2$  have arcs in each pair of sheets at  $x$  which “interleave” as in a generalized saddle (see Figure 6 and [1] for the analytic case). Precisely, let  $\gamma_1, \dots, \gamma_k$  and  $\delta_1, \dots, \delta_l$  be the arcs of  $F_1 \cap \mathcal{F}^{(2)}$  and  $F_2 \cap \mathcal{F}^{(2)}$ , respectively, in a single sheet of  $F_1$  and  $F_2$ , respectively, at  $x$ . By PL minimality, if  $V$  is a unit tangent vector to  $e$  at  $x$ , then  $\sum_{i=1}^k \langle V, \gamma'_i(0) \rangle = 0$  and  $\sum_{n=1}^l \langle V, \delta'_n(0) \rangle = 0$  also. Here we

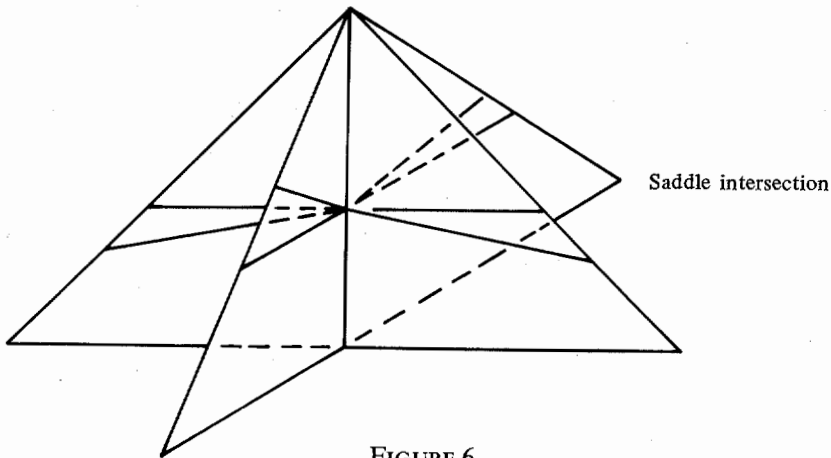


FIGURE 6

are assuming the  $\gamma_i$  and  $\delta_n$  are parametrized so that  $\gamma_i(0) = \delta_n(0) = x$  for all  $1 \leq i \leq k$  and  $1 \leq n \leq l$ . Consequently we see that the  $\gamma_i$  cannot be all "above" or all "below" the  $\delta_n$ , i.e.  $F_1$  and  $F_2$  cannot touch at the single point  $x$  locally, they must cross (see Figure 6).

In conclusion, two PL minimal surfaces  $F_1$  and  $F_2$  can coincide along finite connected unions of arcs in  $\mathcal{F}^{(2)}$ , but can still be distinct by "diverging" at the endpoints of such unions. At such an endpoint which will be in  $\mathcal{F}^{(1)}$ , the picture is either like a generalized saddle or a pair of transversely crossing sheets where the intersection curve has a subarc in  $\mathcal{F}^{(2)}$  (see Figure 7).

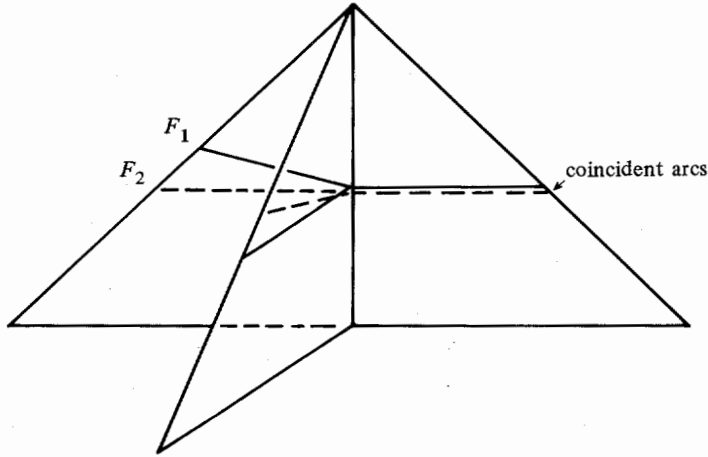


FIGURE 7

To construct barriers, suppose  $M$  is a triangulated 3-manifold with non-empty boundary. Then we can put a metric on  $\mathcal{F}^{(2)}$  so that  $\partial M$  is convex, rather than totally geodesic. We require that the mean curvature vector along an edge of  $\partial M$  always points inwards. More precisely, suppose  $\sigma$  is a 2-simplex in  $\mathcal{F}^{(2)}$  with  $\text{int } \sigma \subset \text{int } M$  and some of the edges of  $\sigma$  are in  $\partial M$ . Then we identify  $\sigma$  with a suitable *convex* subset of the hyperbolic plane, as in Figure 8.

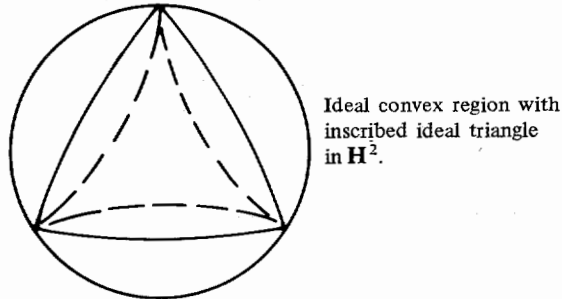


FIGURE 8

Note that if a 2-simplex  $\sigma' \subset \partial M$  has  $e$  as a common edge with  $\sigma$ , then  $\sigma'$  can still be identified with an ideal hyperbolic triangle by choosing an appropriate gluing along  $e$ .

**2.2. Exchange and roundoff trick** (cf. [16] and Lemma 1.2 of [5]).

**Lemma 1.** *Suppose  $F_1, F_2 \subset M$  are two-sided embedded PL least area surfaces in their homotopy classes and assume  $F_1$  intersects  $F_2$  transversely with  $F_1 \cap F_2$  transverse to  $\mathcal{F}$ . Then there are no product regions between  $F_1$  and  $F_2$ , i.e. there are no products  $R \times [0, 1] \subset M$ , where  $R$  is a compact connected surface, with nonempty boundary,  $R_1 = R \times \{0\} \subset F_1$  and  $R_2 = R \times \{1\} \cup \partial R \times [0, 1] \subset F_2$ .*

*Proof.* This is a fundamental technique in the application of least area surfaces to 3-manifold topology. We have presented it in the form of Lemma 1.2 of [5].

Let  $F'_1 = (F_1 - \text{int } R_1) \cup R_2$  and let  $F'_2 = (F_2 - \text{int } R_2) \cup R_1$ . Then  $F'_i$  may be singular (the interior of the product region may intersect  $F_i$ ) but clearly  $F'_i$  is homotopic to  $F_i$ ,  $i = 1, 2$ . Let PL area of a surface  $A$  be denoted by  $|A|$ . If  $|R_2| < |R_1|$  then  $|F'_1| < |F_1|$  and we obtain a contradiction to  $F_1$  being PL least area. The case when  $|R_1| < |R_2|$  is similar. Hence  $|R_1| = |R_2|$  and  $|F_i| = |F'_i|$ ,  $i = 1, 2$ . But  $F'_i$  has corners,  $i = 1, 2$ . In the analytic case this is obvious. In the PL case, since  $F_1$  and  $F_2$  are transversely intersecting PL minimal surfaces, any crossing pair of disk types of  $F_1$  and  $F_2$  in a 3-simplex  $\tau$  of  $\mathcal{F}^{(3)}$  will meet in a single spanning arc or a single point (see Figure 9). This follows because the arcs of  $F_1$  and  $F_2$  in  $\partial\tau$  are hyperbolic geodesics, so if distinct can meet in at most one point. There are two possibilities for an arc of  $\partial R_1 = \partial R_2$  in a 3-simplex  $\tau$ ; either the arc has endpoints in the interiors of 2-simplices of  $\partial\tau$  or the arc lies entirely in  $\partial\tau$ . The latter case is ruled out by the assumption that  $F_1 \cap F_2$  intersects  $\mathcal{F}$  transversely. Similarly we can exclude the case of an isolated point of  $\partial R_1 = \partial R_2$  in  $\partial\tau$ .

