

## Least Area Incompressible Surfaces in 3-Manifolds

Michael Freedman<sup>1</sup>, Joel Hass<sup>2</sup>, Peter Scott

<sup>1</sup> Mathematics Department, University of California at San Diego, La Jolla, CA-92093, USA

<sup>2</sup> Mathematics Department, University of Michigan, Ann Arbor, MI 48109, USA

Let  $M$  be a Riemannian manifold and let  $F$  be a closed surface. A map  $f: F \rightarrow M$  is called *least area* if the area of  $f$  is less than the area of any homotopic map from  $F$  to  $M$ . Note that least area maps are always minimal surfaces, but that in general minimal surfaces are not least area as they represent only local stationary points for the area function. The existence of least area immersions in a homotopy class of maps has been established when the homotopy class satisfies certain injectivity conditions on the fundamental group [18, 17].

In this paper we shall consider the possible singularities of such immersions. Our results show that the general philosophy is that least area surfaces intersect least, meaning that the intersections and self-intersections of least area immersions are as small as their homotopy classes allow, when measured correctly. One should note that evidence supporting this view had been found by Meeks-Yau in their embedding theorems for minimal disks and 2-spheres [13, 14].

Our main result asserts that if a least area immersion is homotopic to an embedding, then it has no self-intersections, which clearly exemplifies the above philosophy. The precise result is the following.

**Theorem 5.1.** *Let  $M$  be a closed,  $P^2$ -irreducible Riemannian 3-manifold and let  $F$  be a closed surface, not  $S^2$  or  $P^2$ . Let  $f: F \rightarrow M$  be a least area immersion such that  $f_*: \pi_1(F) \rightarrow \pi_1(M)$  is injective and such that  $f$  is homotopic to a two-sided embedding  $g$ . Then either*

- (i)  $f$  is embedding, or
- (ii)  $f$  double covers a one-sided surface  $K$  embedded in  $M$  and  $g(F)$  bounds a submanifold of  $M$  which is a twisted  $I$ -bundle over a surface isotopic to  $K$ .

This result answers conjectures of Meeks [12] and Uhlenbeck [22]. Partial results towards Theorem 5.1 have been obtained by Schoen and Shalen in the case of the torus and by Uhlenbeck [22] in the case of hyperbolic manifolds. Some of the ideas in this paper originated in Uhlenbeck's work. In [13] and [14], Meeks and Yau prove their versions of Dehn's Lemma and the Sphere

Theorem for analytic metrics and then approximate a smooth metric by analytic metrics to obtain the smooth case. They also need to use the deep theorem of Lojasiewicz that semi-analytic sets are triangulable. In this paper, we extend a technique due to Meeks and Yau [13] for approximating non-transverse minimal surfaces by transverse surfaces. This enables us to give a comparatively direct proof of our result with no mention of analytic metrics. Our methods could also be used in the Meeks-Yau papers [14] and [13] to give a very substantial simplification of their arguments.

A simple example of case (ii) of the theorem can be obtained by taking the product of the flat Moebius band with the circle [7]. There is a family of totally geodesic tori homotopic to the boundary of this 3-manifold, each of which has the same area and each of which minimizes area in its homotopy class. One member of this family double covers an embedded one-sided torus. An orientable example can be obtained by taking the orientable  $I$ -bundle over the Klein bottle and considering the class of tori homotopic to the boundary.

Theorem 5.1 and its method of proof allow us to prove many more results on intersections and self-intersections of least area immersions. The first result is a generalization of Theorem 5.1 and asserts that if a least area immersion is homotopic to a covering of an embedding then it is a covering of an embedding. Next we consider the self-intersections of a least area immersion which can not be homotoped to a covering of an embedding. The natural way to measure the complexity of the self-intersection of a self-transverse immersion is to count the double curves. This count needs care if the immersion contains curves of triple points and can not be made at all if the immersion is not self-transverse. In the case when the surface involved is the torus or Klein bottle, however, we show in §6 that any least area immersion is self-transverse and must have the minimal possible number of double curves. For higher genus surfaces, both these statements are false. We give in §8 an example of a least area immersion of the closed orientable surface of genus two into a 3-manifold which is self-transverse and has more than the minimal number of double curves. However, there is a different, but still natural, measure of the complexity of the self-intersection of an immersion, which is also defined for immersions which are not self-transverse. Using this measure, we show in §6 that any least area immersion has least self-intersection.

Simultaneously with the above, we also consider the intersection of two least area immersions. The first result we obtain states that if the two least area immersions are homotopic to disjoint maps, then they are either disjoint or their images coincide. If two least area immersions can not be homotoped apart, our results are similar to those on the self-intersections of a single surface. Two least area immersions, one of which is an immersed torus or Klein bottle, always intersect transversely and with the minimal number of double curves. This is false in general, but as before, an alternate method of counting complexity of intersections yields least intersection in all cases.

The organization of the paper is as follows; in §1 we discuss some of the basic properties of least area surfaces. In §2 we prove a crucial special case of the main theorem, when the surface is mapped in by a homotopy equivalence. In §3 we then make some simple but important deductions from this result.

We prove an existence theorem for the non-orientable case and extend the results of the previous section to this case. In §4 we prove a result in 3-dimensional topology which will be used later. In §5, we prove the embedding result, Theorem 5.1, and some related results. In §6, we apply the methods of §5 to show that least area surfaces intersect least. In §7, we consider extensions of our results to the relative case and to reducible 3-manifolds. In §8, we give two examples. The first is a simple example of two embedded least area surfaces in a 3-manifold which intersect non-transversely. The second is an example of a self-transverse least area immersion of a surface of genus two in a 3-manifold, which does not have the least possible number of double curves.

The main results of this paper make sense if the dimensions involved are all reduced by one, and the ideas of our proof still work. For example, Theorem 5.1 becomes the result that if  $f: S^1 \rightarrow F$  is an essential loop on a closed surface  $F$  which is homotopic to a two-sided embedding and is a shortest loop in its homotopy class, then  $f$  is an embedding or double covers a one-sided embedded circle in  $F$ . We have written up the results about shortest loops on surfaces in a separate paper [5]. The reduction in dimension eliminates many technical problems and we hope that [5] will demonstrate clearly the simplicity of the ideas in the present paper.

## §1. Results About Least Area Surfaces

We work in the  $C^\infty$  category. We will state our results for surfaces and manifolds with no boundary, although all the results go through straightforwardly if the 3-manifold is allowed to have a boundary whose mean curvature vector field is zero or inward pointing [15]. This will be discussed further in §7.

We will use the following terminology. A map  $f: F \rightarrow M$  is *incompressible* if it induces an injection on the fundamental group. Note that we depart from general usage by not requiring an incompressible map to be an embedding. Such a map is *two-sided* if its normal bundle is trivial. An immersion  $f$  is *self-transverse* if given two points  $x$  and  $x'$ , with  $f(x) = f(x')$ , there exist small discs about  $x$  and  $x'$  which are embedded by  $f$  and intersect transversely. Note that this is not the same as being in general position. It is possible for a self-transverse immersion to have an arc of triple points or a countable infinite set of triple points, for example.

Let  $F_1$  and  $F_2$  be a pair of embedded two-sided surfaces in a 3-manifold  $M$  which intersect transversely. A compact submanifold  $X$  of  $M$  is said to be a *product region* between  $F_1$  and  $F_2$  in  $M$  if there are compact subsurfaces  $S_1$  of  $F_1$  and  $S_2$  of  $F_2$  with  $S_1 \cap S_2 = \partial S_1 = \partial S_2$  and with  $S_1 \cup S_2 = \partial X$  and such that the pair  $(X, S_i)$  is homeomorphic to the pair  $(S_1 \times I, S_1 \times i)$ . We also require that  $\partial S_i$  be non-empty. Note that  $F_1$  and  $F_2$  are allowed to meet the interior of  $X$ .

When we consider covering spaces it will always be assumed that the Riemannian metric on the cover is induced by the covering map, so that the covering projection is a local isometry. We often refer to a map  $f: F \rightarrow M$  simply as a surface, in a manner similar to the way one talks about paths in manifolds.

Existence results for least area surfaces were established by Schoen-Yau [18] and by Sacks-Uhlenbeck [17]. Let  $M$  be a closed Riemannian  $n$ -manifold and let  $F$  be a closed orientable surface, not  $S^2$ . Let  $g: F \rightarrow M$  be incompressible. Then there is a branched immersion  $f: F \rightarrow M$  whose action on  $\pi_1$  is conjugate to that of  $g$  and  $f$  minimizes area among all such maps. If  $\pi_2(M) = 0$ , then  $f$  and  $g$  must be homotopic. In the case when  $M$  is 3-dimensional, the methods of Osserman [16] and Gulliver [6] show that  $f$  is actually an immersion. This establishes the following result which we state for future reference.

**Theorem 1.1.** *Let  $M$  be a closed  $P^2$ -irreducible Riemannian 3-manifold and let  $F$  be a closed orientable surface not  $S^2$ . If  $g: F \rightarrow M$  is an incompressible map, then there is a least area map  $f: F \rightarrow M$  which is homotopic to  $g$ , and any such can be parametrised as an immersion.*

*Remark.* We shall extend this to the case when  $F$  is not orientable in Theorem 3.2.

We will prove our results about least area surfaces by geometrical arguments involving little analysis. However, our methods are not totally independent of the existence theorem above. Although, all our results have the assumption that one is given a surface of least area, we need Theorem 1.1 in the proof of Lemma 3.3, and this lemma plays a crucial role in all our later results. It would be interesting to know if this dependence on Theorem 1.1 can be removed.

The purpose of the next lemma is to establish a useful property of least area surfaces and at the same time present an argument which will be used throughout the paper, the area swap argument.

**Lemma 1.2.** *Let  $F_1$  and  $F_2$  be a pair of two-sided, least area surfaces, embedded in a 3-manifold  $M$ , which meet transversely. Then there are no product regions between them.*

*Remark.* The result remains true if  $F_1$  and  $F_2$  are disjoint parallel surfaces, as our definition of product region required that  $\partial S_i$  be non-empty.

*Proof of Lemma 1.2.* Suppose there is a product region  $X$  between  $F_1$  and  $F_2$  with  $\partial X = S_1 \cup S_2$ . Consider the two surfaces  $F'_1 = F_1 - S_1 + S_2$  and  $F'_2 = F_2 - S_2 + S_1$ . The new surface  $F'_i$ , while possibly singular, is clearly homotopic to  $F_i$ ,  $i = 1, 2$ . As the total surface area of the two new surfaces equals that of the two old surfaces we must have the inequality

$$\text{area}(F'_i) \leq \text{area}(F_i)$$

for  $i = 1$  or  $2$ .  $F'_i$  however has a folding curve along  $\partial S_i$  whereas a least area map can be parameterized as an immersion and thus can not have a folding curve. Hence  $F'_i$  is not a least area surface, contradicting our assumption that  $F_i$  was. It follows that there can be no product region between  $F_1$  and  $F_2$ .

Lemma 1.2 will often be used to prove that two given least area surfaces are disjoint. However if the surfaces  $F_1$  and  $F_2$  do not meet transversely, then Lemma 1.2 is not applicable. It is quite possible for two least area surfaces to

meet non-transversely. We give a simple example in § 8. But a technique due to Meeks and Yau [14] allows one to circumvent this problem. The local nature of the intersections and self-intersections of minimal surfaces is well understood. See Lemma 1.4 below for details. A minimal immersion of a surface into a 3-manifold either factors through a covering of surfaces or is self-transverse except at a finite number of points. Further, in a neighbourhood of a point of non-transverse intersection of two sheets, the picture is that of a generalised saddle point.

A standard model of this situation in  $\mathbb{R}^3$  is to take one sheet as the  $x_1 x_2$ -plane and the other sheet to be given by  $x_3 = \text{Re}(z^n)$ , where  $z = x_1 + ix_2$ , and  $n$  is an integer,  $n \geq 2$ . Thus the intersection of the two sheets consists of a union of straight lines through the origin. In particular, the self-intersection set has no isolated points. Similar comments apply to the intersections of two minimal surfaces. The proof of the following lemma shows the Meeks-Yau technique in action.

**Lemma 1.3.** *If  $F_1, F_2$  are two-sided embedded incompressible surfaces of least area in a  $P^2$ -irreducible 3-manifold  $M$  and the inclusion of each is a homotopy equivalence, then they are either disjoint or have the same image.*

*Proof.* Assume first that the two surfaces intersect transversely. The methods of Waldhausen [23] or of § 4 show that there exists a component of  $M - \{F_1 \cup F_2\}$  which is a product piece and Lemma 1.2 gives a contradiction.

If they do not intersect transversely and their images do not coincide, we can find a point  $x$  in  $F_1 \cap F_2$  such that a small disk  $D_1$  about  $x$  in  $F_1$  intersects a small disk  $D_2$  about  $x$  in  $F_2$  transversely along an arc  $\alpha$ .

Cutting along this arc gives four half-disks,  $A_1$  and  $A_2$  on  $D_1$ , and  $B_1$  and  $B_2$  on  $D_2$ . These can be pasted together in one of two ways when a “cut and paste” takes place along the arc, either joining  $A_1$  to  $B_1$  and  $A_2$  to  $B_2$  or  $A_1$  to  $B_2$  and  $A_2$  to  $B_1$ . Either of these cut and pastes will produce a pair of folding curves along  $\alpha$ , which can be smoothed out to allow a reduction of area in a small ball about  $x$ . Let  $\epsilon$  be a sufficiently small constant so that either of the above cut and pastes yields an area reduction of at least  $\epsilon$ . We will show that we can perturb  $F_1$  by a small isotopy to get a new embedded surface  $F_3$  which is transverse to  $F_2$ , with  $\text{Area}(F_3) < \text{Area}(F_1) + \epsilon$  and such that  $F_3$  agrees with  $F_1$  on  $D_1$  and  $D_2$ . As  $F_2$  and  $F_3$  are transverse, we know that there is a product region between them and can carry out an area swap to obtain

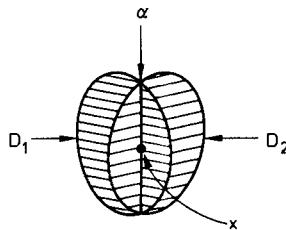


Fig. 1.1

surfaces  $F'_2$  and  $F'_3$  with the same total area but with folding curves. We must choose the point  $x$  to lie on one of these curves. Then we know that smoothing out the folds to obtain surfaces  $F''_2$  and  $F''_3$  will reduce the total area of our surfaces by at least  $\varepsilon$ . Thus

$$\begin{aligned} \text{Area}(F''_2) + \text{Area}(F''_3) &\leq \text{Area}(F'_2) + \text{Area}(F'_3) - \varepsilon \\ &= \text{Area}(F_2) + \text{Area}(F_3) - \varepsilon < \text{Area}(F_2) + \text{Area}(F_1). \end{aligned}$$

Hence either  $\text{Area}(F''_2) < \text{Area}(F_2)$  or  $\text{Area}(F''_3) < \text{Area}(F_1)$ . In either case, this contradicts the least area property of  $F_1$  or  $F_2$ , completing the proof of Lemma 1.3.