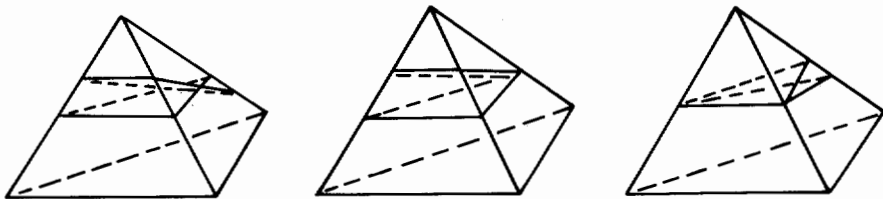


FIGURE 9

So we can conclude that every curve of  $\partial R_1$  transversely crosses the interiors of 2-simplices in  $\mathcal{F}^{(2)}$ . But then  $F'_1 \cap \sigma$ , for  $\sigma$  such a 2-simplex, has an arc with a corner which can be shortened by rounding it off (see Figure 10). This

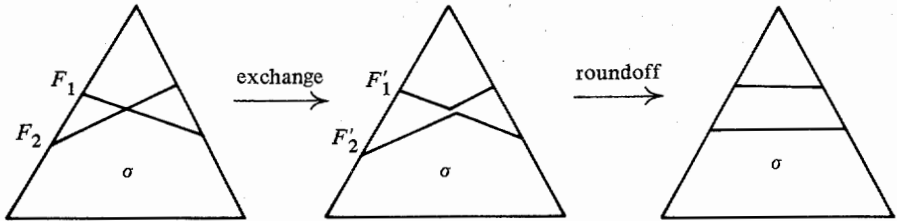


FIGURE 10

contradicts the hypothesis that  $F_1$  is least area, since  $|F_1'| = |F_1|$  and  $F_1'$  is homotopic to  $F_1$ . Therefore there are no product regions. Note that the opposite (vertical) exchange to that shown in Figure 10 would also give that  $F_1'$  and  $F_2'$  could be homotoped to normal surfaces of smaller weight, contrary to the choice of  $F_1$  and  $F_2$  as least weight surfaces.

**2.3. Meeks-Yau trick** (cf. [16] and Lemma 1.3 of [5]). In applications, one often needs to apply the exchange and roundoff principle in situations when  $F_1$  and  $F_2$  may not meet transversely or  $F_1 \cap F_2$  is not transverse to  $\mathcal{S}$ . There are two ways to overcome this in the PL case. It is possible to slightly perturb the metric on  $\mathcal{S}^{(2)}$  so that new PL least area surfaces  $F_1^*$  and  $F_2^*$  can be found which are very close to  $F_1$  and  $F_2$  and intersect transversely; with  $F_1^* \cap F_2^*$  transverse to  $\mathcal{S}$ . This will be done in the Appendix.

The other way is the Meeks-Yau trick. In the analytic case, the idea is to perturb  $F_1$  slightly to  $F_1^*$ , increasing area by  $\varepsilon$ , so that  $F_1^*$  meets  $F_2$  transversely. If there are any product regions between  $F_1^*$  and  $F_2$ , then at least  $2\varepsilon$  in area is saved by exchange and roundoff, a contradiction. We now show how to adapt this trick to the PL minimal case, replacing area by PL area, i.e. essentially length, since weight for normal surfaces is constant under small variations. We will look at a typical situation, as in Lemma 1.3 of [5].

**Lemma 2.** *Suppose  $F_1, F_2$  are two-sided embedded  $\pi_1$ -injective PL least area surfaces in a  $P^2$ -irreducible 3-manifold  $M$  and the inclusion of each is a homotopy equivalence. Then either they are disjoint or have the same image.*

*Proof.* If the surfaces intersect transversely and  $F_1 \cap F_2$  is transverse to  $\mathcal{S}$ , then Waldhausen [26] gives the existence of product regions and the exchange and roundoff trick produces a contradiction.

Suppose  $F_1$  and  $F_2$  have distinct images and either they intersect nontransversely or  $F_1 \cap F_2$  is not transverse to  $\mathcal{S}$ , as in §2.1 above. We choose some point  $x$  in  $F_1 \cap F_2 \cap \mathcal{S}^{(1)}$  and examine the local picture at  $x$ , as in Figures 11a and 11b. We can perturb  $F_1$  and  $F_1^*$  near  $x$  as in Figures 12a and 12b respectively. (There are two possible directions for the perturbation, with similar pictures and argument.) Now in both cases, by suitable perturbations



near all points of nontransversality, we can assume  $F_1^*$  is transverse to  $F_2$  and  $F_1^* \cap F_2$  is transverse to  $\mathcal{T}$ . Also the length of  $F_1^*$  is given by  $l(F_1^*) = l(F_1) + \varepsilon$ , where  $\varepsilon$  is arbitrarily small.

An exchange can now be performed between  $F_1^*$  and  $F_2$ , since there will be product regions by [26]. There are two possibilities in each of Figures 12a and 12b; since these are so similar we restrict attention to Figure 12a. We obtain new surfaces  $F'_1$  and  $F'_2$ , as in Figures 13a and 13b, as the possible results of

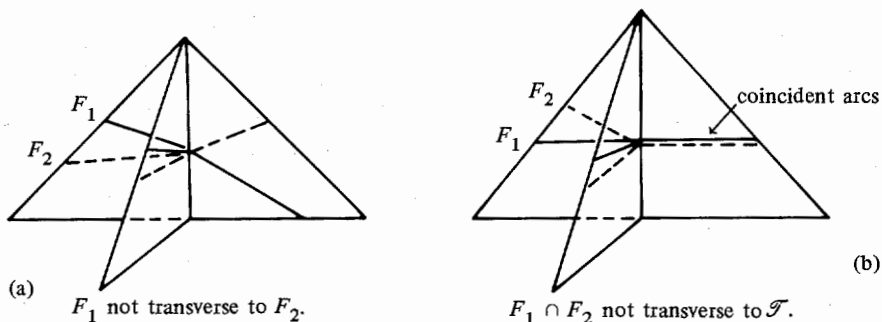


FIGURE 11

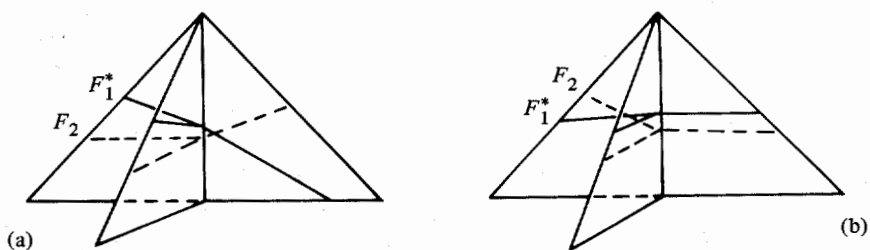


FIGURE 12

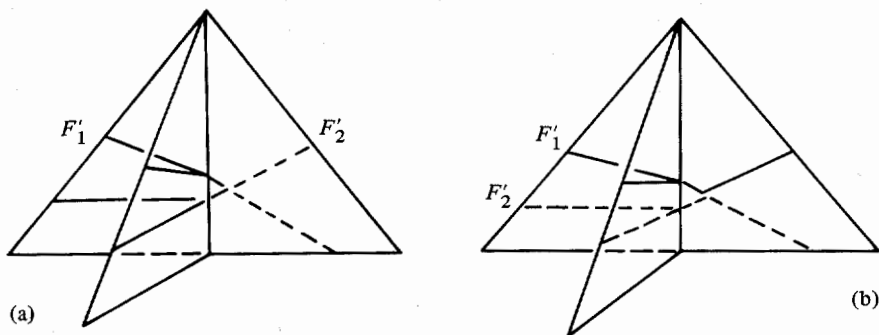


FIGURE 13

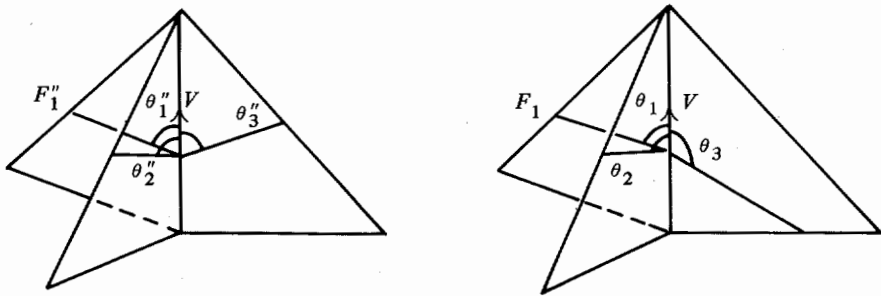


FIGURE 14

an exchange between  $F_1^*$  and  $F_2$ . In the case of Figure 13a,  $F_1'$  is not a normal surface so its weight can be reduced. This contradicts the assumption that  $F_1$  and  $F_2$  are PL least area, since all the surfaces are in the same homotopy class and must have the same weight.

If  $F_1'$  and  $F_2'$  intersect as in Figure 13b, then we can firstly smooth out the corners to obtain  $F_1''$  from  $F_1'$  as in Figure 14. We have drawn  $F_1$  in Figure 14 as well to emphasize that the angles  $\theta_1''$  and  $\theta_2''$  are approximately equal to  $\theta_1$  and  $\theta_2$  respectively, but  $\theta_3''$  is much smaller than  $\theta_3$ . The angles are measured relative to the unit vector  $V$  along the edge of  $\mathcal{T}^{(1)}$  containing  $x$ . Let  $f: F_1'' \rightarrow M$  be the map of  $F_1''$  and let  $y$  be chosen so that  $f(y) = x$ . Clearly the mean curvature  $H(y)$  of  $F_1''$  at  $y$  points in the direction  $V$  and also the length of  $H(y)$  can be bounded away from zero, independent of  $\epsilon$ . Hence we can reduce the length of  $F_1''$  by  $2\epsilon$  by sliding the surface along  $\mathcal{T}^{(1)}$  in the direction  $V$ . This gives a new surface  $F_1^+$  satisfying  $l(F_1^+) = l(F_1'') - 2\epsilon$ . Hence

$$\begin{aligned} l(F_1) + l(F_2) &= l(F_1^*) + l(F_2) - \epsilon = l(F_1') + l(F_2') - \epsilon \\ &> l(F_1'') + l(F_2') - \epsilon = l(F_1^+) + l(F_2') + \epsilon > l(F_1^+) + l(F_2'). \end{aligned}$$

Consequently one of  $F_1^+$  and  $F_2'$  must have smaller length than  $F_1$  or  $F_2$ , contradicting these latter surfaces being PL least area. This completes the proof.

To complete the existence of PL minimal and PL least area surfaces, we will need the following fundamental result of Kneser [13] and Haken [7] (see also [11], [9]).

**Theorem 3.** (a) Assume  $M$  is  $P^2$ -irreducible and let  $f: F \rightarrow M$  be a  $\pi_1$ -injective surface or an essential disk. Then there is a least weight normal surface  $g: F \rightarrow M$  homotopic to  $f$ . If  $f$  is embedded then there is a normal surface  $g: F \rightarrow M$  which is isotopic to  $f$  and has least weight in the isotopy class of  $f$ .

(b) Suppose  $M$  is any 3-manifold which contains essential 2-spheres. If  $\pi_2(M) \neq \{1\}$  then there exists a normal 2-sphere  $g: S^2 \rightarrow M$  which is noncontractible and  $g$  has least weight in  $\pi_2(M)$ . If  $\pi_2(M) = \{1\}$  then there is an essential embedded 2-sphere  $g: S^2 \rightarrow M$  which is least weight amongst all essential 2-spheres. (Such 2-spheres necessarily bound fake 3-balls but do not bound 3-balls in  $M$ .)

*Proof.* (a) We note that in the moves (cf. [9] or [7]) to attain normality, there are compressions of the surface inside 3-simplices. These can be achieved by homotopies or isotopies so long as  $M$  is  $P^2$ -irreducible. Amongst the normal surfaces homotopic or isotopic to  $f$ , we can choose one of least weight.

(b) Here the compressions will split a 2-sphere into two 2-spheres, at least one of which must be essential. If both 2-spheres are essential then if one is discarded, we do not remain in the same homotopy or isotopy class. However we stay in the class of essential 2-spheres. Again once normal essential 2-spheres have been constructed, one of least weight can be chosen.

### 3. Applications of PL minimal and PL least area surfaces

The following results from [5], [23], and [14] can now be proved using the properties outlined in §2, plus the original methods, where PL minimal (respectively PL least area) is substituted for minimal (respectively least area) throughout. Also the results discussed in [12] using least weight normal surfaces, namely [16]–[20] but *not* [15] can also be obtained in this way. We need an extra property of PL least area surfaces to establish the main application of [15].

Throughout this section, whenever we consider covering spaces it will always be assumed that the triangulation and the metric on the 2-skeleton of the cover is induced by the covering projection. This implies that the projection or lift of a PL minimal surface is PL minimal, under any covering, by the equivalence of mean curvature zero and PL minimality. We call an immersion  $f: F \rightarrow M$  *locally self-transverse* if given points  $y, y'$  with  $f(y) = f(y')$ , there are small disks about  $y, y'$  in  $F$  which are embedded by  $f$  and intersect transversely.

**Theorem 4** (cf. Theorems 3.1 and 7.2 of [5]). *Let  $M$  be a  $P^2$ -irreducible 3-manifold which covers a compact 3-manifold. Let  $f: F \rightarrow M$  be a  $\pi_1$ -injective map. Then there exists a PL least area map in the homotopy class of  $f$ .*

*Proof.* If  $M$  is compact, this follows immediately by Theorems 3(a), 1, and 2, and the observation that there are only finitely many ways of constructing a normal surface of fixed weight. (The number of disk types, as in Figure 1, is bounded and there are finitely many 3-simplices to choose from.) Hence one

can find a PL least area map from amongst the finitely many PL minimal surfaces of least weight in the homotopy class of  $f$ .

The general case follows as in [5]. If  $M$  covers a compact 3-manifold  $M'$  then by the compact version there is a PL least area map  $f'$  in the homotopy class of  $\pi f$ , where  $\pi$  is the covering projection. Then  $f'$  lifts to the desired PL least area map.

**Remark 4.** The assumption that  $M$  covers a compact 3-manifold is needed here but not in Theorem 3. The reason is that weight of normal surfaces is discrete, so least weight normal surfaces exist trivially. However if there are infinitely many in a homotopy class, for noncompact  $M$ , then there may not be one of minimal length.

**Theorem 5** (cf. Theorem 3.4 of [5]). *Let  $M$  be a  $P^2$ -irreducible 3-manifold which covers a compact 3-manifold. Let  $f: F \rightarrow M$  be a two-sided  $\pi_1$ -injective map. Let  $M_1$  be any covering of  $M$  and let  $F_1$  be a finite-sheeted covering of  $F$  such that  $f_1: F_1 \rightarrow M_1$  is a lift of  $f$ . Then  $f_1$  is least area if and only if  $f$  is least area.*

*Proof.* See [5].

**Theorem 6** (cf. Theorem 5.1 of [5]). *Let  $M$  be a compact  $P^2$ -irreducible 3-manifold and let  $f: F \rightarrow M$  be a PL least area  $\pi_1$ -injective map which is homotopic to a two-sided embedding  $g$ . Then either:*

- (i)  $f$  is an embedding or
- (ii)  $f$  double covers a one-sided surface  $K$  properly embedded in  $M$  and  $g(F)$  disconnects  $M$  into two compact regions, one of which is a twisted  $I$ -bundle over a surface isotopic to  $K$ .

*Proof.* See [5].

**Theorem 7** (cf. Theorem 6.2 of [5]). *Let  $M$  be a compact  $P^2$ -irreducible 3-manifold and let  $f_i: F_i \rightarrow M$  be two-sided PL least area  $\pi_1$ -injective maps,  $i = 1, 2$ . If  $f_1$  and  $f_2$  can be homotoped to have disjoint images, then  $f_1(F_1)$  and  $f_2(F_2)$  are disjoint or coincide. In the latter case  $f_1$  and  $f_2$  are coverings of an embedded surface in  $M$ .*

We now summarize the approach in §6 of [5] to the problem of counting intersections and self-intersections of PL least area surfaces. This will be important for our final results.