It remains to explain how to isotop  $F_1$  and how to choose  $x$  on an appropriate double curve. The idea is very simple but, for clarity, we defer the precise details to the proofs of Lemmas 1.4 and 1.5. Recall that a standard model for the situation in a neighbourhood of a point  $0$  at which  $F_1$  and  $F_2$  intersect non-transversely is to take  $F_1$  as the  $x_1x_2$ -plane in  $\mathbb{R}^3$  and  $F_2$  to be given by  $x_3 = \text{Re}(z^n)$ , where  $z = x_1 + ix_2$  and  $n \geq 2$ . Thus  $0$  is the origin of this model. It is easy to perturb  $F_1$  slightly in this model by pushing  $F_1$  upwards (or downwards) near  $0$ , and we can perform a similar isotopy of  $F_1$  near each point of non-transverse intersection with  $F_2$ . The new surface  $F_3$  will have area close to that of  $F_1$  and will be transverse to  $F_2$ . A further crucial fact is that the curves forming  $F_3 \cap F_2$  differ from  $F_1 \cap F_2$  only near the non-transverse points of  $F_1 \cap F_2$  and that the change which occurs is as shown in Fig. 1.2. Case a) or b) will occur depending on whether we push  $F_3$  up or down near  $0$ . This last fact is what allows one to choose the point  $x$  appropriately. One might be unable to choose  $x$  appropriately if the change from  $F_1 \cap F_2$  to  $F_3 \cap F_2$  introduced some small circle near  $0$ , for example. The clearest way to see how to choose  $x$  is actually to choose many points  $x_i$ , as follows. Recall that  $F_1 \cap F_2$  is a 1-manifold except at a finite number of points where the picture consists of  $n$  lines with a common point. Hence  $F_1 \cap F_2$  with the non-manifold points removed has only finitely many components and we choose a point  $x_i$  on each such component. As before, each  $x_i$  gives rise to a positive number  $\varepsilon_i$  which is a lower bound for the area decrease of  $F_1 \cup F_2$  obtained by cut and paste on the double curve through  $x_i$ ; and we let  $\varepsilon$  denote the minimum of the  $\varepsilon_i$ 's. When we perturb  $F_1$  to  $F_3$ , the fact that the double curves alter as described means that each component of  $F_3 \cap F_2$  contains at least one

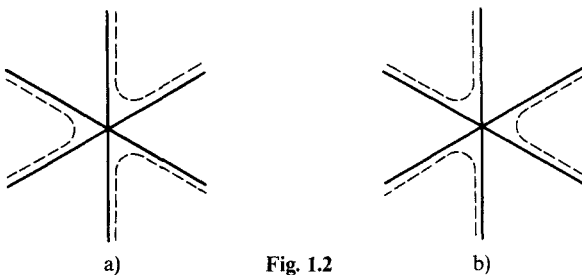


Fig. 1.2

$x_i$ , and this allows the previous argument to apply to give the required contradiction.

Later in this paper, we will need a refined version of this technique. Before discussing this refinement, however, we need to establish more precisely the local description of the situation when two sheets of minimal surface meet non-transversely. Here is the statement we need.

**Lemma 1.4.** *Let  $M$  be a Riemannian 3-manifold and let  $F$  be a closed surface. Let  $f: F \rightarrow M$  be a minimal immersion and let  $a$  and  $a'$  be distinct points of  $F$  such that  $f(a) = f(a')$  and the two sheets of  $f(F)$  through  $f(a)$  have a common tangent plane. Choose a chart, with coordinates  $(x_1, x_2, x_3)$ , in the smooth structure on  $M$  so that the origin  $0$  is at  $f(a)$  and the  $x_1x_2$ -plane is tangent to both sheets of  $f(F)$  at  $f(a)$ . Let  $V$  and  $V'$  be sufficiently small neighbourhoods in  $F$  of  $a$  and  $a'$  respectively, so that  $f(V)$  and  $f(V')$  are the graphs of smooth functions  $\phi$  and  $\phi'$  on some neighbourhood  $U$  of  $0$  in the  $x_1x_2$ -plane.*

*Then either  $\phi = \phi'$  on some neighbourhood of  $0$  in which case the two sheets coincide in this neighbourhood of  $0$ , or there is a  $C^1$ -change of coordinates in the  $x_1x_2$ -plane which transforms the function  $\phi - \phi'$  into  $c \cdot \text{Re}(z^N)$  on some neighbourhood of  $0$ , where  $z = x_1 + ix_2$  and  $N$  is some integer,  $N \geq 2$ , and  $c$  is some non-zero constant.*

*Remark.* It follows that either  $f: F \rightarrow M$  factors through a covering of surfaces or the non-transversal points are isolated, and in a neighbourhood of a non-transversal point the double curve picture is  $C^1$ -equivalent to  $n$  straight lines in  $\mathbb{R}^2$  with a common point.

*Proof of Lemma 1.4.* This proof is based on the proof of Lemma 2 of [14]. Let  $P$  denote the function  $\phi - \phi'$  on  $U$  and suppose that  $P$  is not identically zero on any neighbourhood of  $0$ . Since both  $\phi$  and  $\phi'$  satisfy the minimal surface equation in  $M$ , it follows that  $P$  satisfies a second order linear homogeneous elliptic P.D.E. with smooth coefficients, say  $L(P) = 0$ . Also  $P$  vanishes at the origin. A theorem of Bers [1] asserts that, as  $P$  is not identically zero on any neighbourhood of  $0$ , there is a non-zero homogeneous polynomial  $\rho_N(x)$  of degree  $N \geq 2$  such that

$$P(x) = \rho_N(x) + O(|x|^{N+\epsilon}), \quad \text{where } 0 < \epsilon < 1,$$

and

$$\frac{\partial^l}{\partial x_1^r \partial x_2^s} P(x) = \frac{\partial^l}{\partial x_1^r \partial x_2^s} \rho_N(x) + O(|x|^{N-l+\epsilon}),$$

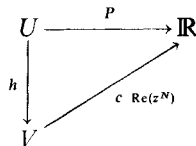
for all  $l \leq N$ , and  $\rho_N(x)$  satisfies the equation  $L_0(\rho_N) = 0$ , where  $L_0$  is obtained from the second-order part of  $L$  by replacing each coefficient  $a(x)$  by the number  $a(0)$ . Thus  $\rho_N(x)$  satisfies a second order linear homogeneous elliptic P.D.E. with constant coefficients.

By a linear change of coordinates in the  $x_1x_2$ -plane, we can assume that  $\rho_N(x)$  is harmonic i.e. satisfies  $\Delta \rho_N = 0$ . A further orthogonal change of coordinates in the  $x_1x_2$ -plane will arrange that  $\rho_N(x) = c \cdot \text{Re}(z^N)$ , where  $z = x_1 + ix_2$  and  $c$  is constant. Hence the only critical point of  $\rho_N$  is the origin. Now a linear coordinate change will not affect the order of the error term  $P(x) - \rho_N(x)$ ,

so we still have  $P(x) = \rho_N(x) + O(|x|^{N+\epsilon})$ , where  $0 < \epsilon < 1$ , and the same comment applies to the order of the  $l$ -th partial derivatives of  $P(x) - \rho_N(x)$ , for  $l \leq N$ , where  $\rho_N(x)$  is now equal to  $c \cdot \text{Re}(z^N)$ . In particular, all the  $l$ -th partial derivatives at 0 of  $P(x)$  are equal to the corresponding derivatives at 0 of  $\rho_N(x)$ , when  $l \leq N$ . Thus the  $N$ -jets at 0 of  $P(x)$  and  $\rho_N(x)$  are equal. Of course, this jet is zero apart from the terms of order  $N$ . One can now see that the result we need is a standard type of result in singularity theory. We have a given  $N$ -jet  $\rho_N(x)$  and wish to show that given any smooth function  $P(x)$  with this  $N$ -jet at 0, there is a  $C^1$ -change of coordinates in the  $x_1 x_2$ -plane transforming  $P(x)$  into  $\rho_N(x)$  in a neighbourhood of 0, i.e. we want to show that the  $N$ -jet of  $\rho_N(x)$  is  $C^1$ -sufficient. The result we need is that if  $\rho_N(x)$  is a homogeneous polynomial of degree  $N$  and if there are positive numbers  $\epsilon$  and  $\delta$  such that  $|\nabla \rho_N(x)| \geq \epsilon |x|^{N-\delta}$ , for all  $x$  in some neighbourhood of 0, then  $\rho_N(x)$  is  $C^1$ -sufficient. This is proved by Cheng [2]. See also p. 153 of Lu's book [11]. Now  $|\nabla \text{Re}(z^N)| = N|z|^{N-1}$ , so  $c \cdot \text{Re}(z^N)$  clearly satisfies the required condition. Note that, in general,  $C^1$ -sufficiency involves a  $C^1$ -change of coordinates in source and target. However, because our functions are real valued, a change of coordinates in the source suffices. This is also in [2] and [11]. This completes the proof of Lemma 1.4.

We now have a precise picture of the situation when two sheets of minimal surfaces are tangent. Next we must consider perturbations of this situation as already discussed in the proof of Lemma 1.3. Note that if we perform a smooth isotopy in our model, we may only obtain a  $C^1$ -isotopy in the manifold  $M$ . This is not a serious problem, as areas still change continuously under a  $C^1$ -isotopy so the arguments of Lemma 1.3 can be applied using a  $C^1$ -surface  $F_3$ . Also a  $C^1$ -surface can be perturbed to a smooth surface by an  $\epsilon$ -isotopy, so one could still obtain a smooth surface  $F_3$  with all the properties required for the proof of Lemma 1.3. However, for our more refined applications it will be convenient to define our isotopy in a chart for the smooth structure on  $M$ .

**Lemma 1.5.** *Let  $U$  be a neighbourhood of the origin in the  $x_1 x_2$ -plane and let  $\phi$  and  $\phi'$  be smooth functions on  $U$  with  $\phi - \phi'$  denoted  $P$  such that  $P(x) = c \cdot \text{Re}(z^N) + O(|x|^{N+\epsilon})$  where  $\epsilon > 0$ . Let  $h$  be a  $C^1$ -diffeomorphism of  $U$  with a neighbourhood  $V$  of the origin in the  $y_1 y_2$ -plane such that the diagram below commutes.*



Let  $r$  be a positive real number such that  $U$  contains the disc of radius  $r$  center 0 and let  $\psi$  be a smooth function on  $U$  such that  $\psi(x) = 0$  if  $|x| \geq r$  and  $\psi(x) = 1$  if  $|x| \leq 1/2r$ . For any real number  $t$ , let  $\phi_t(x) = \phi(x) + t \cdot \psi(x)$ .

Then there is a  $t_0 > 0$  such that if  $0 < |t| \leq t_0$ , the graph  $\Gamma(\phi_t)$  of  $\phi_t$  meets  $\Gamma(\phi')$  transversely in curves obtained from  $\Gamma(\phi) \cap \Gamma(\phi')$  as shown in Fig. 1.2.

*Proof.* First observe that if  $|x| \geq r$ , then  $\phi_t(x) = \phi(x)$ , so that

$$\Gamma(\phi_t) \cap \Gamma(\phi') = \Gamma(\phi) \cap \Gamma(\phi')$$



when  $|x| \geq r$  for all  $t$ . Next consider  $|x| \leq 1/2r$ . For such  $x$ , we have  $\phi_t(x) = \phi(x) + t$ . Thus  $\Gamma(\phi_t) \cap \Gamma(\phi')$  is simply a level curve of  $P$ , namely the curve  $P(x) = t$ . But  $h$  defines a  $C^1$ -diffeomorphism of the level curves of  $P$  with the level curves of  $c \cdot \text{Re}(z^N)$ . Hence  $\Gamma(\phi_t) \cap \Gamma(\phi')$  has the required shape when  $|x| \leq 1/2r$ , for all  $t$ . Finally, when  $1/2r \leq |x| \leq |r|$ , we use the fact that  $\Gamma(\phi)$  is transverse to  $\Gamma(\phi')$ . Thus there is  $t_0 > 0$  such that when  $0 \leq t \leq t_0$ ,  $\Gamma(\phi_t)$  is still transverse to  $\Gamma(\phi')$ , when  $1/2r \leq |x| \leq |r|$ , and  $\Gamma(\phi_t) \cap \Gamma(\phi')$  is isotopic to  $\Gamma(\phi) \cap \Gamma(\phi')$ . These three facts together complete the proof of Lemma 1.5.

Now let  $M$  be a Riemannian 3-manifold,  $F$  a closed surface and  $f: F \rightarrow M$  a minimal immersion which does not factor through a covering of surfaces. We want to perturb  $f$  to be self-transverse. Lemma 1.4 shows that the non-transverse points of  $f$  are isolated. If two sheets  $V$  and  $V'$  of  $f(F)$  pass through a point 0 and have common tangent planes, then Lemma 1.5 shows how to regularly homotop  $f$  so that  $V$  and  $V'$  become transverse, and this homotopy can be supported on as small a neighbourhood of 0 as desired. Hence, if no other sheets of  $f(F)$  pass through 0, we can choose our regular homotopy of  $f$  so that the double curves of  $f$  do not alter except in  $V$  and  $V'$ . If a sheet  $W$  of  $f(F)$  passes through 0 but is transverse to  $V$  and  $V'$ , then a suitably small isotopy of  $V$  will only alter  $V \cap W$  by an isotopy in  $V$  and  $W$ . Only finitely many sheets of  $f(F)$  can pass through 0, so a suitably small isotopy of  $V$  will only alter  $V \cap W_i$  by an isotopy for every sheet  $W_i$  of  $f(F)$  through 0 which is transverse to  $V$  and  $V'$ .

Finally we may have many sheets of  $f(F)$  through 0, all tangent to  $V$ . In this case, choose a chart in the smooth structure of  $M$  such that the common tangent plane at 0 is the  $x_1x_2$ -plane. For a suitably small neighbourhood  $U$  of the origin in the  $x_1x_2$ -plane, each sheet is given as the graph of some smooth function on  $U$ . For a suitably small neighbourhood of 0, the intersection of  $V$  with each of the other sheets  $W_i$  tangent at 0 is  $C^1$ -equivalent to the standard model as proved in Lemma 1.4. Now we deform the sheet  $V$  exactly as in Lemma 1.5. For a suitably small such deformation, the intersection of  $V$  with each  $W_i$  will alter as described in Lemma 1.5. Now we have one less sheet through 0 and can repeat the argument using still smaller deformations. The following is a precise statement of what we have achieved. For convenience, we have taken  $t_0$  to be 1, which can always be arranged by a scale change in the  $t$ -parameter.