Assume for simplicity that our PL least area maps do not factor through nontrivial coverings of surfaces. Let  $f_1: F_1 \rightarrow M$  be a  $\pi_1$ -injective PL least area surface and let  $M_1$  denote the covering of  $M$  with group  $\pi_1(F_1)$ , one of the conjugacy class of subgroups of  $\pi_1(M)$  determined by  $f_1$ . Then there is a lift of  $F_1$  to  $M_1$  which is an embedding, by Theorems 5 and 6. We denote this lift by  $F_1$  again. Let us call the image of a lift of  $f_1(F_1)$  to  $M_1$  a *component* of the pre-image of  $f_1(F_1)$ . Then we count the *self-intersections* of  $f_1$  as the

number  $D(f_1)$  of the components of the pre-image of  $f_1(F_1)$  which meet  $F_1$ . Similarly the number of intersections between two  $\pi_1$ -injective PL least area maps  $f_i: F_i \rightarrow M, i = 1, 2$ , is given by the number  $D(f_1, f_2)$  of the components of the pre-image of  $f_2(F_2)$  in  $M_1$  which meet  $F_1$ .

**Theorem 8** (cf. Theorem 6.3 of [5]). *Let  $M$  be a compact  $P^2$ -irreducible 3-manifold. Let  $f'_i: F_i \rightarrow M$  be two-sided  $\pi_1$ -injective maps and let  $f_i$  be a PL least area map homotopic to  $f'_i$  for  $i = 1, 2$ . Then*

- (i)  $D(f_1, f_2) \leq D(f'_1, f'_2)$ ,
- (ii)  $D(f_1) \leq D(f'_1)$ .

We now turn to results of Scott [23] and Meeks-Scott [14].

**Theorem 9** (cf. Theorem 3.1 of [23]). *Let  $N$  be a closed irreducible Seifert fiber space with infinite fundamental group. Let  $M$  be a closed orientable, irreducible 3-manifold such that  $\pi_1(M)$  is isomorphic to  $\pi_1(N)$ . Then  $M$  and  $N$  are homeomorphic.*

**Remark 5.** The argument in [23] is a beautiful analysis of the intersection of least area tori in a suitable common covering space of  $M$  and  $N$ . The same method works using PL least area tori.

**Corollary 2** [23]. *If  $M$  is a closed orientable irreducible 3-manifold with infinite fundamental group which is finitely covered by a Seifert fiber space, then  $M$  must be a Seifert fiber space.*

**Theorem 10** [14, Theorem 2.1]. *Let  $M$  be a closed 3-manifold with a geometric structure of type  $H^2 \times R, \overline{Sl_2R}, Nil, E^3$  or Sol. Then any finite group action on  $M$  preserves the geometric structure.*

**Remark 6.** See [25] or [24] for discussions of geometric structures. As in Remark 5, the only simplification occurring is that existence of least area tori is replaced by that of PL least area tori.

Our next theorem is a “local” analysis of the singularities of PL minimal surfaces, namely as in [5], such surfaces have least possible intersections and self-intersections in normal homotopy classes. Note the absence of any assumptions guaranteeing nontrivial homotopy properties of the surfaces, like  $\pi_1$ -injectivity of their maps. Also it is not necessary to suppose that the surfaces are two-sided.

We begin with the following lemma that if a certain neighborhood of  $\mathcal{S}^{(1)}$  is removed from  $M$ , then any normal surface becomes  $\pi_1$ -injective. Let  $N(\mathcal{S}^{(1)})$  be the small special neighborhood of the 1-skeleton of  $\mathcal{S}$  in  $M$  given by choosing an open 3-ball containing each open edge in  $\mathcal{S}^{(1)}$  and adding in all the vertices in  $\mathcal{S}^{(0)}$ , as in Figure 15. Let  $M_0 = M - N(\mathcal{S}^{(1)})$ . Clearly  $\text{int } M_0$  is an open handlebody and  $\partial M_0$  is a collection of open annuli, one for each edge in  $\mathcal{S}^{(1)}$ .

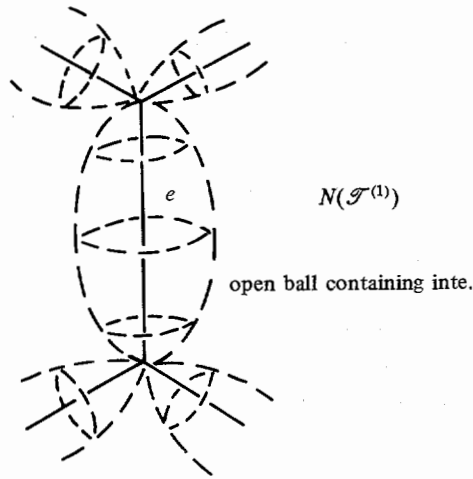


FIGURE 15

Let  $f: F \rightarrow M$  be any normal surface. We can suppose without loss of generality that  $f(F)$  meets  $N(\mathcal{T}^{(1)})$  in a finite set of small embedded disks, since  $f$  is transverse to  $\mathcal{T}^{(1)}$ . Let  $F_0 = f^{-1}(M_0)$ . then we can also assume without loss of generality that for any  $g$  in the normal homotopy class  $\mathcal{N}(f)$ ,  $g^{-1}(M_0) = f^{-1}(M_0) = F_0$ . Let  $f_0 = f|_{F_0}$ ,  $g_0 = g|_{F_0}$ , and  $\sigma_0 = \sigma \cap M_0$ , where  $\sigma$  is any simplex in  $\mathcal{T}$ . Also let  $\mathcal{T}_0^{(2)}$  denote  $\mathcal{T}^{(2)} \cap M_0$ .

**Lemma 3.**  $f_0: F_0 \rightarrow M_0$  is  $\pi_1$ -injective.

*Proof.* Suppose this is not the case. Then there is a map of the disk  $\phi: D \rightarrow M_0$  satisfying one of the following two possibilities. The first is that there is a noncontractible loop in  $F_0$  whose image under  $f_0$  is  $\phi(\partial D)$ . The second is that there are arcs  $\lambda, \mu, \nu$  with  $\partial D = \lambda \cup \mu, \partial\lambda = \partial\mu, \nu$  in  $F_0$  and  $\phi(\lambda) \subset \partial M_0, \phi(\mu) = f_0(\nu)$ , where  $\nu$  is nontrivial in  $\pi_2(F, \partial F)$ . These cases will be referred to by saying that  $f_0$  is compressible or  $\partial$ -compressible respectively. We assume without loss of generality that  $\phi$  is transverse to  $\mathcal{T}_0^{(2)}$  and obtain a contradiction by induction on the number of components of  $\phi^{-1}(\mathcal{T}_0^{(2)})$ .

To start the induction, we show  $\phi^{-1}(\mathcal{T}_0^{(2)})$  cannot be empty. Since  $f$  is normal,  $F$  meets each 3-simplex  $\tau$  of  $\mathcal{T}$  in the seven disk types, as in Figure 1. So if  $\tau_0 = \tau \cap M_0$ , then  $F_0$  intersects  $\tau_0$  in properly embedded disks, so cannot be compressible or  $\partial$ -compressible in  $\tau_0$ .

It now suffices to reduce the number of arcs and loops in  $\phi^{-1}(\mathcal{T}_0^{(2)})$ . Note these curves are all disjoint and embedded. Any loops can be removed by elementary surgeries of  $\phi$ . So we will concentrate on an arc  $\gamma$  of  $\phi^{-1}(\sigma_0)$ , where  $\sigma$  is a 2-simplex in  $\mathcal{T}^{(2)}$ . There are three pictures for  $\phi(\gamma)$  in  $\sigma_0$ , as in Figures 16a, b, c. Also there are two cases to consider for the arc  $\gamma$  in  $D$  as in Figures

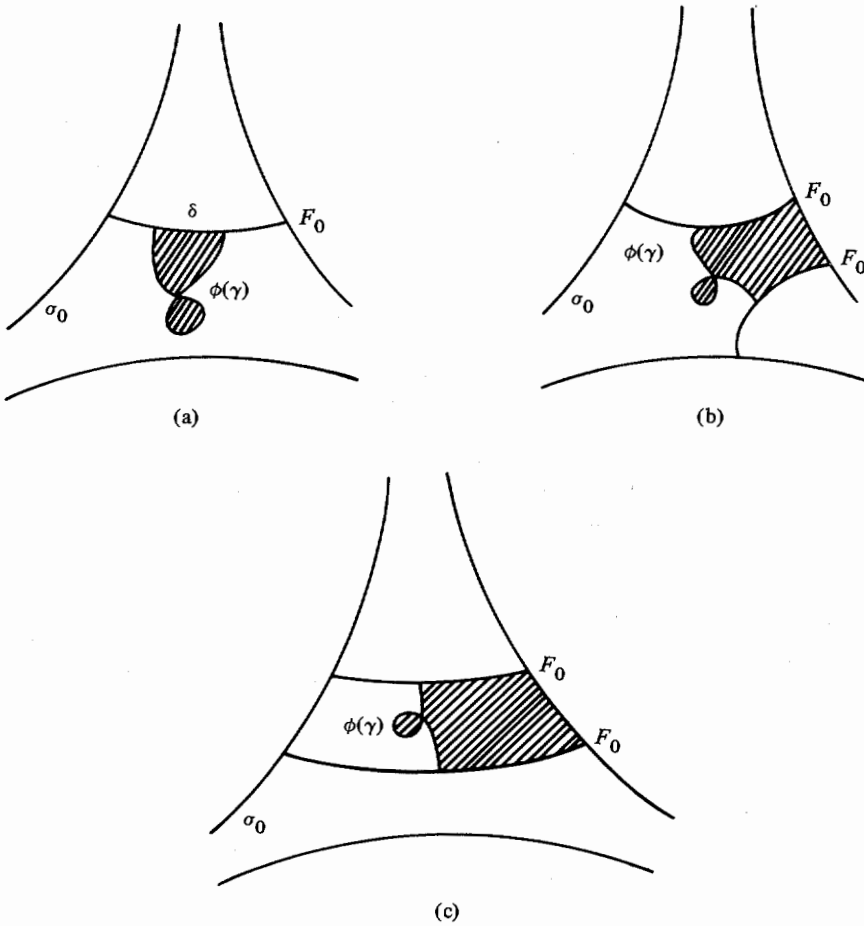


FIGURE 16

17i, ii, depending on whether  $\phi$  is a compression or a  $\partial$ -compression. We will let  $D_0, D_1$  denote the closures of the components of  $D - \gamma$ .

In the situation of Figure 16a, there is an obvious shaded homotopy in  $\sigma_0$  between  $\phi(\gamma)$  and  $\delta$ , an arc of  $F_0$ . If this is in combination with Figure 17i, we modify  $\phi$  by replacing either  $\phi|D_0$  or  $\phi|D_1$  by this homotopy. One of the two new maps of the disk, which we denote by  $\phi^*$ , will have the property that  $\phi^*|\partial D$  is the image of a noncontractible loop in  $F_0$ . By a small perturbation off of  $\phi_0$ , we obtain that  $\phi^{*-1}(\mathcal{T}_0^{(2)})$  has fewer components than  $\phi^{-1}(\mathcal{T}_0^{(2)})$ , as desired. In the case of Figure 17ii, we replace  $\phi|D_0$  by the shaded homotopy in  $\sigma_0$  and the argument is similar. Note that we retain the property that  $\phi$  is a

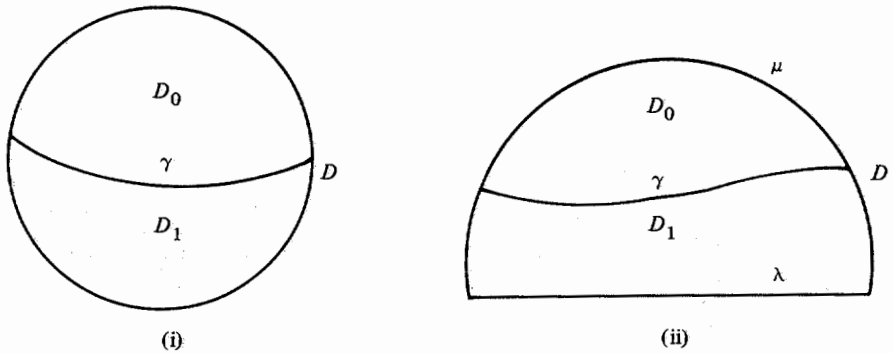


FIGURE 17

$\partial$ -compression here; in particular the new arc  $\nu$  in  $F_0$  with  $f_0(\nu) \subset \phi^*(D)$  cannot be homotopic in  $F_0$  relative to  $\partial\nu$  into  $\partial F_0$ , since the ends of  $\nu$  are on two *different* curves of  $\partial F_0$ . Because  $F$  is normal, every point of  $F \cap \mathcal{T}^{(1)}$ , with *multiplicity*  $n$ , yields  $n$  disks of  $F \cap N(\mathcal{T}^{(1)})$  and  $n$  boundary curves of  $F_0$ .

Next we consider Figures 16b and 16c which are alike. (Notice that although the arcs of  $F_0$  in these pictures have been drawn as disjoint, they can cross since  $F_0$  may be singular. This does not affect the method though.) In the situation of Figure 17i, either  $\phi|D_0$  or  $\phi|D_1$  can be substituted for by the shaded homotopy in Figures 16b or 16c. This converts a compression into a  $\partial$ -compression. Finally in the case of Figure 17ii, we replace  $\phi|D_1$  by the shaded homotopy in Figures 16b or 16c. This preserves the disk as a  $\partial$ -compression. In all the above possibilities, the number of components in  $\phi^{-1}(\mathcal{T}_0^{(2)})$  is decreased after a small push off  $\sigma_0$  and the proof is complete.

Let us assume, for convenience, that all PL minimal and normal surfaces  $f: F \rightarrow M$  have the property that  $f$  is in general position relative to  $\mathcal{T}^{(1)}$ . So as well as  $F$  meets  $\mathcal{T}^{(1)}$  transversely, every point of  $F \cap \mathcal{T}^{(1)}$  lies in a single sheet of  $F$ , i.e. occurs with multiplicity one. The *self-intersections* of  $f$  will be counted by  $S(f) = D(f_0)$  and the *intersections* of normal surfaces  $f: F \rightarrow M$  and  $f': F' \rightarrow M$  by  $S(f, f') = D(f_0, f'_0)$ . Note that as  $f, f'$  are in general position relative to  $\mathcal{T}^{(1)}$ ,  $f$  and  $f_0$  (respectively  $f'$  and  $f'_0$ ) have the same double curves, i.e. no double curve passes through  $\mathcal{T}^{(1)}$  or  $N(\mathcal{T}^{(1)})$ , if this neighborhood is chosen small enough. In the Appendix we outline a simple method to ensure that any finite set of PL minimal surfaces are in general position relative to each other and to  $\mathcal{T}$ .

**Theorem 11.** *Let  $f: F \rightarrow M$  and  $f': F' \rightarrow M$  be normal surfaces in a compact 3-manifold. If  $g$  and  $g'$  are the PL minimal surfaces in the normal homotopy classes of  $f$  and  $f'$  respectively, then*

$$S(g) \leq S(f) \quad \text{and} \quad S(g, g') \leq S(f, f').$$



**Corollary 3** (cf. Theorem 6). *If  $f: F \rightarrow M$  is an embedded normal surface in a compact 3-manifold, then the PL minimal surface  $g$  in  $\mathcal{N}(f)$  is either an embedding or a double cover of an embedded surface. In the latter case  $f(F)$  bounds a twisted I-bundle over a nonorientable surface isotopic to  $g(F)$ .*

**Corollary 4** (cf. Theorem 7). *Suppose  $f: F \rightarrow M$  and  $f': F' \rightarrow M$  are normal surfaces with disjoint images in a compact 3-manifold. Then the PL minimal surfaces  $g$  in  $\mathcal{N}(f)$  and  $g'$  in  $\mathcal{N}(f')$  have images which are either disjoint or coincide. If  $g$  and  $g'$  have the same image, they are coverings of embeddings.*

**Remark 7.** It is easy to see from the following proof of Theorem 11 that the corollaries do not depend on the simplifying assumption that the surfaces are in general position relative to  $\mathcal{T}^{(1)}$ .