**Lemma 1.6.** *Let  $M$  be a Riemannian 3-manifold, let  $F$  be a closed surface and let  $f_0: F \rightarrow M$  be a minimal immersion which does not factor through a covering of surfaces. Let  $\varepsilon$  be a positive constant. Then there is a regular homotopy  $f_t: F \rightarrow M$ ,  $0 \leq t \leq 1$ , such that*

- (i)  $f_t$  coincides with  $f_0$  outside an  $\varepsilon$ -neighbourhood of the points of non-transverse intersection,
- (ii)  $f_t$  is self-transverse, if  $t > 0$ ,
- (iii) the double curves of  $f_t$ ,  $t > 0$ , are obtained from those of  $f_0$  as shown in Fig. 1.2.

*Remarks.* The maps  $f_t$  need not be in general position. However, general position maps  $F \rightarrow M$  are certainly dense in the space of all maps  $F \rightarrow M$ . For

given  $t$ , a general position map close enough to  $f_t$  will have isotopic double curves i.e. corresponding double curves in  $F$  will be individually isotopic in  $F$ . Of course the entire double sets need not be isotopic but this will be irrelevant. Hence, it is clear that we can find general position maps  $F \rightarrow M$  with double curves isotopic to those of  $f_1$  and with area as close as we like to the area of  $f$ .

We will want to apply the same technique to a general minimal immersion  $f$  as we used in Lemma 1.3 when considering two embedded minimal surfaces. As in that argument, one chooses point  $x_i$  on the double curves of  $f(F)$ , so that there is one on each segment of each double curve, where the segments are formed by removing the points of nontransversal intersection. If there is a curve of  $n$ -tuple points, one needs to regard this as  $\binom{n}{2}$  double curves with the same image and we choose  $\binom{n}{2}$  distinct points  $x_i$  one for each intersecting pair of sheets. For each point  $x_i$  consider two sheets passing through  $x_i$  and ignore the others. One chooses a ball  $B_i$  center  $x_i$  and obtains a number  $\epsilon_i$  which is a lower bound for the area decrease in  $f$  obtained by cut and paste of the two sheets through  $x_i$  along their double curve. Let  $\epsilon$  denote the minimum of the  $\epsilon_i$ 's. One can find a general position map  $f' : F \rightarrow M$  with double curves isotopic to those of  $f_1$  such that  $\text{Area}(f') < \text{Area}(f) + \epsilon$  and for each  $i$ ,  $f'$  agrees with  $f$  on the two 2-discs of  $F$  mapped into  $B_i$ . Thus any cut and paste on  $f'(F)$  will yield surfaces which can be deformed to have total area strictly less than the area of  $f$ .

## §2. Least Area Homotopy Equivalences

**Theorem 2.1.** *Let  $M$  be an oriented Riemannian 3-manifold without boundary of the homotopy type of some closed oriented surface  $F$ , not  $S^2$ . If  $f : F \rightarrow M$  is a least area, immersed homotopy equivalence then  $f$  is an embedding.*

*Proof.* We will start by assuming that  $f$  is in general position. Set  $N =$  regular neighborhood of  $f(F)$ ,  $N \subset M$ . Since the natural map  $g : F \rightarrow N$  factors  $f$  which is a homotopy equivalence,  $g_j : H_j(F; \mathbb{Z}_2) \rightarrow H_j(N; \mathbb{Z}_2)$  is an injection,  $j = 1$  or  $2$ . We begin by supposing  $g_1$  is also onto. All homology and intersection theory will be with coefficients  $\mathbb{Z}_2$  unless specified to the contrary.

It is well known that there must be an embedding of  $F$  in  $M$  which is a homotopy equivalence. From this the following fact is easily deduced:

*Fact.*  $M$  has two ends and the generator  $\beta \in H_2(M)$  is characterized by having intersection number = 1 with any line  $l$  running from one end to the other.

It follows that any such  $l$  meets  $f(F)$ . Thus any  $l$  meets  $N$ . Thus any  $l$  meets  $\partial N$ . So  $\partial N$  separates the ends of  $M$ . It follows that some possibly disconnected union of boundary components  $A \subset \partial N$  must carry  $\beta$ ,  $\text{inc}_*[A] = \beta \in H_2(M)$ . Since  $A$  is bordant via  $N$  to the sum of all other boundary components, if we set  $B = \partial N - A$ ,  $A$  and  $B$  each carry  $\beta$ . Thus we have:

$$\text{inc}_2[A] = \text{inc}_2[B] = g_2[F] \in H_2(N). \tag{*}$$

Call this homology class  $\alpha$ ; inclusion carries  $\alpha$  to  $\beta$ .

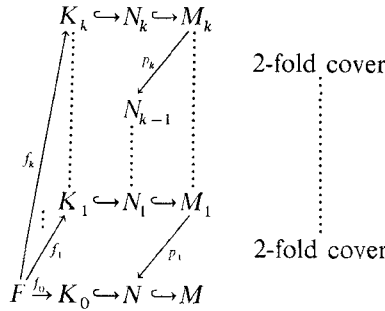
At this point there are two cases to consider:

*Case 1.*  $g_1: H_1(F) \rightarrow H_1(N)$  is onto and, by a previous remark,  $g_1$  is an isomorphism.

*Case 2.*  $g_1$  is not onto.

In case 1 we are content to work with the manifold  $N$ . In case 2 we must use a tower argument to pass to a manifold  $N_k$ . First we construct  $N_k$  and second we will derive an appropriate version of the formula (\*) for  $N_k$ . As  $g_1$  is not onto, there is a connected double covering space  $M_1$  of  $N$  such that  $f: F \rightarrow N$  lifts to a map  $f_1: F \rightarrow M_1$ . Let  $N_1$  be the regular neighbourhood of  $f_1(F)$ . By repeating this argument for  $N_1$ , one constructs in the usual way [8] a tower of double coverings which must terminate for the usual reason. We set  $K_i = f_i(F)$ .

Here is the complete tower:



In case 1,  $N_k$  would simply be  $N$ . In any case,  $H_1(F) \rightarrow H_1(N_k)$  is an isomorphism.

$K_i$  is a quotient space of  $F$  with quotient map  $f_i$ ,  $0 \leq i \leq k$ . The quotient map identifies pairs of normally immersed circles in  $F$  and at the normal crossings triples are identified.

**Lemma 2.2.** All intersection numbers  $x \cdot f_k[F]$  are zero for  $x \in H_1(N_k)$ .

*Proof.* Let  $c \subset F$  be a loop. Because the normal bundle of  $F \rightarrow N_k$  is trivial, the intersection of  $c$  in  $N_k$  with  $f_k[F]$  can be computed as  $\sum_i c \cdot \gamma_i$  where  $\gamma_i$  is the collection of immersed curves identified by  $f_k$ . But the  $\gamma_i$ 's come in pairs and since  $f_k$  factors a homotopy equivalence the pairs of immersed circles identified under  $f_k$  must be freely homotopic in  $F$ , thus  $\sum_i c \cdot \gamma_i = 0$ . The proof is completed by remarking that any class  $x$  is in the image of some loop  $c$  since  $H_1(F) \rightarrow H_1(N_k)$  is an isomorphism. ■

Here is our generalization of (\*):

**Corollary 2.3.**  $N_k - K_k$  admits a two coloring. Let  $A_k$  denote the black and  $B_k$  the white boundary components.  $A_k$  and  $B_k$  each carry  $f_k[F]$ .

*Proof.* We follow an argument given in [4]. Color one point of  $N_k - K_k$  white and let the color of any other point be determined by the intersection number with  $K_k$  (or  $f_k[F]$ ). Lemma 2.2 shows that this procedure is well defined. The second assertion is proved by constructing a deformation retraction  $N_k \rightarrow K_k$  which carries both  $A_k$  and  $B_k$  onto  $K_k$  as  $Z_2$ -homology cycles. ■

If  $k=0$  (case 1) we may regard  $A_k = A$  and  $B_k = B$ .

**Lemma 2.4.**  $A_k$  consists of the disjoint union of a surface  $F_k$  diffeomorphic to  $F$  and spheres,  $A_k = F_k \sqcup S^2$ 's. Similarly,  $B_k = G_k \cup S^2$ 's.

*Proof.* In the exact homology sequence of the pair  $(N_k, \partial N_k)$  the two maps

$$H_2(N_k, \partial N_k) \xrightarrow{p} H_1(\partial N_k) \xrightarrow{q} H_1(N_k)$$

are dual,  $p$  being the Poincaré dual of  $H^1(N_k) \rightarrow H^1(\partial N_k)$  which is  $\text{hom } q$ . Thus  $\text{image } p \cong \text{kernel } q \cong \text{coker } p$ . Set  $d = \text{dimension } H_1(N_k) = \text{dimension } H_1(F) = 2 \text{ genus } F$ . Now  $\text{dimension (image } p) \leq \dim H_2(N_k, \partial N_k) = d$  so

$$\text{dimension } H_1(A_k) + \text{dimension } H_1(B_k) = \text{dimension } H_1(\partial N_k) \leq 2d. \quad (**)$$

Mapping down the tower we find maps  $A_k \rightarrow M \simeq F$  and  $B_k \rightarrow M \simeq F$  of degree 1 (mod 2). In particular, the maps  $A_k \rightarrow F$  and  $B_k \rightarrow F$  are of non-zero degree, and so  $A_k$  and  $B_k$  must each have a component whose genus equals or exceeds the genus of  $F$ . Consulting (\*\*) we see that this component must have the same genus as  $F$  and map to  $F$  with degree one (mod 2) and the remaining components of  $A_k$  and  $B_k$  must be spheres.

We need the following criterion for recognizing when a map from  $F$  to  $M$  is a homotopy equivalence.

**Lemma 2.5.** A map  $F \xrightarrow{h} M$  which induces an isomorphism on homology with  $Z_2$ -coefficients and is an embedding is necessarily a homotopy equivalence.

*Proof.* Compose with the homotopy inverse of  $f: F \rightarrow M$ . The composite map  $F \rightarrow F$  is of odd degree. As it induces an isomorphism on  $H_1(F)$ , this degree must be one except possibly when  $F$  is the torus. In this case we use the fact that  $h$  is an embedding to conclude that  $h$  is degree one. This is because divisible integral classes in an oriented  $n$ -manifold cannot be the image of the fundamental class of a connected embedded  $(n-1)$ -manifold. It follows that  $h$  must be a homotopy equivalence in all cases.

If  $f_k$  is not an embedding, Meeks and Yau [14] show how to choose the regular neighborhood  $N_k$  so that  $\text{area } \partial N_k < 2 \text{ area } f_k(F)$ . Thus either  $F_k$  or  $G_k$  (say  $F_k$ ) will have less area than  $f_k(F)$ . If  $f_k$  is an embedding we omit the regular neighborhood altogether and will use the same symbol  $F_k$  for  $f_k(F) = K_k$ . So in either case  $\text{area } F_k \leq \text{area } f_k(F)$ .

Under the first projection  $p_k: F_k \rightarrow N_{k-1}$  pairs of disjointly embedded circles may be identified. Since the composition down the tower  $F_k \rightarrow M$  is a  $Z_2$ -homology equivalence, the identified circles must be homologous. Thus such a pair bounds a region  $R_i \subset F$ , and  $R_i$  may be cut out and glued back to obtain a new surface  $F'_k$  diffeomorphic to  $F_k$ . But now when  $p_k$  is restricted to  $F'_k$  a double curve can be eliminated by a small, area reducing perturbation called

“rounding the corner” [14]. If this is done, sequentially for all the pairs then we will call the resulting embedded surface  $F_{k-1} \subset N_{k-1}$ . Since the cutting and gluing did not affect the  $Z_2$  homology class,  $p_k[F_k] = [F_{k-1}] \in H_2(N_{k-1})$  we may repeat the procedure to obtain  $F_{k-2}$ . Proceeding down the tower we finally arrive at  $F_0 \subset N \subset M$ . The inclusion is a  $Z_2$ -homology equivalence so by Lemma 2.5 it is a homotopy equivalence.

Let  $x_k: N_k \rightarrow M$  be the composition which descends the tower diagram. Covering projections are local isometries and so do not change area. The procedure that eliminates double curves – cut-glue-round corners – may always be taken to reduce area. Since  $F_0$  results from  $F_k$  by a sequence of these two procedures  $\text{area}(F_0) \leq \text{area}(F_k)$ . Equality holds if and only if  $x_k|_{F_k}$  is 1–1, that is only if there are no double curves to eliminate. But  $\text{area}(F_k) \leq \text{area}(f_k(F))$  with equality if and only if  $f_k$  is an embedding. Also  $\text{area}(f_k(F)) = \text{area}(f_0(F))$ . By our least area hypothesis  $\text{area}(f_0(F)) \leq \text{area}(F_0)$  forcing the two earlier inequalities to be equalities. As a result  $f_k$  and  $x_k|_{f_k(F)}$  are both 1–1. Thus  $f = x_k \circ f_k$  is also 1–1. This completes the proof of Theorem 2.1, in the case when  $f$  is in general position.

If  $f: F \rightarrow M$  is not in general position, we perturb it to a general position map  $f'$  as described in Lemma 1.6 and the remarks which follow it. We then apply all the foregoing arguments to  $f'$ . The arguments up to and including Lemma 2.5 do not refer to area and so they all apply.

Immediately after Lemma 2.5, we say that if  $f_k$  is not an embedding, we can find a regular neighbourhood  $N_k$  of  $f_k(F)$  so that  $\text{Area } \partial N_k < 2 \text{Area } f_k$ . Of course,  $\text{Area } f_k = \text{Area}(f)$ . In our new situation, the last sentence of §1 shows that if the corresponding map  $f'_k$  is not an embedding, we can find a regular neighbourhood  $N_k$  of  $f'_k(F)$  so that  $\text{Area}(\partial N_k) < 2 \text{Area}(f)$ . Now the rest of the argument applied to the maps  $f'_k$  yields a contradiction unless  $f'$  is an embedding. It follows that  $f$  must be an embedding which completes the proof of Theorem 2.1.