*Proof.* We assume first that  $f$  and  $f'$  are two-sided surfaces. By Lemma 3, we know that since  $f: F \rightarrow M$  is normal, it follows that  $f_0: F_0 \rightarrow M_0$  is  $\pi_1$ -injective. The strategy is to apply Theorem 8. To do this, a suitable Riemannian metric must be introduced on the 2-skeleton of  $M_0$ . Notice that although  $M_0$  is not compact, its “triangulation” has only finitely many “simplices” so the previous arguments apply. We need to choose the metric so that the given PL minimal surfaces  $g, g'$  yield surfaces  $g_0, g'_0$  which are PL minimal for  $M_0$ , after small perturbations.

If  $\tau$  is any 3-simplex of  $\mathcal{T}$ , then  $\tau_0$  is obtained by truncating  $\tau$  along its edges. These truncated 3-simplices give a cell decomposition of  $M_0$  which we will denote by  $\mathcal{T}_t$  and call the triangulation of  $M_0$ . Notice that  $\tau_0$  has four faces which are 2-simplices and 6 faces which are 2-gons, all of which have their vertices removed. Assume also that the special neighborhood  $N(\mathcal{T}^{(1)})$  which is to be removed from  $M$  has arbitrarily small size  $\epsilon$ , i.e.  $N(\mathcal{T}^{(1)}) \cap \mathcal{T}^{(2)}$  has points at distance up to  $\epsilon$  from  $\mathcal{T}^{(1)}$  in the metric on  $\mathcal{T}^{(2)}$ .

A 2-simplex  $\sigma$  of  $\mathcal{T}$  becomes a 2-simplex  $\sigma_0$  of  $\mathcal{T}_t$ , when intersected with  $M_0$ . We will identify  $\sigma_0$  with the same ideal hyperbolic triangle as  $\sigma$  (see Figure 2). This can be done so that if  $\lambda$  is any arc of  $g(F) \cap \sigma_0$  or of  $g'(F') \cap \sigma_0$ , then every point of  $\lambda$  is within  $m\epsilon$  of  $\mu$ , the unique hyperbolic geodesic in  $\sigma_0$  with the same endpoints as  $\lambda$ , where  $m$  is some constant (see Figure 18). In addition, every pair of such arcs  $\lambda$  either coincide or meet transversely in single points, as they are geodesics in the original metric on  $\sigma$ . Consequently if  $\epsilon$  is small enough, the intersection picture is identical for the new family of geodesics  $\mu$ , as for the arcs  $\lambda$  of  $g(F) \cap \mathcal{T}_0^{(2)}$  (see Figure 18).

Next we identify all the 2-gons in  $\mathcal{T}_t^{(2)}$  with ideal convex hyperbolic 2-gons, satisfying the distance from any point on one boundary arc to the other boundary arc is at most  $\epsilon$ , as in Figure 19. Suppose  $e$  is any edge of  $\mathcal{T}_t^{(1)}$  and  $x$  is a point on  $e$  at which say  $g_0(F_0)$  crosses  $e$ . (The case for  $g'_0(F'_0)$  is

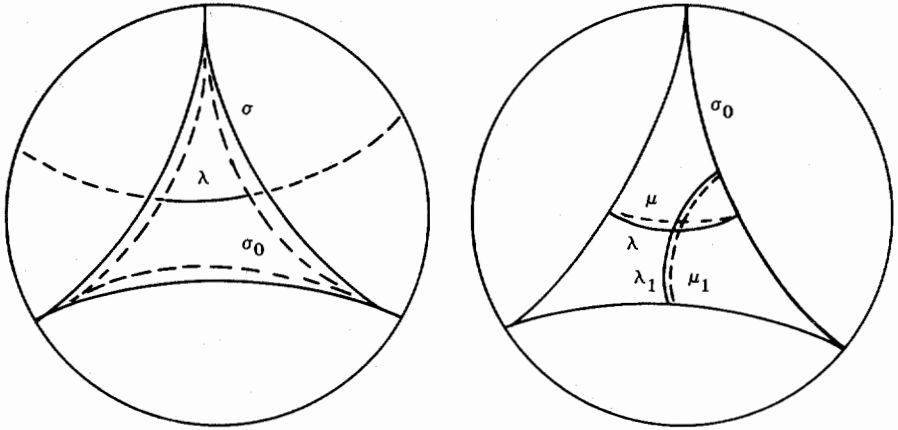
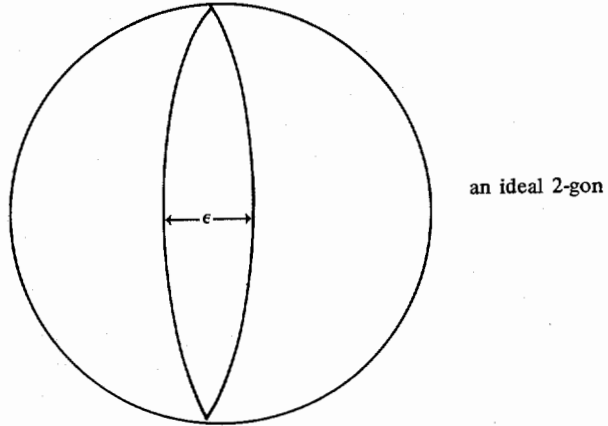


FIGURE 18



an ideal 2-gon

FIGURE 19

identical.) Then there are clearly exactly three arcs  $\lambda_1, \lambda_2, \lambda_3$  of  $g_0(F_0) \cap \mathcal{T}_i^{(2)}$  with endpoints at  $x$ , by our assumptions that  $g$  is in general position relative to  $\mathcal{T}^{(1)}$  and  $N(\mathcal{T}^{(1)})$  is small (see Figure 20).

**Claim.** The metric on  $\mathcal{T}_i^{(2)}$  can be chosen so that if  $\mu_1, \mu_2, \mu_3$  are the geodesic arcs with the same endpoints as  $\lambda_1, \lambda_2, \lambda_3$ , then the angles  $\theta_i$ ,  $1 \leq i \leq 3$ , between  $\mu_i$  and a unit vector  $V$  along  $e$  satisfy  $\cos \theta_1 + \cos \theta_2 + \cos \theta_3 = 0$ .

*Reason.* Let  $C$  be the loop of  $g_0(\partial F_0)$  through  $x$ , let  $D$  be the disk of  $g_0(F_0) \cap N(\mathcal{T}^{(1)})$  bounded by  $C$  and let  $z$  be the unique point where  $D$  crosses  $\mathcal{T}^{(1)}$ . Since  $g$  is PL minimal, the sum of the cosines of the angles of the edges of  $g(F) \cap \mathcal{T}^{(2)}$  which have endpoints at  $z$ , with the edge of  $\mathcal{T}^{(1)}$

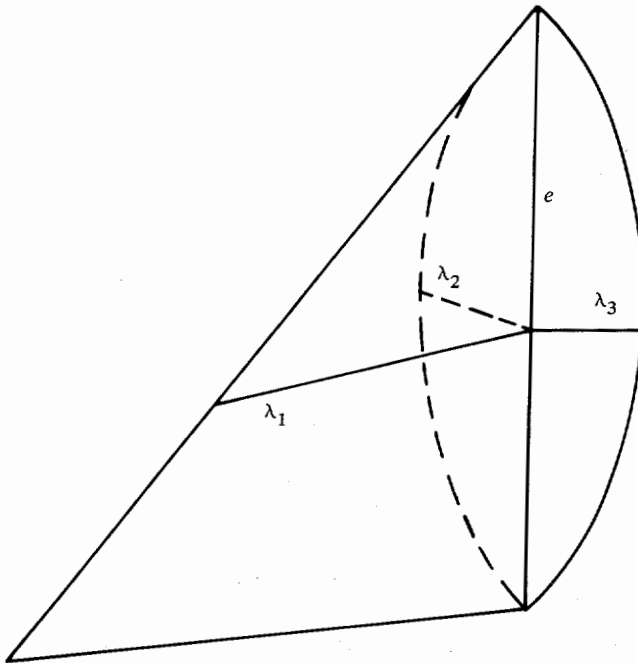


FIGURE 20

containing  $z$ , is zero. By continuity, the corresponding sum for the edges of  $g_0(F_0) \cap \mathcal{T}_0^{(2)}$  which have endpoints on  $C$ , with the appropriate edges of  $\mathcal{T}_t^{(1)}$ , is nearly zero; in fact, the absolute value of the sum can be bounded by  $n\epsilon$ , where  $n$  is a constant independent of  $C$ .

Since our 2-gons are chosen to be very thin, if an arc  $\lambda$  of  $g_0(F_0)$  crossing any 2-gon is chosen "obliquely," then it is still relatively short but the angles  $\theta$  and  $\theta'$  of the corresponding geodesic  $\mu$  with the boundary edges become close to 0 and  $\pi$  respectively (see Figure 21). Also by the Gauss-Bonnet Theorem, we see that  $\pi - \theta + \theta'$  is close to the area of one of the complementary regions of  $\mu$  in the 2-gon, so  $\theta + \theta'$  is approximately equal to  $\pi$ . Hence  $|\theta + \theta' - \pi| < p\epsilon$  for some constant  $p$ . Finally by proceeding along the loop  $C$ , we can easily choose the geodesic arcs crossing the 2-gons with appropriate angles so as to achieve  $\cos \theta_1 + \cos \theta_2 + \cos \theta_3 = 0$  at all except the last point  $x$ . At this point, we must have  $|\cos \theta_1 + \cos \theta_2 + \cos \theta_3|$  is of the order of  $\epsilon$ . But then enlarging (shrinking) one of the pair of 2-gons at  $x$  enables us to increase (decrease) one of these angles  $\theta_i$ ,  $1 \leq i \leq 3$ , keeping all other angles fixed. So we can always achieve that the surface  $h_0: F_0 \rightarrow M_0$  (or  $h'_0: F'_0 \rightarrow M_0$ ), with  $h_0(F_0) \cap \mathcal{T}_t^{(2)}$  (or  $h'_0(F'_0) \cap \mathcal{T}_t^{(2)}$ ) given by the geodesic arcs  $\mu_i$  as above, is a

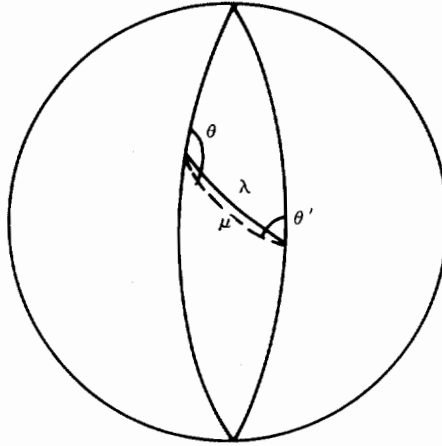


FIGURE 21

PL minimal surface  $and h_0$  has the same self-intersection picture as  $g_0$  ( $h_0, h'_0$  have an identical intersection pattern to  $g_0, g'_0$ ).

By Theorem 8, PL least area surfaces  $k_0, k'_0$  in the homotopy classes of  $f_0, f'_0$  minimize intersections and self-intersections. It remains to identify  $k_0$  and  $k'_0$ . Certainly PL least area surfaces must have least weight. The weight of  $f_0$  is clearly the number of points where  $f_0(\partial F_0)$  meets edges of  $\mathcal{T}_t^{(1)}$ . It is easy to see this number cannot be decreased by any homotopy of  $f_0$ . Hence the boundary curves of  $k_0(F_0)$  must be *normally* homotopic to those of  $f_0(F_0)$ , and similarly for  $k'_0$  and  $f'_0$ . But  $k_0(F_0)$  is complete determined by  $k_0(\partial F_0)$ , since all the arcs of  $k_0(F_0) \cap \mathcal{T}_0^{(2)}$  are hyperbolic geodesics and so are uniquely specified by their endpoints. So we see that  $k_0$  is normally homotopic to  $f_0$ , and likewise for  $k'_0$  and  $f'_0$ . But now by Theorem 2, there is a unique PL minimal surface in  $\mathcal{N}(f_0)$  and so  $k_0 = h_0$ . Hence we conclude that  $D(g_0) = D(h_0) = D(k_0)$  and  $D(g_0, g'_0) = D(h_0, h'_0) = D(k_0, k'_0)$ . By Theorem 8,  $D(k_0) \leq D(f_0)$  and  $D(k_0, k'_0) \leq D(f_0, f'_0)$ . Therefore two-sided PL minimal surfaces have smallest numbers of intersections and self-intersections in normal homotopy classes, as claimed.

To complete the theorem, we must consider the one-sided case. Assume that  $f: F \rightarrow M$  and  $f': F' \rightarrow M$  are one-sided surfaces. (The case when, for example,  $f$  is one-sided and  $f'$  is two-sided follows by the same method.) Let  $\pi: \tilde{F} \rightarrow F$  be the double covering such that if  $\tilde{f} = f \cdot \pi$ , then  $\tilde{f}: \tilde{F} \rightarrow M$  has image homotopic to the boundary of a regular neighborhood of  $f(F)$  in  $M$ . Clearly  $\tilde{f}$  is also a normal surface. Finally let  $g: F \rightarrow M$  and  $\tilde{g}: \tilde{F} \rightarrow M$  be the

PL minimal surfaces in the normal homotopy classes of  $f$  and  $\tilde{f}$  respectively. Also  $\tilde{f}'$ ,  $g'$ , and  $\tilde{g}'$  are defined analogously, using  $f'$  instead of  $f$ .

Since the surface  $g \cdot \pi: \tilde{F} \rightarrow M$  is obviously PL minimal, its mean curvature is identically zero and by Theorem 2 we conclude that  $\tilde{g} = g \cdot \pi$ . But then we can apply the previous argument in the two-sided case, obtaining  $D(\tilde{g}, \tilde{g}') \leq D(\tilde{f}, \tilde{f}')$  and  $D(\tilde{g}) \leq D(\tilde{f})$ . Combining these, the outcome is that  $D(g, g') \leq D(f, f')$  and  $D(g) \leq D(f)$ . (Note that if  $M_2$  (respectively  $M_1$ ) is the cover of  $M$  corresponding to  $\pi_1(\tilde{F})$  (respectively  $M_1$ ) is the cover of  $M$  corresponding to  $\pi_1(F)$ ), then  $M_2$  is the double cover of  $M_1$ .) This finishes the proof.

**Remark 8.** We sketch an alternative approach for proving Corollary 3, via the techniques of [12], rather than [5]. This is a considerably more elementary method. Note that Corollary 3 is needed to prove the next result, Theorem 12.

As above, let us assume that  $f$  is an embedded two-sided normal surface. We wish to show the PL minimal surface  $g$  normally homotopic to  $f$  is either an embedding or a double cover of an embedded nonorientable surface.

Since  $f$  is embedded, in any 3-simplex of  $\mathcal{T}^{(3)}$ ,  $f$  can have at most one quadrilateral type (see Figure 1). But then if  $g$  is not a cover of an embedding, the double curves of  $g$  will all be regular (see [12]). In fact irregular double curves occur precisely where two different quadrilateral types cross. Now a regular exchange (see [12] can be performed along all the double curves of  $g$  to give an embedded normal surface  $h$ . As  $h$  is embedded and has exactly the same disk types as both  $g$  and  $f$ , it follows readily that  $h$  is normally homotopic to  $f$ . But  $h$  has corners and has the same PL area as  $g$ . Smoothing out the corners gives a surface in  $\mathcal{N}(f)$  with smaller PL area than  $g$ , a contradiction.

The case when  $f$  is one-sided follows exactly as before from the two-sided case.

We can now give another proof (see also [12]) of the main application of [15].

**Theorem 12.** *Let  $M$  be a  $P^2$ -irreducible 3-manifold. Then any covering of  $M$  is  $P^2$ -irreducible.*

*Proof.* Let  $\tilde{M}$  be any covering of  $M$  and assume  $\tilde{M}$  is not  $P^2$ -irreducible. By the sphere theorem (cf. for example [9]), since  $M$  is  $P^2$ -irreducible,  $\pi_2(M) = \{1\}$ . Therefore,  $\pi_2(\tilde{M}) = \{1\}$  also, and since  $\tilde{M}$  is not  $P^2$ -irreducible, there must be an embedded 2-sphere  $S$  which is essential in  $\tilde{M}$ , i.e. bounds a fake 3-ball  $R$ . Let  $\pi: \tilde{M} \rightarrow M$  be the covering projection and let  $N$  be the compact submanifold of  $M$  obtained by taking a regular neighborhood of  $\pi(R)$  and adding 3-balls to any boundary 2-spheres of this neighborhood.