**Corollary 2.6.** *Let  $F$  be a closed surface, not  $S^2$  or  $P^2$  and let  $M$  be a  $P^2$ -irreducible Riemannian 3-manifold. Let  $f_i: F \rightarrow M$  be two least area immersions,  $i = 1, 2$ . If  $f_1$  is homotopic to  $f_2$  by a homotopy fixing a base point then  $f_1(F) = f_2(F)$ .*

*Proof.* Let  $M'$  be the cover of  $M$  determined by  $f_{i\#} \pi_1(F)$ . Consider the lift of  $f_1$  to  $M'$ . Since  $f_2$  is homotopic to  $f_1$  it too will lift to  $M'$  by Theorem 2.1. Since the homotopy fixes a point the two lifted maps will intersect. Both lifts are least area and must be embeddings. Lemma 1.3 shows that there is a product region between the two lifts if they intersect transversely, which then yields a contradiction. The assumption of transversality is dealt with as before.

### §3. Coverings of Least Area Maps

We start this section by using the results of §2 to extend Theorem 1.1 to the case where  $F$  is non-orientable and two-sided.

**Theorem 3.1.** *Let  $M$  be a  $P^2$ -irreducible 3-manifold that covers a closed Riemannian 3-manifold and let  $F$  be a closed surface not  $S^2$  or  $P^2$ . Let  $f: F \rightarrow M$  be a*

two-sided incompressible map. Then there exists a least area immersion homotopic to  $f$ .

*Proof.* If  $M$  and  $F$  are orientable this is known [17, 18] for the case when  $M$  is compact. This extends to the case when  $M$  covers a compact 3-manifold  $M'$  by noting that a least area map  $f'$  in the homotopy class of maps  $\pi f$ , where  $\pi$  is the covering projection, exists by the compact version of the theorem.  $f'$  lifts to a least area immersion in the homotopy class of  $f$ .

If  $F$  and  $M$  are non-orientable let  $\tilde{F}$  and  $\tilde{M}$  denote their orientable double coverings and let  $\tilde{f}: \tilde{F} \rightarrow \tilde{M}$  denote the map covering  $f$ . This map exists because of our assumption that  $f$  was two-sided. Let  $M_F$  denote the covering of  $M$  with  $\pi_1(M_F) = f_*(\pi_1(F))$  and let  $\tilde{M}_F$  denote the covering of  $\tilde{M}$  with  $\pi_1(\tilde{M}_F) = \tilde{f}_*(\pi_1(\tilde{F}))$ . Clearly  $\tilde{M}_F$  is the orientable double covering of  $M_F$ . See Fig. 3.1. Let  $\tau$  denote the covering involution of  $\tilde{M}_F$ . Let  $f_1: F \rightarrow M_F$  be the lift of  $f$  and let  $\tilde{f}_1: \tilde{F} \rightarrow \tilde{M}_F$  be the map covering  $f_1$ . The existence theorem for the orientable case shows that  $\tilde{f}: \tilde{F} \rightarrow \tilde{M}$  is homotopic to a least area map and hence so is  $\tilde{f}_1: \tilde{F} \rightarrow \tilde{M}_F$ . Let  $\tilde{g}: \tilde{F} \rightarrow \tilde{M}_F$  denote the least area map. Theorem 2.1 shows that  $\tilde{g}$  is an embedding. Now Lemma 1.3 shows that  $\tilde{g}(\tilde{F})$  and  $\tau\tilde{g}(\tilde{F})$  are either disjoint or equal, as both surfaces are two-sided in  $\tilde{M}_F$ . In either case  $\tilde{g}(\tilde{F})$  projects to an embedded incompressible surface in  $M_F$ . But any such surface must have fundamental group equal to  $\pi_1(F)$ , since  $f_1(F)$  is two-sided in  $M_F$  [3, 9, 19]. We must therefore have  $\tau\tilde{g}(\tilde{F}) = \tilde{g}(\tilde{F})$ , thus determining an embedding  $g$  of  $F$  in  $M_F$  which is homotopic to  $f_1(F)$ . This embedding must be a least area map for if it were homotopic to a map  $h: F \rightarrow M$  of less area, the double covering map  $\tilde{h}: \tilde{F} \rightarrow \tilde{M}_F$  would be a map of less area than  $\tilde{g}$  contradicting the fact that  $\tilde{g}$  is elast area. This uses the obvious fact that the area of  $\tilde{g}: \tilde{F} \rightarrow \tilde{M}_F$  is exactly twice the area of  $g: F \rightarrow M_F$ . This least area embedding  $g: F \rightarrow M_F$  projects into  $M$  to give a least area map from  $F$  to  $M$  homotopic to the original map  $f$ . This concludes the proof of Theorem 3.1.

We can use essentially the same argument to extend Theorem 2.1 to the non-orientable case.

**Theorem 3.2.** *Let  $M$  be a  $P^2$ -irreducible Riemannian 3-manifold and let  $F$  be a closed surface not  $S^2$  or  $P^2$ . Let  $f: F \rightarrow M$  be a two-sided least area map which is a homotopy equivalence. Then  $f$  is an embedding.*

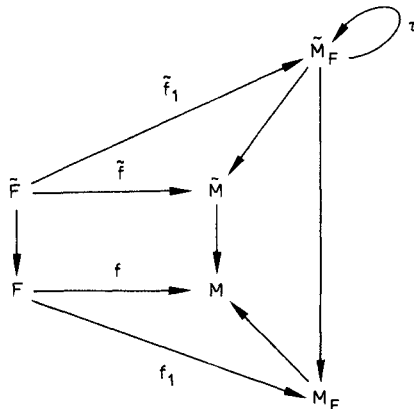


Fig. 3.1

The same argument also yields some results about the behaviour of coverings of least area maps.

**Lemma 3.3.** *Let  $M$  be a 3-manifold which covers a closed Riemannian 3-manifold. If  $f: F \rightarrow M$  is a two-sided homotopy equivalence,  $\tilde{M}$  is a finite cover of  $M$  and  $\tilde{F}$  is the corresponding cover of  $F$ , so that the following diagram commutes,*

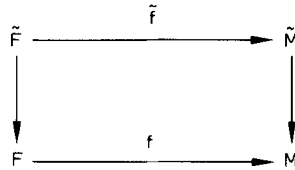


Fig. 3.2

then  $f$  is least area if and only if  $\tilde{f}$  is least area.

*Proof.* Clearly if  $\tilde{f}$  is least area, so is  $f$ . If not, there exists a map  $g: F \rightarrow M$  of smaller area which is homotopic to  $f$  and a corresponding lift  $\tilde{g}$  which is of smaller area than  $\tilde{f}$ . This argument uses the obvious fact that the area of  $\tilde{g}(\tilde{F})$  is  $d$  times the area of  $g(F)$  and the area of  $\tilde{f}(\tilde{F})$  is  $d$  times the area of  $f(F)$ , where  $d$  is the degree of the covering map  $\tilde{M} \rightarrow M$ .

Conversely, assume first that the cover is regular and  $f$  has least area. Let a least area map in the homotopy class of  $\tilde{f}$  be  $\tilde{g}$ . (Note the use of Theorem 3.1 to ensure the existence of  $\tilde{g}$ .) Then  $\tilde{g}$  is an embedding of  $\tilde{F}$ , and the covering translates of  $\tilde{g}(\tilde{F})$  all intersect trivially by Lemma 1.3, as each is two-sided. This implies that  $\tilde{g}(\tilde{F})$  projects down to a cover of an embedded surface in  $M$ . This surface,  $F_1$ , is homotopic to  $F$  as it is an embedded incompressible surface in  $M$ . If  $\tilde{g}$  has less area than  $\tilde{f}$ , then the area of  $F_1$  is smaller than that of  $f(F)$ , yielding a contradiction.

Assume now that the cover is not regular. Then  $\pi_1(\tilde{M})$  is not a normal subgroup of  $\pi_1(M)$ . However there exists a subgroup of  $\pi_1(\tilde{M})$  which is of finite index and normal in  $\pi_1(M)$ ; just intersect all the conjugates of  $\pi_1(\tilde{M})$ . There exists a corresponding regular cover  $\tilde{\tilde{M}}$  and a corresponding lift  $\tilde{\tilde{f}}: \tilde{\tilde{F}} \rightarrow \tilde{\tilde{M}}$ . See Fig. 3.3. As  $f$  is least area, it follows that  $\tilde{\tilde{f}}$  has least area. It then also follows from the above that  $\tilde{f}$  has least area.

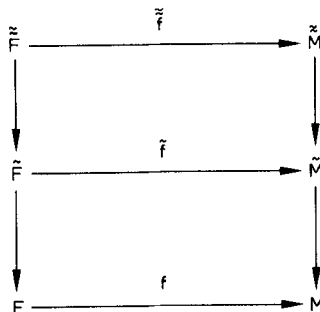


Fig. 3.3

**Theorem 3.4.** *Let  $M$  be a  $P^2$ -irreducible 3-manifold which covers a compact Riemannian 3-manifold. Let  $f: F \rightarrow M$  be a two-sided map which induces an injection of the fundamental group. Let  $M_1$  be a covering of  $M$  and let  $F_1$  be a finite  $k$ -fold cover of  $F$  such that  $f_1: F_1 \rightarrow M_1$  is a lift of  $f$ . Then  $f$  is least area if and only if  $f_1$  is least area.*

*Remark.*  $M_1$  may be an infinite covering of  $M$  or even  $M$  itself.

*Proof.* Consider the following diagram:

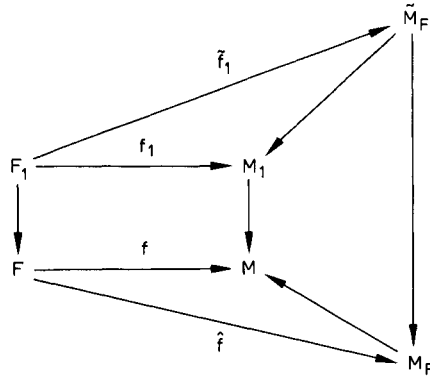


Fig. 3.4

$M_F$  is the covering space of  $M$  whose fundamental group corresponds to the map  $f$ , and similarly  $\tilde{M}_F$  is the covering space of  $M$  corresponding to  $f_1$ .

Clearly  $f$  is least area if and only if  $\tilde{f}$  is, and similarly for  $f_1$  and  $\tilde{f}_1$ . By Lemma 3.3,  $\tilde{f}$  is least area if and only if  $\tilde{f}_1$  is, as each induces a homotopy equivalence. Hence  $f$  is least area if and only if  $f_1$  is least area, as required.

Note that the above result does not assume that the maps  $f$  and  $f_1$  are embeddings.

**§ 4. Some 3-dimensional Topology**

In § 1 we discussed the importance of finding product regions between minimal surfaces. We will need to know when these exist even in the case where the minimal surfaces under consideration are not compact. The next lemma generalizes results of Waldhausen [23].

**Lemma 4.1.** *Let  $M$  be a  $P^2$ -irreducible 3-manifold without boundary and let  $F_1$  and  $F_2$  be incompressible, two-sided, possibly non-compact, embedded surfaces in  $M$  with no boundary. Suppose that  $F_1$  and  $F_2$  intersect transversely, that the natural map  $\pi_1(F_1) \rightarrow \pi_1(M)$  is an isomorphism and that  $F_1$  separates  $M$  into two components whose closures are  $A$  and  $B$ . If  $F_2 \cap A$  or  $F_2 \cap B$  has a non-empty compact component then there is a product region between  $F_1$  and  $F_2$  in  $M$ .*

*Proof.* First observe that the result is trivial if  $F_1$  or  $F_2$  is  $S^2$ . Note that in all cases  $F_1 \cap F_2$  consists of a disjoint collection of lines and circles.



*Case 1.* There is a circle  $C$  in  $F_1 \cap F_2$  which is null homotopic in  $M$ .

As  $F_1$  is incompressible,  $C$  must also be null-homotopic in  $F_1$ . Hence  $C$  bounds a 2-disk  $D_1$  in  $F_1$ . We choose  $C$  to be innermost in  $F_1$  so that  $D_1$  meets no other component of  $F_1 \cap F_2$ . As  $F_2$  is also incompressible,  $C$  bounds a 2-disk  $D_2$  in  $F_2$ . The disks together form a 2-sphere which is embedded because  $C$  was innermost. This sphere bounds a 3-ball  $X$  in  $M$  as  $M$  is irreducible and  $X$  is clearly a product region between  $F_1$  and  $F_2$ .

*Case 2.* There are no circles in  $F_1 \cap F_2$  which are null homotopic in  $M$ .

In this case we claim that each component of  $F_2 \cap A$  or  $F_2 \cap B$  is incompressible in  $A$  on  $B$ . Let  $S$  be a component of  $F_2 \cap A$ . If  $S$  is compressible in  $A$ , the loop theorem provides an embedded essential circle  $C$  in  $S$  which is null-homotopic in  $A$ . Hence  $C$  is null homotopic in  $M$  and must bound a 2-disk in  $F_2$ . As  $C$  is essential in  $S$ , the 2-disk must contain a boundary component  $C'$  of  $S$ . Then  $C'$  is a circle in  $F_1 \cap F_2$  which is null homotopic in  $M$ , contradicting our hypothesis.

Now suppose that  $S_2$  is a compact component of  $F_2 \cap A$ . We will show that there is a surface  $S_1$  in  $F_1$  and a product region  $X$  in  $M$  with  $\partial X = S_1 \cup S_2$ . In the following arguments  $Z_2$  coefficients are assumed for the homology groups.

As  $F_1$  is incompressible in  $M$ ,  $A$  is  $P^2$ -irreducible. Hence  $A$  is aspherical. It follows that the inclusion of  $F_1$  into  $A$  is a homotopy equivalence. In particular,  $H_2(A, F_1) = 0$ . Now  $S_2$  represents an element of  $H_2(A, F_1)$ . Hence there is a compact submanifold  $X$  of  $A$  and a compact subsurface  $S_1$  of  $F_1$  such that  $\partial X = S_1 \cup S_2$ . Let  $Y$  denote the closure of  $A - X$ . As  $A$  is not compact and  $X$  is compact, it follows that  $Y$  is not compact.