Then  $M$  is again  $P^2$ -irreducible and we can reduce this problem to the study of the restriction of  $\pi$  to the component  $\tilde{N}$  of  $\pi^{-1}(N)$  which contains  $R$ .

As usual we triangulate  $N$  and put a Riemannian metric on the 2-skeleton of the triangulation. The triangulation and metric can then be lifted to  $\tilde{N}$ . Amongst the embedded essential 2-spheres in  $N$ , one of least PL area can be found. In fact by Theorem 3(b), there are such 2-spheres of smallest weight. Projecting to  $N$ , we see there are finitely many possibilities and amongst these we can choose one of least length. This lifts to the desired 2-sphere in  $\tilde{N}$ .

Finally by the usual exchange argument (see §2.2 and Lemma 3 of [15]), it is straightforward to show that if  $\eta: \tilde{N} \rightarrow \tilde{N}$  is any covering transformation and  $S$  is a PL least area essential 2-sphere, then  $S \cap \eta(S) = \emptyset$  or  $S = \eta(S)$ . Hence  $S$  projects either one-to-one or two-to-one to  $N$ . In the first case,  $\pi(S)$  is a two-sided projective plane in  $N$ , contradicting  $N$  is  $P^2$ -irreducible. This establishes the theorem.

### Appendix

We indicate a technique for perturbing the Riemannian metric on  $\mathcal{F}^{(2)}$  so that any finite collection of PL minimal surfaces become transverse to each other and to  $\mathcal{F}$ . The idea is that all the basic properties in §2 depend only on the metric having negative Gaussian curvature. So a small change from constant curvature  $-1$  is allowable. Also the PL minimal condition is equivalent to all the arcs in  $\mathcal{F}^{(2)}$  being geodesics, plus a restriction on the angles between these arcs and  $\mathcal{F}^{(1)}$  at their ends. We will vary these arcs so that the angles remain unchanged.

Suppose that  $\alpha$  is a geodesic arc in a hyperbolic triangle  $\sigma$  in  $\mathcal{F}^{(2)}$  and  $\alpha$  has endpoints  $x_1, x_2$  with angles  $\theta_1, \theta_2$  respectively between  $\alpha$  and the arcs of  $\partial\sigma$  (see Figure 22). We put two small “bumps” in  $\sigma$  at points  $z_1, z_2$  close to  $\alpha$  by locally multiplying the metric at  $z_i, i = 1, 2$ , by a function which has value  $1 + \varepsilon$  at  $z_i$  and 1 outside a small disk centered at  $z_i$ . We also can assume this function has very small gradient everywhere. Then there is a new geodesic arc  $\alpha'$  with endpoints  $x'_1$  and  $x_2$ , where  $x'_1$  is very close to  $x_1$ . Also  $\alpha'$  has the *same* angles  $\theta_1$  and  $\theta_2$  with the arcs of  $\partial\sigma$  as  $\alpha$ , for suitable choice of  $z_1, z_2$  and the bump functions.

Using this device, we can alter the position of any PL minimal surface by a small amount and achieve the desired transversality properties for a finite collection of such surfaces.

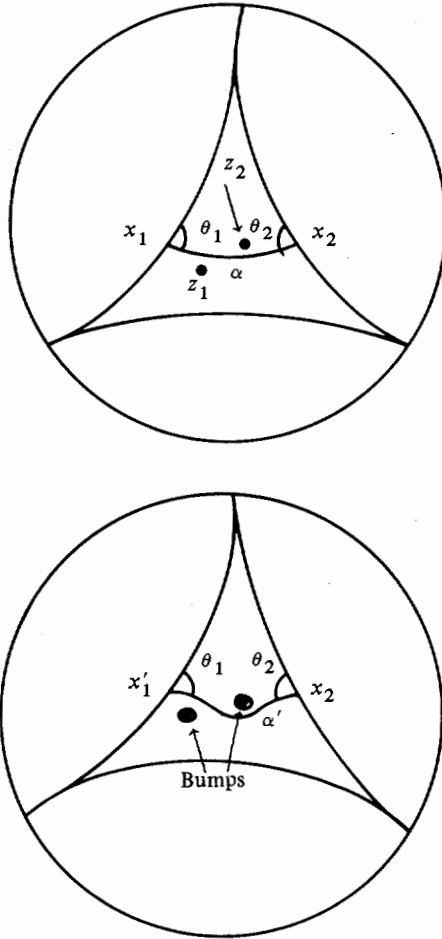


FIGURE 22

## References

- [1] L. Bers, *Isolated singularities of minimal surfaces*, Ann. of Math. (2) **53** (1951) 364–386.
- [2] J. Cheeger & D. Ebin, *Comparison theorems in Riemannian geometry*, North-Holland, Amsterdam, 1975.
- [3] M. Dunwoody, *An equivariant sphere theorem*, preprint.
- [4] A. Edmonds, *A topological proof of the equivariant Dehn's lemma*, Trans. Amer. Math. Soc. **297** (1986) 605–615.
- [5] M. Freedman, J. Hass & P. Scott, *Least area incompressible surfaces in 3-manifolds*, Invent. Math. **71** (1983) 609–642.
- [6] R. Gulliver & P. Scott, *Least area surfaces can have excess triple points*, preprint.
- [7] W. Haken, *Theorie der Normal Flächen*, Acta. Math. **105** (1961) 245–375.
- [8] J. Hass & P. Scott, *The existence of least area surfaces in 3-manifolds*, preprint.

- [9] J. Hempel, *3-manifolds*, Princeton University Press, Princeton, NJ, 1976.
- [10] D. Hilbert, *Über das Dirichletsche Prinzip*, Jber. Deutsch. Math.-Verein. **8** (1900) 184–188.
- [11] W. Jaco, *Normal surfaces and the projective solution space*, Enseignement Math., to appear.
- [12] W. Jaco & J. H. Rubinstein, *PL equivariant surgery and invariant decompositions of 3-manifolds*, preprint.
- [13] H. Kneser, *Geschlossene Flächen in dreidimensionalen Mannigfaltigkeiten*, Jber. Deutsch. Math.-Verein. **38** (1929) 248–260.
- [14] W. Meeks & P. Scott, *Finite group actions on 3-manifolds*, preprint.
- [15] W. Meeks, L. Simon & S.-T. Yau, *Embedded minimal surfaces, exotic spheres and manifolds with positive Ricci curvature*, Ann. of Math. (2) **116** (1982) 621–659.
- [16] W. Meeks & S.-T. Yau, *Topology of three-dimensional manifolds and the embedding problems in minimal surface theory*, Ann. of Math. (2) **112** (1980) 441–485.
- [17] ———, *The equivariant Dehn's Lemma and Loop Theorem*, Comment. Math. Helv. **56** (1981) 225–239.
- [18] ———, *The classical plateau problem and the topology of three dimensional manifolds*, Topology **21** (1982) 409–442.
- [19] ———, *The existence of embedded minimal surfaces and the problem of uniqueness*, Math. Z. **179** (1982) 151–168.
- [20] ———, *The equivariant Loop Theorem for three-dimensional manifolds and a review of the existence theorems for minimal surfaces*, The Smith Conjecture (J. Morgan and H. Bass, eds.), Academic Press, Orlando, 1984, 153–163.
- [21] J. Milnor, *Morse Theory*, Princeton University Press, Princeton, NJ, 1963.
- [22] R. Schoen & S.-T. Yau, *Existence of incompressible minimal surfaces and the topology of 3-dimensional manifolds with non-negative scalar curvature*, Ann. of Math. (2) **110** (1979) 127–142.
- [23] P. Scott, *There are no fake Seifert fibre spaces with infinite  $\pi_1$* , Ann. of Math. (2) **117** (1983) 35–70.
- [24] ———, *The geometries of 3-manifolds*, Bull. London Math. Soc. **15** (1983) 40–487.
- [25] W. Thurston, *The geometry and topology of 3-manifolds*, Lecture Notes, Princeton University Press, Princeton, NJ.
- [26] F. Waldhausen, *On irreducible 3-manifolds which are sufficiently large*, Ann. of Math. (2) **87** (1968) 56–88.

OKLAHOMA STATE UNIVERSITY  
UNIVERSITY OF MELBOURNE, AUSTRALIA