Let  $\tilde{M}$  denote the covering of  $M$  with  $\pi_1(\tilde{M}) = \pi_1(S_2)$  so that  $S_2$  lifts into  $\tilde{M}$ . Let  $\tilde{A}$  denote the pre-image of  $A$  and let  $\tilde{F}_1$  denote the pre-image of  $F_1$ . As above we have  $H_2(\tilde{A}, \tilde{F}_1) = 0$ . As  $S_2$  represents an element of this group we have a component  $\tilde{X}$  of  $\tilde{A} - S_2$  whose closure is compact. As  $\tilde{X}$  must be a covering of  $X$  or  $Y$  and as  $Y$  is not compact we see that  $\tilde{X}$  is a covering of  $X$ . Now  $\partial \tilde{X}$  consists of  $S_2$  and a subsurface of  $\tilde{F}_1$ . Hence the projection map  $\tilde{X} \rightarrow X$  is of degree one as it is of degree one restricted to the pre-image of  $S_2$ . It follows that the projection map is a homeomorphism. Now the equations

$$\pi_1(S_2) = \pi_1(\tilde{A}) = \pi_1(\tilde{X})^*_{\pi_1(S_2)} \pi_1(\tilde{Y})$$

show that  $\pi_1(\tilde{X}) = \pi_1(S_2)$ . The relative  $h$ -cobordism theorem [21] shows that the pair  $(\tilde{X}, S_2)$  is homeomorphic to the pair  $(S_2 \times I, S_2 \times \{0\})$ . Hence  $X$  is a product region between  $F_1$  and  $F_2$  in  $M$  as required and we have proved Lemma 4.1.

*Remark.* One could simplify the proof of the above Lemma if one knew that  $A$  were homeomorphic to  $F_1 \times [0, 1)$ . In that case it is clear that  $\pi_1(Y) = \pi_1(F_1)$  as  $Y$  contains  $F_1 \times \{K\}$  for suitably large  $K$ , so that the equation

$$\pi_1(F_1) = \pi_1(A) = \pi_1(X)^*_{\pi_1(S_2)} \pi_1(Y)$$

implies at once that  $\pi_1(X) = \pi_1(S_2)$  which is all that we need to show that  $X$  is a product region. In fact, in most of the situations which we will concern ourselves with,  $A$  will indeed be a product.

## §5. The Embedding Theorem

In this section we prove the main result of the paper which tells us that least area surfaces are embedded whenever possible. The precise result is the following.

**Theorem 5.1.** *Let  $M$  be a closed,  $P^2$ -irreducible Riemannian 3-manifold and let  $F$  be a closed surface not  $S^2$  or  $P^2$ . Let  $f: F \rightarrow M$  be a least area incompressible 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  $K$  embedded in  $M$  and  $g(F)$  bounds a submanifold of  $M$  which is a twisted  $I$ -bundle over a surface isotopic to  $K$ .*

*Proof.* Our method will show that  $f$  covers an embedded surface  $K$  in  $M$ . Clearly  $\pi_1(F)$  is a subgroup of  $\pi_1(K)$  up to conjugacy. Now it is proved in [9] or [19] that if  $F$  is a two-sided embedded surface in  $M$  and  $\pi_1(F) \subseteq \pi_1(K)$  then either  $\pi_1(F) = \pi_1(K)$  or  $\pi_1(F)$  has index two in  $\pi_1(K)$  and  $F$  bounds in  $M$  a twisted  $I$ -bundle over  $K$ . Hence knowing that  $f$  covers an embedded surface is enough to prove the theorem.

We will assume that  $f$  is self-transverse. At the end of the section we will explain how to obtain the general case.

We use some general facts about covering spaces of  $M$ . Consider an incompressible embedding  $g: F \rightarrow M$  homotopic to  $f$ . In a covering  $\tilde{M}$  of  $M$ , the pre-image of  $g(F)$  will consist of various disjoint, embedded, incompressible surfaces each of which is a covering of  $F$ . If we homotop  $g$  to the map  $f$ , the number and homeomorphism type of the surfaces in  $\tilde{M}$  will not alter, but each of the surfaces will be homotoped to a new possibly singular surface in  $\tilde{M}$ . It is also possible that distinct surfaces may meet each other or even become coincident. In this case these surfaces will not, strictly speaking, be the components of the pre-image of  $g(F)$ . However it will be convenient to refer to them as components. The self-intersections and mutual intersections of the components reflect the singularities of the map  $f$ . As  $f$  is self-transverse all these intersections are also transverse.

Now we consider some specific covering spaces of  $M$ . Let  $M_F$  denote the covering of  $M$  with  $\pi_1(M_F) = f_* \pi_1(F)$  and let  $f_F: F \rightarrow M_F$  denote the lift of  $f$ .  $f_F$  is a least area map because  $f$  is least area. Since  $M$  and therefore  $M_F$  are aspherical, it follows that  $f_F$  is a homotopy equivalence. Hence Theorem 3.2 tells us that  $f_F$  is an embedding. To simplify the notation we will denote the image of  $f_F$  by  $F$  also.

Let  $\tilde{M}$  denote the universal cover of  $M$  and let  $\tilde{F}$  denote the pre-image in  $\tilde{M}$  of  $F$  in  $M_F$ .  $\tilde{F}$  is homeomorphic to the universal cover of  $F$  and so is an embedded plane in  $\tilde{M}$ . Hence the pre-image in  $\tilde{M}$  of  $f(F)$  in  $M$  consists of a collection of embedded planes which are the translates of  $\tilde{F}$  under the action of  $\pi_1(M)$  on  $\tilde{M}$ . The stabilizer of  $\tilde{F}$  is a surface group which contains  $\pi_1(F)$  as a subgroup of index one or two, this number being two only if  $f$  factors through a 2-fold cover.

Showing that  $f$  covers an embedded surface is equivalent to showing that the distinct translates of  $\tilde{F}$  by  $\pi_1(M)$  are disjoint. Let us consider  $\tilde{F}$  and one of

its translates  $g\tilde{F}$  such that  $g\tilde{F}$  is distinct from  $\tilde{F}$ . The stabilizer of  $g\tilde{F}$  is the conjugate by  $g$  of the stabilizer of  $\tilde{F}$ . In particular, the stabilizer of  $g\tilde{F}$  contains  $g\pi_1(F)g^{-1}$ . Let  $G$  denote the subgroup  $\pi_1(F) \cap g\pi_1(F)g^{-1}$  of  $\pi_1(M)$ . Let  $M_G$  denote the cover of  $M$  with  $\pi_1(M_G) = G$  and let  $M_{gF}$  denote the cover with  $\pi_1(M_{gF}) = g\pi_1(F)g^{-1}$ .  $M_G$  covers each of  $M_F$  and  $M_{gF}$ . As  $g\tilde{F}$  is an embedded surface in  $M$  and as  $g\pi_1(F)g^{-1}$  leaves the surface invariant, the image of  $g\tilde{F}$  in  $M_{gF}$  must be an embedded surface homeomorphic to  $F$ . We shall abuse notation slightly and call this surface  $gF$ .

The images in  $M_G$  of  $\tilde{F}$  and  $g\tilde{F}$  are two embedded surfaces,  $F_1$  and  $F_2$ , each with fundamental group isomorphic to  $G$ . It follows that  $F_1$  separates  $M_G$  into two components, let  $A$  and  $B$  denote the closures of these two components.

We next show that either  $F_2 \cap A$  or  $F_2 \cap B$  is compact. To see this, recall that  $f$  is homotopic to an embedding  $g$ . This homotopy will induce homotopies of  $F_1$  and  $F_2$  in  $M_G$  so that  $F_1$  and  $F_2$  become disjoint. As  $F$  is compact, there is an upper bound  $d$  to the distance moved in  $M$  by any point during the homotopy of  $f$  to  $g$ . Hence in the induced homotopies of  $F_1$  and  $F_2$  in  $M_G$  no point of  $F_1$  or  $F_2$  moves a distance greater than  $d$ . Hence  $F_2 \cap A$  or  $F_2 \cap B$  must lie within a neighborhood  $W$  of  $F_1$  consisting of all points of distance less than  $2d$  away from  $F_1$ . Now consider the projection  $p: M_G \rightarrow M_F$ . The image of  $W$  in  $M_F$  must be compact.

Since  $g\pi_1(F)g^{-1}$  is a subgroup of index one or two in  $\text{Stabilizer}(g\tilde{F})$ , it follows that  $G = \pi_1(F) \cap g\pi_1(F)g^{-1}$  is also of index one or two in  $\pi_1(F) \cap \text{Stabilizer}(g\tilde{F})$ . Therefore  $p: F_2 \rightarrow M_F$  factors through a covering of degree at most two. It follows that  $p$  restricted to  $F_2$  is a proper map. The simplest way to see this is to triangulate  $F$  and  $M$  so that  $f$  is simplicial and triangulate all covering spaces in the natural way. Then  $p: F_2 \rightarrow M_F$  is at most 2-to-1 on 2-simplices. It follows that  $F_2 \cap A$  or  $F_2 \cap B$  must be compact as claimed.

Now Lemma 4.1 tells us that either  $F_1$  and  $F_2$  are disjoint or there exists a product region  $X$  between  $F_1$  and  $F_2$ . In the first case,  $\tilde{F}$  and  $g\tilde{F}$  must also be disjoint, as required. So we now assume that  $F_1$  and  $F_2$  intersect and will arrive at a contradiction. As in previous situations we denote the boundary of  $X$  by  $S_1 \cup S_2$  where  $S_i$  lies in  $F_i$ .

If both  $F_1$  and  $F_2$  were closed surfaces we could use the area swapping arguments of Lemma 1.2 to get a contradiction, but neither one need be compact. If  $X$  were to project injectively into  $M_F$  and  $M_{gF}$  we could still make this argument work by comparing areas of  $F - S_1 + S_2$  in  $M_F$  and  $gF - S_2 + S_1$  in  $M_{gF}$ . Unfortunately there is no apparent way to show that  $X$  injects into these spaces. However the situation is rescued by injecting  $X$  into an intermediate space. For this purpose we need some special properties of surface groups. The main theorem of [20] states that surface groups are subgroup separable, or LERF. We use this result in the form of Lemma 1.4 of [20].

**Lemma 5.2.** *Let  $Y$  be a Hausdorff topological space with a regular covering  $\tilde{Y}$  and covering group  $\Gamma$ . Then  $\Gamma$  is subgroup separable if and only if given a finitely generated subgroup  $G$  of  $\Gamma$  and a compact subset  $X$  of  $\tilde{Y}/G$ , there is a finite cover  $Y'$  of  $Y$  such that the projection  $\tilde{Y}/G \rightarrow Y$  factors through  $Y'$  and  $X$  projects injectively into  $Y'$ .*

We will apply this to our situation by taking  $Y = M_F$ ,  $\tilde{Y} = \tilde{M}$ ,  $\Gamma = \pi_1(F)$  and, of course,  $\tilde{Y}/G = M_G$ . It would seem that we need to show that  $G$  is finitely generated. This is true for the above situation, but this fact will not be needed. For if  $G$  is not finitely generated, we can find a finitely generated subgroup  $H$  of  $G$  such that our product region  $X$  in  $M_G$  lifts to  $M_H$ . We do this by simply choosing  $H$  equal to  $\pi_1(X)$ . We can then replace  $G$  by  $H$  in what follows. Hence we will apply Lemma 5.2 assuming that  $G$  is finitely generated.

As  $\pi_1(F)$  is subgroup separable, there is a finite covering  $M'$  of  $M_F$  such that the projection  $M_G \rightarrow M_F$  factors through  $M'$  and  $X$  projects injectively into  $M'$ . Similarly there exists a finite cover  $M''$  of  $M$ , such that the projection  $M_G \rightarrow M_{gF}$  factors through  $M''$  and  $X$  projects injectively into  $M''$ . Let  $F'$  denote the pre-image in  $M'$  of  $F$  in  $M_F$  and let  $F''$  denote the pre-image in  $M''$  of  $gF$  in  $M_{gF}$ . Each of  $F'$  and  $F''$  is an embedded closed surface. Lemma 3.3 tells us that both are least area. We now use an area swapping argument to get a contradiction, just as in Lemma 1.2, using the (possibly singular) surfaces  $F' - S_1 + S_2$  in  $M'$  and  $F'' - S_2 + S_1$  in  $M''$ . This completes the proof of Theorem 5.1 when  $f$  is self-transverse.

To handle the case when  $f$  is not self-transverse we repeat the above argument and use Lemma 1.6. In finding a product region  $X$  in  $M_G$  we used Lemma 4.1 which assumed transverse intersection of the surfaces  $F_1$  and  $F_2$ . In general this need not be true. See §1. Lemma 1.6 allows us to find a family of maps  $f_t$  homotopic to  $f$  and identical to  $f$  outside a neighborhood of the non-transersal points of  $f$  so that  $f_t$  satisfies the conditions given in the lemma. Going through the steps of the proof of Theorem 5.1, but using  $f_t$  instead of  $f$ , we eventually wind up with a product region  $X_t$  in  $M_G$ . We now apply the subgroup separability of  $\pi_1(F)$  using a neighborhood of  $X_t$  such that the corresponding product region  $X_t$  obtained from  $f_t$  lies in the neighborhood, for all  $t > 0$ . Then there is a  $k$ -fold covering  $M'$  of  $M_F$  and an  $l$ -fold covering  $M''$  of  $M_{gF}$  such that the neighborhood of  $X_t$  projects injectively into  $M'$  and  $M''$ . For each  $t$ ,  $0 < t \leq 1$ ,  $X_t$  projects injectively into  $M'$  and  $M''$ . Recall that  $F'$  denotes the pre-image in  $M'$  of  $F$  in  $M_F$  and  $F''$  denotes the pre-image in  $M''$  of  $gF$  in  $M_{gF}$ . When  $t$  is non-zero, we denote the corresponding surfaces by  $F'(t)$  and  $F''(t)$ . We have the product region  $X_t$  and can consider the new surfaces  $F'(t) - S_1(t) + S_2(t)$  in  $M'$  and  $F''(t) - S_2(t) + S_1(t)$  in  $M''$ . Now  $F'(t)$  and  $F''(t)$  do not alter with  $t$  except in small neighbourhoods of the non-transversal points. Hence there is a positive constant  $\varepsilon$ , independent of  $t$ , such that, after rounding corners, our area swap achieves a reduction of at least  $\varepsilon$  in  $\text{Area } F'(t) + \text{Area } F''(t)$ .

Let  $A(t) = \text{Area } f_t - \text{Area } f$ . It follows that

$$\text{Area } F'(t) - \text{Area } F' = k \cdot A(t) \quad \text{and} \quad \text{Area } F''(t) - \text{Area } F'' = l \cdot A(t).$$

Hence

$$\text{Area } F'(t) + \text{Area } F''(t) = \text{Area } F' + \text{Area } F'' + (k + l) \cdot A(t).$$

If we choose  $t$  so small that  $(k + l) \cdot A(t) < \varepsilon$ , we obtain a contradiction to our assumption that  $F'$  and  $F''$  are least area.

**Theorem 5.3.** *Let  $F$  be a closed surface, not  $S^2$  or  $P^2$  and let  $M$  be a  $P^2$ -irreducible Riemannian 3-manifold. Let  $f: F \rightarrow M$  be a two-sided least area immersion which is homotopic to a map  $g$  which is a composition of two maps  $g_1: F \rightarrow F'$  and  $g_2: F' \rightarrow M$ , where  $g_1$  is a covering map of degree  $k$  and  $g_2$  is a two-sided embedding. Then  $f$  is a composition of two maps  $f_1: F \rightarrow F''$  and  $f_2: F'' \rightarrow M$ , where  $f_2$  is an embedding and  $f_1$  is a covering map of degree  $k$  or  $2k$  and if the latter holds then  $f_2$  is a one sided embedding.*

*Proof.* The map  $g: F \rightarrow M$  has no transverse self-intersections. The proof of Theorem 5.1 applies to show that  $f$  also has no transverse self-intersections, so that  $f$  covers an embedded surface  $F''$  in  $M$ . We claim that  $F'$  is homotopic to  $F''$  or to a double cover of  $F''$  which will complete the proof. To see this, consider the covering  $M''$  of  $M$  with  $\pi_1(M'') = \pi_1(F'')$ . Clearly  $f: F \rightarrow M$  lifts to  $f'': F \rightarrow M''$  and  $f''$  covers an embedded copy of  $F''$  in  $M''$ . Hence  $g: F \rightarrow M$  also lifts to  $M''$  and  $g$  must cover an embedded surface  $L$  in  $M''$ , where  $L$  is some covering of  $F'$  and hence is two-sided. But the embedded surface  $L$  can only be homotopic to  $F''$  or to a double covering of  $F''$  [9, 19], and in the second case  $F''$  must be one-sided in  $M''$ . It follows that if  $F''$  is two-sided in  $M$ , then the projection of  $L$  into  $M$  is homotopic to  $F''$  and if  $F''$  is one-sided in  $M$  then the projection of  $L$  into  $M$  is homotopic to a two-sided embedded surface forming the boundary of a regular neighborhood of  $F''$ . In either case projection of  $L$  into  $M$  is homotopic to a two-sided embedding. As  $\pi_1(L) \subset \pi_1(F')$  and as  $F'$  is two-sided, it follows that  $\pi_1(L) = \pi_1(F)$  [9, 19] which establishes the claim.

**Theorem 5.4.** *Let  $F$  be a closed surface, not  $S^2$  or  $P^2$  and let  $M$  be a  $P^2$ -irreducible Riemannian 3-manifold. Let  $f: F \rightarrow M$  be a two-sided least area immersion which is homotopic to a map  $g$  which is a composition of two maps  $g_1: F \rightarrow F'$  and  $g_2: F' \rightarrow M$ , where  $g_1$  is a covering map of degree  $k$ ,  $g_2$  is two-sided and  $k$  is maximal. Then  $f$  is a composition of two maps  $f_1: F \rightarrow F''$  and  $f_2: F'' \rightarrow M$ , where  $f_1$  is a covering map of degree  $k$  or  $2k$  and if the latter holds then  $f_2$  is a one-sided mapping.*

*Proof.* Consider the space  $M_{F'}$  which is finitely covered by  $M_F$ .  $f$  lifts to  $M_{F'}$  where it is homotopic to a  $k$ -fold cover of the lift of  $g_2$ . The map  $g_2$  is a homotopy equivalence into  $M_{F'}$  and therefore homotopic to an embedding. Theorem 5.3 now gives that the lift of  $f$  to  $M_{F'}$  is either a  $k$ -fold or  $2k$ -fold cover of an embedding. The result follows by projecting this map to  $M$ .

**§6. Intersections of Least Area Surfaces**

Let  $M$  be a compact,  $P^2$ -irreducible Riemannian 3-manifold. Let  $F_1$  and  $F_2$  be closed surfaces, not  $S^2$  or  $P^2$  and let  $f_i: F_i \rightarrow M$  be a two-sided least area incompressible map,  $i=1,2$ . We shall examine the self-intersections of  $f_1$  and the intersection of  $f_1$  with  $f_2$ .

Pick a subgroup  $\pi_1(F_i)$  of  $\pi_1(M)$  from the conjugacy class of subgroups determined by  $f_i$ . Let  $M_i$  denote the covering of  $M$  with group  $\pi_1(F_i)$  and let  $F_i$  denote the image in  $M_i$  of the lift of  $f_i$ . Note that this image is an embedded two-sided copy of  $F_i$ , by Theorem 3.4. Let  $\tilde{M}$  denote the universal covering space of  $M$  and let  $\tilde{F}_i$  denote the pre-image in  $\tilde{M}$  of  $F_i$  in  $M_i$ . As  $F_i$  is an

embedded surface in  $M_i$ ,  $\tilde{F}_i$  is an embedded plane. The full pre-image of  $f_i(F_i)$  in  $\tilde{M}$  consists of embedded planes which are the translates of  $\tilde{F}_i$  by  $\pi_1(M)$ . The stabilizer of  $\tilde{F}_i$  is a surface group which contains  $\pi_1(F_i)$  and so the stabilizer of  $g\tilde{F}_i$  contains  $g\pi_1(F_i)g^{-1}$ . Let  $M_{g,i}$  denote the cover of  $M$  with  $\pi_1(M_{g,i}) = g\pi_1(F_i)g^{-1}$  and let  $g\tilde{F}_i$  denote the image of  $\tilde{F}_i$  in  $M_{g,i}$ . Let  $G = \pi_1(F_i) \cap g\pi_1(F_j)g^{-1}$  and let  $M_G$  denote the corresponding cover of  $M$ . Let  $L_i$  and  $gL_j$  denote the images in  $M_G$  of  $\tilde{F}_i$  and  $g\tilde{F}_j$ .  $L_i$  and  $gL_j$  are embedded surfaces.

**Lemma 6.1.** *Suppose that  $L_i$  meets  $gL_j$  transversely. Then there can be no product regions between  $L_i$  and  $gL_j$ .*

*Proof.* If there is a product region  $X$  in  $M_G$  between  $L_i$  and  $gL_j$ , we can use the argument at the end of the proof of Theorem 5.1 to get a contradiction. As in that proof, we can assume that  $G$  is finitely generated. Then the subgroup separability of  $\pi_1(F_i)$  and  $\pi_1(F_j)$  implies that  $X$  projects injectively into suitable finite coverings of  $M_i$  and  $M_{g,j}$ . This then gives a contradiction by the area swapping argument.

Our first result on intersections of least area surfaces deals with the case when the expected intersection is empty.

**Theorem 6.2.** *Let  $M$  be a compact,  $P^2$ -irreducible, Riemannian 3-manifold. Let  $F_1$  and  $F_2$  be closed surfaces, not  $S^2$  or  $P^2$  and let  $f_i: F_i \rightarrow M$  be a two-sided, least area, incompressible (possibly singular) map,  $i=1,2$ . 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 are equal. In the second case,  $f_1$  and  $f_2$  are coverings of an embedded surface in  $M$ .*

*Proof.* As usual it suffices to consider the case when  $f_1$  and  $f_2$  intersect transversely. The surfaces  $L_1$  and  $gL_2$  in  $M_G$  intersect transversely or trivially. There is a homotopy of the original maps  $f_1$  and  $f_2$  to maps with disjoint images. It follows as in the proof of Theorem 5.1 that either  $L_1$  and  $gL_2$  are disjoint or there is a product region between  $L_1$  and  $gL_2$  in  $M_G$ .

Lemma 6.1 shows that a product region can not exist so that  $L_1$  and  $gL_2$  are disjoint or identical. As this result holds for any choice of  $g$  in  $\pi_1(M)$ , it follows that  $f_1(F_1)$  and  $f_2(F_2)$  are disjoint or equal, as claimed.

Now we consider further the case when  $f_1(F_1) = f_2(F_2)$ . We choose base-points and identify  $\pi_1(F_1)$  and  $\pi_1(F_2)$  with subgroups of  $\pi_1(M)$ . The common image of  $f_1$  and  $f_2$  is a singular surface  $K$  and  $\pi_1(F_1)$  and  $\pi_1(F_2)$  are subgroups of  $\pi_1(K)$  of finite index. Hence  $\pi_1(F_1) \cap \pi_1(F_2)$  is also a subgroup of  $\pi_1(K)$  of finite index. Let  $F_3$  be the corresponding covering of  $K$  and let  $f_3: F_3 \rightarrow M$  be the composite of this covering with the immersion of  $K$  in  $M$ . This is a least area map by Theorem 3.4. The hypothesis that  $f_1$  and  $f_2$  can be homotoped apart implies that  $f_3$  can be homotoped to maps  $f'_3$  and  $f''_3$  which are disjoint. The methods of Theorem 5.1 can now be applied to show that  $f_3$  can have no self-intersections, so that  $f_3$  covers an embedded surface  $K$  in  $M$ . It follows that  $f_1$  and  $f_2$  both cover the same surface  $K$ , which completes the proof of Theorem 6.2.

We now consider the situation where  $f_1$  and  $f_2$  cannot be homotoped apart. At the same time, we will consider the self-intersections of a single least

area map  $f_1$ . We will show that least area surfaces intersect as little as possible as stated in the introduction. First we need to discuss how to measure the intersections of singular surfaces. We will consider the case of a single least area map  $f_1$  but the case of two maps is similar.

Note that any least area map  $F \rightarrow M$  is a composite  $F \rightarrow F' \rightarrow M$  where  $F \rightarrow F'$  is a covering and  $F' \rightarrow M$  is an immersion which is self-transverse except at isolated points. It will be convenient to assume that  $f_1: F_1 \rightarrow M$  does not factor through a non-trivial covering of surfaces. If  $f_1$  is in general position, a natural measure of its self-intersections is the number of double curves  $d$  in its image, or the number of double curves in the domain  $F_1$  which will equal  $2d$  as  $f_1$  is two-sided. However the example we give in § 8 shows that least area maps do not necessarily have the least possible number of double curves, so this measure is not suitable for our purposes. A further problem with this measure is that it needs care when  $f_1$  is locally self-transverse, but not in general position, as  $f_1$  could have curves of triple points. Finally, if  $f_1$  is not locally self-transverse one cannot measure its self-intersections at all.

The following seems to be the best way to measure the self-intersections of  $f_1$ . We consider the lift of  $F_1$  into  $M_1$  and count the number of components of the pre-image in  $M_1$  of  $f_1(F_1)$  which meet  $F_1$ . This number is denoted  $D(f_1)$ . Note that if  $f_1$  is in general position and if each component of the pre-image of  $f_1(F_1)$  meets  $F_1$  in a single curve, then  $D(f_1)$  is the same as the number of double curves  $2d$  of  $f_1$ . But, in general,  $D(f_1)$  is less than  $2d$ . This definition applies even when  $f_1$  factors through some covering of surfaces.

When we have two least area immersion  $f_i: F_i \rightarrow M$ , we define  $D(f_1, f_2)$  to be the number of components of the pre-image in  $M_1$  of  $f_2(F_2)$  which meet  $F_1$ . This number is symmetric i.e.  $D(f_1, f_2) = D(f_2, f_1)$ , because one can interpret both numbers in  $\tilde{M}$ . Let  $X_{12} = \{g \in \pi_1(M) : g\tilde{F}_2 \text{ meets } \tilde{F}_1\}$ . Clearly  $X_{12}$  is a union of double cosets  $\pi_1(F_1)g\pi_1(F_2)$  and  $D(f_1, f_2)$  is the number of these double cosets. Similarly  $D(f_2, f_1)$  counts the number of double cosets  $\pi_1(F_2)h\pi_1(F_1)$  in  $X_{21}$ . But the map  $g \rightarrow g^{-1}$  defines a natural bijection between  $X_{12}$  and  $X_{21}$  and induces a bijection of double cosets, showing that  $D(f_1, f_2) = D(f_2, f_1)$ .

**Theorem 6.3.** *Let  $M$  be a compact,  $P^2$ -irreducible, Riemannian 3-manifold. Let  $F_i$  be a closed surface, not  $S^2$  or  $P^2$ , let  $f'_i: F_i \rightarrow M$  be a two-sided, incompressible map, and let  $f_i$  be a 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)  $D(f_i) \leq D(f'_i)$ .

*Proof.* We will prove (i) only, as the proof of (ii) is similar. As usual, let  $F_1$  denote the lift of  $f_1$  to  $M_1$  and let  $\tilde{F}_1$  denote the pre-image in  $\tilde{M}$  of  $F_1$ . If (i) is false, there must be a homotopy of  $f_1$  and  $f_2$  which removes the intersection of some component of the pre-image of  $f_2(F_2)$  with  $F_1$ . In  $\tilde{M}$ , we find a translate  $g\tilde{F}_2$  of  $\tilde{F}_2$  whose intersection with  $\tilde{F}_1$  is removed by this homotopy. Let  $G = \pi_1(F_1) \cap g\pi_1(F_2)g^{-1}$ , and let  $L_1$  and  $gL_2$  denote the images in  $M_G$  of  $\tilde{F}_1$  and  $g\tilde{F}_2$ . The methods of § 5 produce a product region  $X$  in  $M_G$  between  $L_1$  and  $gL_2$ , and this product region can be injected into finite covers of  $M_1$  and  $M_{g,2}$ . This leads to a contradiction as before.

The above result shows that least area surfaces intersect as little as possible, but gives us no help in determining the size of the intersection in examples. The following result gives some information which is relevant to all examples. We continue using the notation established at the start of this section. It is most convenient to formulate this result in terms of intersections in  $\tilde{M}$  rather than in  $M_1$ .

**Lemma 6.4.** (i) *If  $\pi_1(F_i) \cap g\pi_1(F_j)g^{-1} = 1$ , then  $\tilde{F}_i \cap g\tilde{F}_j$  is empty.*

(ii) *If  $\pi_1(F_i) \cap g\pi_1(F_j)g^{-1}$  is a closed infinite surface group then  $\tilde{F}_i$  and  $g\tilde{F}_j$  are disjoint or equal.*

*Remark.* This gives us information about the intersection of two least area maps, and about the self-intersections of a single map.

*Proof.* (i) Let  $\tilde{A}$  and  $\tilde{B}$  denote the closures of the two components of  $\tilde{M} - \tilde{F}_i$ . Our hypothesis implies that the projection  $\tilde{M} \rightarrow M_i$  when restricted to  $g\tilde{F}_j$  is a proper map. As  $F_i$  is compact, it follows that  $\tilde{F}_i \cap g\tilde{F}_j$  is compact. But  $g\tilde{F}_j$  is a plane and so has only one end. Hence one of  $g\tilde{F}_j \cap \tilde{A}$  or  $g\tilde{F}_j \cap \tilde{B}$  is compact.

The surfaces  $\tilde{F}_i$  and  $\tilde{F}_j$  cannot be equal as their stabilizers have trivial intersection. As usual, it suffices to consider the case when all intersections are transverse so that  $\tilde{F}_i$  is transverse to  $g\tilde{F}_j$ . As one of  $g\tilde{F}_j \cap \tilde{A}$  and  $g\tilde{F}_j \cap \tilde{B}$  is compact Lemma 4.1 implies that either  $\tilde{F}_i$  and  $g\tilde{F}_j$  are disjoint or there is a product region between them. Lemma 6.1 shows that a product region cannot occur and so they are disjoint as required.

(ii) Let  $G = \pi_1(F_i) \cap g\pi_1(F_j)g^{-1}$ . This is a closed surface group by assumption, hence the surfaces  $L_i$  and  $gL_j$  in  $M_G$  are closed surfaces. Now each is a least area surface by Theorem 3.5, hence they are disjoint or equal by Lemma 1.3. Hence  $\tilde{F}_i$  and  $g\tilde{F}_j$  are also disjoint or equal.

It follows at once from the above result that if  $f_1: F_1 \rightarrow M$  is a least area map and if  $\pi_1(F_1) \cap g\pi_1(F_1)g^{-1}$  is always trivial or a closed surface group then  $f_1$  covers an embedding.

The next special case we consider is when  $G = \pi_1(F_i) \cap g\pi_1(F_j)g^{-1}$  is infinite cyclic. In this case each of the surfaces  $L_i$  and  $gL_j$  in  $M_G$  must be an annulus or Moebius band, and can certainly give rise to double curves unlike the previous cases. However, we can prove the following result.

**Lemma 6.5.** *If  $G = \pi_1(F_i) \cap g\pi_1(F_j)g^{-1}$  is infinite cyclic, then either  $\tilde{F}_i$  and  $g\tilde{F}_j$  are disjoint or they intersect transversely in a single line whose stabiliser contains  $G$ .*

Before proving this result, we will discuss some of its consequences.

If  $F_i$  is a torus or Klein bottle then every subgroup of  $\pi_1(F_i)$  is trivial, infinite cyclic or the fundamental group of a closed surface, namely the torus or Klein bottle. This observation combined with the two preceding lemmas at once gives us the following surprising result.

**Theorem 6.6.** *Let  $M$  be a closed,  $P^2$ -irreducible, Riemannian 3-manifold. Let  $F_1$  be a torus or Klein bottle and let  $F_2$  be a closed surface, not  $S^2$  or  $P^2$ . For  $i = 1, 2$ , let  $f_i: F_i \rightarrow M$  be a two-sided, least area, incompressible map which does not factor through a covering of surfaces. Then*



- (i)  $f_1$  is locally transverse to  $f_2$ ,
- (ii)  $f_1$  is locally self-transverse.

Lemma 6.5 also allows us to refine Theorem 6.3 and show that least area tori are not only self-transverse but also have the least possible number of double curves. The precise results is the following.

If  $f_1$  and  $f_2$  are two immersions in general position, we define  $d(f_1, f_2)$  to be the number of double curves in their intersection and define  $d(f_1)$  to be the number of double curves in the self-intersection of  $f_1$ . If the maps are only locally transverse we count curves of multiplicity  $k$  with multiplicity  $(k - 1)$ . Note that this count is to be made in the source  $F_1$  of  $f_1$  and not in  $M$ .

**Theorem 6.7.** *Let  $M$  be a closed,  $P^2$ -irreducible, Riemannian 3-manifold. Let  $F_1$  be a torus or Klein bottle and let  $F_2$  be a closed surface, not  $S^2$  or  $P^2$ . Let  $f'_i: F_i \rightarrow M$  be a two-sided, incompressible map in general position, and let  $f_i$  be a 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)  $d(f_1) \leq d(f'_1)$ .

*Proof.* Theorem 6.6 ensures that  $d(f_1)$  and  $d(f_1, f_2)$  are defined. Now Lemmas 6.4 and 6.5 show that  $d(f_i)$  equals  $D(f_i)$  and  $d(f_1, f_2)$  equals  $D(f_1, f_2)$ . Theorem 6.3 then proves the required inequalities.

Now we return to the proof of Lemma 6.5, which we restate for the convenience of the reader.

**Lemma 6.5.** *If  $G = \pi_1(F_i) \cap g\pi_1(F_j)g^{-1}$  is infinite cyclic, then either  $\tilde{F}_i$  and  $g\tilde{F}_j$  are disjoint or they intersect transversely in a single line whose stabiliser contains  $G$ .*

*Proof.* We start by considering the case when all intersections are assumed to be transverse.

As usual we consider the surfaces  $L_i$  and  $gL_j$  in  $M_G$ . Each has a cyclic fundamental group and so is isomorphic to an annulus or Moebius band. If  $L_i$  or  $gL_j$  is a Moebius band then the intersection is empty, as an open Moebius band has only one end and the method of Lemma 6.4(i) applies. If there is an inessential circle in  $L_i \cap gL_j$ , there will be a product region in  $M_G$  between  $L_i$  and  $gL_j$  by the proof of Lemma 4.1. If there is more than one circle in  $gL_j \cap L_i$  there will be a compact annulus in  $gL_j$  cut off by two adjacent circles. Lemma 4.1 shows that there would then be a product region in  $M_G$  between  $L_i$  and  $gL_j$ . We conclude that  $gL_j \cap L_i$  consists of a single essential circle if it is not empty and the conclusion follows.

Now we consider the general situation. We will discuss the case of two maps, taking  $i = 1, j = 2$ . The case when  $i = j$  is proved in the same way.

Recall that  $f_1$  and  $f_2$  are transverse except at isolated points and that at these points the intersection looks like the intersection of the graph of  $\text{Re}(z^n)$  with the  $z$ -plane. The same remarks apply to the intersection of  $L_1$  and  $gL_2$  in  $M_G$ . Suppose that  $L_1$  and  $gL_2$  intersect non-transversely at some point  $Q$ . As they intersect, each is an annulus. The intersection is a finite graph, with at least one vertex, the one at  $Q$ , having  $2n$  prongs coming into it, with  $n \geq 2$ . Lemma 1.6 gives us a small perturbation of  $f_1$  to a map  $f'_1$  transverse to  $f_2$ . This will lift to a small perturbation of  $L_1$  to an annulus  $L'_1$  transverse to  $gL_2$ . The picture for the intersection near  $Q$  is shown below.

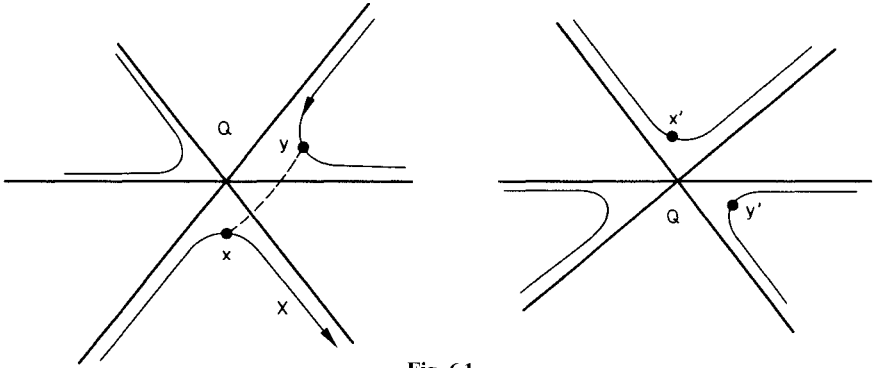


Fig. 6.1

The homotopy of  $f_1$  described in Lemma 1.6 was fixed outside a small neighbourhood of the points of non-transverse intersection. It will be convenient here to make our homotopy move all of  $f_1$ . Consider the lift of  $f_1$  to an embedding of  $F_1$  in  $M_1$ . As  $F_1$  is two-sided in  $M_1$ , we have two choices of an isotopy of  $F_1$  in  $M_1$  pushing  $F_1$  along the normals in  $M_1$ . This gives us two homotopies of  $f_1$  by projection into  $M$ . Let  $\Gamma_1$  and  $\Gamma_2$  be the two new intersections of  $L_1$  with  $gL_2$ . Each will be a collection of disjoint simple closed curves as  $L_1$  and  $gL_2$  are transverse. Note that  $\Gamma_1$  and  $\Gamma_2$  are disjoint from each other and from  $\Gamma$  by construction.

The proof of Theorem 5.1 shows that if we make our perturbation of  $f_1$  small enough, we can treat  $f'_1$  as if it is still a least area map. In particular, we see that both  $\Gamma_1$  and  $\Gamma_2$  must be connected, by the first part of the proof of this lemma. We can now derive a contradiction by considering Fig. 6.1. Start at the point  $x$  of  $\Gamma_1$  shown in the figure and run along the arc  $X$ . One must eventually return to the point  $y$ , as  $\Gamma_1$  is connected. We can construct a new simple closed curve by taking this arc together with a short arc joining  $x$  to  $y$  as in the figure. Clearly this new curve does not meet  $\Gamma_2$ . But the points  $x'$  and  $y'$  of  $\Gamma_2$  are separated by this new curve as every simple closed curve on an annulus separates. It follows that  $\Gamma_2$  is disconnected. This contradiction completes the proof of Lemma 6.5.

## §7. Extensions

As noted in the introduction, the assumption that the 3-manifold has no boundary can be replaced with the assumption that it has a smooth boundary whose mean curvature vector is zero or inward pointing. In fact even the assumption of smoothness can be weakened to allow a piecewise smooth boundary as long as it is suitably behaved. The precise conditions are given in [15] and are:

$M$  is a compact subdomain of a smooth triangulated manifold  $M'$  such that  $\partial M$  is a two dimensional subcomplex of  $M'$  consisting of surfaces  $\{H_1, \dots, H_l\}$  which have the following properties:

(1) Each  $H_i$  is a  $C^2$  surface in  $M'$  whose mean curvature is non-negative with respect to the inward normal.

(2) Each surface  $H_i$  is a compact subset of some surfaces  $H'_i$  in  $M'$  where  $H'_i \cap M = H_i$  and  $\partial H'_i$  lies on  $\partial M'$ .

We will henceforth take the above to be the meaning of a 3-manifold having positive mean curvature on the boundary.

The assumption that the 3-manifold had no boundary was not used in any of the 3-manifold arguments of §1 through §6. To extend our results, it is only necessary to know that the Schoen-Yau existence result extends to the case of a 3-manifold with boundary of the above type. This is proved in [15].

It is also possible to weaken the irreducibility condition on the 3-manifold. Let  $f: F \rightarrow M$  be an incompressible two-sided map into a reducible manifold. The Schoen-Yau existence theorem proves the existence of an area minimizing map among all maps inducing a conjugate action on the fundamental group of  $F$ . If we minimize in this sort of class rather than in a homotopy class of maps, we can obtain analogous embedding results. For example, the statement of Theorem 5.1 in this context would be;

**Theorem 7.1.** *Let  $M$  be a closed Riemannian 3-manifold and let  $F$  be a closed surface not  $S^2$  or  $P^2$ . Let  $f: F \rightarrow M$  be a map inducing an injection of the fundamental group. Assume there exists a two-sided embedding  $g: F \rightarrow M$  such that  $f_*(\pi_1(M)) = \gamma g_*(\pi_1(M)) \gamma^{-1}$  where  $\gamma$  is some path in  $M$ , and assume that  $f$  minimizes area among all such maps. Then either*

- (i)  $f$  is an embedding, or
- (ii)  $f$  double covers a one-sided surface embedded in  $M$ .

For the rest of this section, we will consider the question of least area surfaces with boundary. For this discussion it will be convenient to assume that the 3-manifolds involved are irreducible, but the results can all be extended to the reducible case as for closed surfaces. When one considers surfaces with boundary, there are two natural ideas of a least area map. In the first case which we call the fixed boundary case, one considers a map  $g: (F, \partial F) \rightarrow (M, \partial M)$  and the class of all maps which are homotopic to  $g$  rel  $\partial F$ . We will restrict our discussion of this case to the situation where  $g$  embeds  $\partial F$  in  $\partial M$ . In the second case, which we call the free boundary case, one considers the class of all maps of  $F$  into  $M$  which are properly homotopic to  $g$ . The following result is the existence theorem in the fixed boundary case. It is due to Lemaire [10].

**Theorem 7.2.** *Let  $f: (F, \partial F) \rightarrow (M, \partial M)$  be a proper incompressible map of a surface with boundary into a  $P^2$ -irreducible Riemannian 3-manifold whose boundary has non-negative mean curvature such that  $f(\partial F)$  is a collection of disjoint simple closed curves. Then there exists a least area map  $g$  which is homotopic to  $f$  (by a homotopy fixing the image of the boundary) and  $g$  can be parameterized as an smooth immersion.*

The next existence statement is for the free boundary case. One clearly needs to assume a boundary incompressibility condition to avoid having the

minimizing surface degenerate. Define a mapping  $f(F, \partial F) \rightarrow (M, \partial M)$  to be boundary incompressible if the natural map  $\pi_1(F, \partial F) \rightarrow \pi_1(M, \partial M)$  is injective.

Presumably, the statement we give below can be proved by combining the methods of Schoen and Yau [18] with the methods of Meeks and Yau [15], but we have not done this. However, it seems worth starting the existence result which appears to be needed in order to apply our results.

**Existence Statement 7.3.** *Let  $f: (F, \partial F) \rightarrow (M, \partial M)$  be a proper map of a surface  $F$  into a  $P^2$ -irreducible 3-manifold whose boundary has non-negative mean curvature. Assume that  $f$  is incompressible and boundary incompressible. Then there exists a least area map  $g$  among all maps which are homotopic to  $f$  by a proper homotopy, and  $g$  can be parameterized as a smooth immersion.*

On the assumption that this statement holds, one can now work through the first six sections of this paper and prove all our results in both cases. Of course, some minor modifications are needed in the proofs and in the statements of some results, and we briefly discuss these.

In §1, we introduced the idea of a product region and Lemma 1.2 showed that such regions cannot exist under certain circumstances. In the fixed boundary case, no changes are needed but in the free boundary case, we need to generalise our definition of product region to cover the case where the region meets the boundary of the 3-manifold. In this case, the boundary of the product region will consist of the surfaces  $S_1$  and  $S_2$  in  $F_1$  and  $F_2$  respectively together with a subsurface of  $\partial M$  which is a product region between  $\partial F_1$  and  $\partial F_2$ . Lemma 1.2 needs no changes to its proof. Lemmas 1.3, 1.4, 1.5 and 1.6 also need essentially no change, but one needs to note that a least area immersion can fail to be transverse on the boundary of the surface.

In §2, we proved our embedding result for homotopy equivalences. The proofs in the fixed and free boundary cases are similar to the closed case with relative homology groups used instead of absolute homology groups. Note that in the fixed boundary case the phenomenon of double covering a one-sided surface is ruled out by the condition that the least area map be an embedding on its boundary. Note also that in the free boundary case, one needs to assume that the given homotopy equivalence is properly homotopic to an embedding, as this is not automatic as it was in the closed case.

§3 needs only minor alterations to take account of the preceding remarks, and §4 also goes through essentially unchanged, except that one needs to consider arcs in  $F_1 \cap F_2$  as well as circles. Using these revised versions of Sects. 1–4, the proof of the main embedding result in §5 needs only trivial alterations.

Finally, all the results in §6 apply in both the fixed and free boundary cases, except for the following points. Lemma 6.4(i) is false as the universal cover of a surface with boundary cannot have one end. Also Lemma 6.4(ii) is irrelevant. The conclusion of Lemma 6.5 needs to include the new possibility that  $F_i$  and  $gF_j$  could intersect in a single arc properly embedded in each. This forces some changes to the proof of the lemma but the ideas are the same. (This possibility can only arise in the free boundary case.) Lemma 6.5 was used to show that any least area torus or Klein bottle is self-transverse and has the

minimal number of double curves. The crucial property of these surfaces was that any subgroup of their fundamental groups is trivial, infinite cyclic or the fundamental group of a closed surface. In the bounded cases, we need to add the annulus and Moebius bands to this list of surfaces which are automatically transverse.

There is an interesting application of these results to hierarchies of Haken 3-manifolds. Let  $M$  be a Riemannian, Haken manifold whose boundary is empty or has non-negative mean curvature. Then any hierarchy for  $M$  is isotopic to one consisting of least area surfaces in the fixed boundary sense. This is because a least area surface is automatically minimal and so has mean curvature zero. Hence cutting  $M$  along such a surface preserves the property that  $\partial M$  has non-negative mean curvature. If each surface in the hierarchy is boundary incompressible, the least area surfaces can be chosen to be least area in their free homotopy classes.

## §8. Two Examples

Our first example is a very simple one of two embedded least area surfaces with non-transverse intersection. Recall that this is impossible if either surface is a torus or Klein bottle, by Theorem 6.6. However, our construction can give examples for any surfaces of higher genus.

Let  $F$  be a closed surface with negative Euler characteristic and some metric, and let  $M$  denote  $F \times S^1$  with the product metric. It is well known that  $M$  contains many embedded, incompressible surfaces not isotopic to  $F$ , where  $F$  denotes  $F \times \{*\}$  for some point  $*$  in  $S^1$ . Choose such a surface and let  $L$  denote a least area surface homotopic to it. Thus  $L$  is embedded and not homotopic to  $F$ . We claim that  $L$  must meet some  $F \times \{pt\}$  non-transversely. As every  $F \times \{pt\}$  is a least area surface, this gives the required example. Consider the projection of  $M$  onto  $S^1$  and restrict to a map  $L \rightarrow S^1$ . As  $L$  cannot equal  $F \times \{pt\}$ , this map will have critical points corresponding precisely to the points where  $L$  meets some  $F \times \{pt\}$  non-transversely. But if this map had no critical points, it would follow that  $L$  would be a bundle over  $S^1$  with this map as the bundle projection. Hence, so long as we choose  $L$  not to be a torus or Klein bottle, we will obtain the required points of non-transverse intersection.

In Sect. 6 we saw that a least area incompressible map from a torus or Klein bottle will have the fewest double curves in its homotopy class [Theorem 6.7]. For surfaces of genus greater than one examples may be produced to show that this no longer holds; we must be satisfied with the weaker minimality result, Theorem 6.3. Explicitly we consider  $F$  to be the double torus  $T^2 \# T^2$  and  $(M, \langle, \rangle)$  to be a certain Riemannian, Haken 3-manifold with mean curvature of  $\partial M = 0$ . We find a general position immersion  $f: F \rightarrow M$  which induces an injection on  $\pi_1$  and has one double curve,  $d(f) = 1$ . However any immersion  $g$  of least area in the homotopy class of  $f$  must be in general position and have two double curves,  $d(g) = 2$ .

The manifold  $M$  embeds smoothly (but not isometrically) in  $R^3$ . We will describe it pictorially in two ways (Diagram 8.1). Begin by setting  $N$  equal to a

regular neighborhood of  $j: F \hookrightarrow R^3$  and  $N' = \text{neib.}(j'(F))$  where  $j$  and  $j'$  are the immersions indicated below. (The parameterizations of  $j(F)$  and  $j'(F)$  are not indicated but will soon be constrained.)

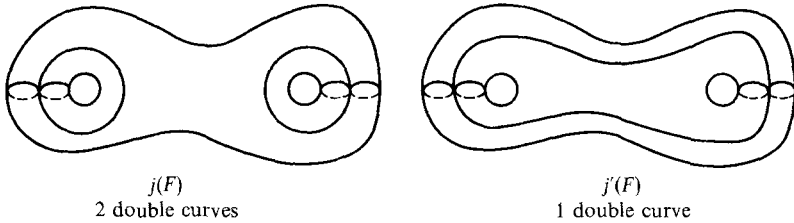


Diagram 8.1

The underlying smooth manifold  $M$  can be obtained by attaching a 2-handle to either  $N$  or  $N'$  as indicated below.  $M = N \cup H = N' \cup H'$ .

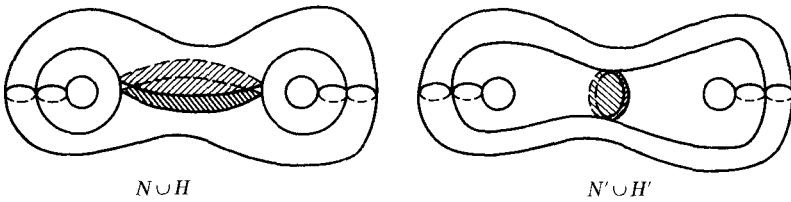


Diagram 8.2

We constrain  $j$  and  $j'$  and our identification of  $N \cup H$  with  $N' \cup H'$  so that the compositions  $F \xrightarrow{j} N \xrightarrow{\text{inc}} M$  and  $F \xrightarrow{j'} N' \xrightarrow{\text{inc}'} M$  are homotopic.

Call these compositions  $f$  and  $f'$ . Our plan is to construct two metrics  $\langle, \rangle$  and  $\langle, \rangle'$  on  $M$ .  $\langle, \rangle$  will be “large” on  $H$  and  $\langle, \rangle'$  will be large on  $H'$ . Considering these two metrics, we show that any least area  $g$  is in general position and that the number of double curves  $d(g)$  is 2 for  $\langle, \rangle$  and one for  $\langle, \rangle'$ . Intuitively, we imagine that the metric on the 2-handle drives  $f$  off  $H$  or  $H'$  (resp.) forcing  $g$  to assume a position similar to  $j$  or  $j'$  (respectively).

The spaces  $N, N'$ , and  $M$  may be replaced by homotopy equivalent 2-complexes  $\bar{N}, \bar{N}'$ , and  $\bar{M}$  formed from  $F$  by attaching certain (1-complexes)  $\times$  interval along (1-complex)  $\times$  boundary (interval). These 2-complexes are homotopy equivalent to the spines of the spaces they replace and even have regular neighborhoods homeomorphic to  $N, N'$ , and  $M$  respectively. The existence of the inclusions  $F \subset \bar{N}, F \subset \bar{N}', F \subset \bar{M}$  make  $\bar{N}, \bar{N}'$ , and  $\bar{M}$  more convenient for us to work with than the actual spines.

We may set  $\bar{N} = F \cup \text{two annuli}$ ,  $\bar{N}' = F \cup \text{one annulus}$ , and  $\bar{M} = F \cup e \times I$  or  $\bar{M} = F \cup \theta \times I$ . The two descriptions of  $\bar{M}$  correspond to the identifications  $M = N \cup H$  and  $M = N' \cup H'$ ;  $e$  is the eyeglasses 1-complex:  $\circ - \circ$  and  $\theta$  represents the 1-complex shape like the letter  $\theta$ . The obvious homotopy equivalences give us the following isomorphisms.

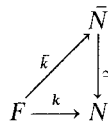
**Lemma 8.1.**

$$\begin{aligned} \pi_1(N) &\cong \pi_1(\bar{N}) \cong (\pi_1(F) * \mathbb{Z}) * \mathbb{Z} \\ \pi_1(N') &\cong \pi_1(\bar{N}') \cong \pi_1(F) * \mathbb{Z} \\ \pi_1(M) &\cong \pi_1(\bar{M}) \cong \pi_1(F) * \text{Free group of rank 2} \quad \blacksquare \end{aligned}$$

An immediate consequence is that the various inclusions induce an injection of  $\pi_1(F)$ .

**Lemma 8.2.** *Let  $k: F \rightarrow N$  be a map such that the composite  $F \rightarrow N \subset M$  is homotopic to the original map  $F \rightarrow M$ . Then  $k$  is homotopic to the original map  $j: F \rightarrow N$ .*

*Proof.* Replace  $N$  by  $\bar{N}$  and  $k$  by  $\bar{k}$  making this diagram commute:



We must show  $\bar{k}$  is homotopic to the inclusion  $i: F \subset \bar{N}$ .

Make  $\bar{k}$  transverse to the two center circles  $\alpha \cup \beta$  of the annuli  $\bar{N} - F$ . A standard argument (see Chap. 3 [8] for example) changes  $\bar{k}$  by a homotopy to eliminate inessential circles from  $\bar{k}^{-1}(\alpha \cup \beta)$ .

It is easily checked that  $H_2(\bar{N}; \mathbb{Z}) \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$ . The first generator is  $i_*[F]$ . The other two are both represented by embedded tori, each consisting of one of the annuli attached to  $F$  and one annulus embedded in  $F$  (with common boundaries). The handle  $H$  kills an element of infinite order in  $H_1(N; \mathbb{Z})$  so the composition  $\bar{N} \xrightarrow{\cong} N \rightarrow M$  induces an isomorphism on  $H_2(\bar{N}; \mathbb{Z})$ . Since  $\text{inc} \circ k$  is homotopic to  $F \hookrightarrow M$ , we deduce that  $k_*[F] = (1, 0, 0) \in H_2(\bar{N}; \mathbb{Z})$ .

The number of circles of  $\bar{k}^{-1}(\alpha)$  and  $\bar{k}^{-1}(\beta)$  (counted according to normal orientation) are equal to the second and third components of  $k_*[F] \in H_2(\bar{N}; \mathbb{Z})$  and so must both be zero. Since  $\bar{k}$  is an injection on  $\pi_1$  the circles  $\bar{k}^{-1}(\alpha)$  are freely homotopic and disjointly embedded, thus they are parallel.  $\bar{k}^{-1}(\alpha)$  lies in an annulus  $A \subset F$  with  $\partial A \subset \bar{k}^{-1}(\alpha)$ . Similarly there is an annulus  $B \subset F$  with  $\bar{k}^{-1}(\beta) \subset B$  and  $\partial B \subset \bar{k}^{-1}(\beta)$ . Since  $\alpha$  is not freely homotopic to  $\beta$  (in  $\bar{N}$ ) neither  $A \subset B$  nor  $B \subset A$  thus  $A$  and  $B$  are disjoint.

It is well known that any map of a torus into a surface of higher genus (actually consider maps into  $\bar{N} - \alpha \cup \beta$ ) factors up to homotopy through a map of a circle. Using this first homotope  $\bar{k}$  so that  $\bar{k}^{-1}(\alpha) = A$  and  $\bar{k}^{-1}(\beta) = B$  and then eliminate the entire inverse image with a final homotopy so that  $\bar{k}^{-1}(\alpha \cup \beta) = \phi$ .

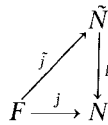
Composing  $\bar{k}$  with the strong deformation retraction  $\bar{N} - \alpha \cup \beta \rightarrow F$  we have homotoped  $\bar{k}$  into  $F$ . This map must be homotopic to  $\text{id}_F$  since we know both maps become homotopic in  $M$  and  $\pi_1(F) \rightarrow \pi_1(M)$  is an injection. In particular  $\bar{k}$  is homotopic to the inclusion  $i: F \subset \bar{N}$ .

*Note.* Similarly, one can also prove that if  $k': F \rightarrow N'$  is a map such that the composite  $F \rightarrow N' \subset M$  is homotopic to the original map  $F \rightarrow M$ , then  $k'$  is homotopic to  $j': F \rightarrow N'$ .

Let  $\tilde{N}, \tilde{N}'$ , and  $\tilde{M}$  be the covering spaces determined by  $\pi_1(F)$ . The lift of  $F$  to each of these spaces is an embedding, denoted simply by  $F$  and the inclusions  $F \subset \tilde{N}, F \subset \tilde{N}'$ , and  $F \subset \tilde{M}$  are homotopy equivalences. There are two distinct (but isotopic) possibilities for  $F \subset \tilde{M}$  arising from the two descriptions of  $M$  (see Diagram 8.2). Diagram 8.3 illustrates how the surface  $F$  in  $\tilde{N}, \tilde{N}'$  or  $\tilde{M}$  meets the other inverse images of  $j(F), j'(F), \text{inc} \circ j(F)$  and  $\text{inc}' \circ j'(F)$ . The accuracy of the illustration may be checked by: (1) constructing the covering spaces by hand or (2) computing the intersection of the various inclusions of  $\pi_1(F)$  with its conjugates.

The following comments should make this checking easier. If  $f: F \rightarrow M$  is an incompressible map and if  $\tilde{M}$  is a covering of  $M$ , then the components of the pre-image in  $\tilde{M}$  of  $f(F)$  in  $M$  correspond bijectively to the double cosets  $\pi_1(\tilde{M})g\pi_1(F)$  in  $\pi_1(M)$ , and the component corresponding to  $\pi_1(\tilde{M})g\pi_1(F)$  has fundamental group isomorphic to  $\pi_1(\tilde{M}) \cap g\pi_1(F)g^{-1}$ . In our situation, the 3-manifold's fundamental group is a *HNN* extension based on  $\pi_1(F)$ . A useful point here is that in a *HNN* extension  $A *_C$ , the subgroup  $A \cap gAg^{-1}$  lies in a conjugate of  $C$  unless  $g$  lies in  $A$  when, of course,  $A \cap gAg^{-1} = A$ .

Suppose  $M$  is given some metric and we find that a least area  $g$  (in the homotopy class of  $f$ ) happens to lie entirely in  $N$ . Lemma 8.2 says that  $g: F \rightarrow N$  is homotopic to  $j: F \rightarrow N$ , as maps into  $N$ . In order to study the double points of  $g$  by the techniques of Sect. 6 it suffices to have a description of the lift

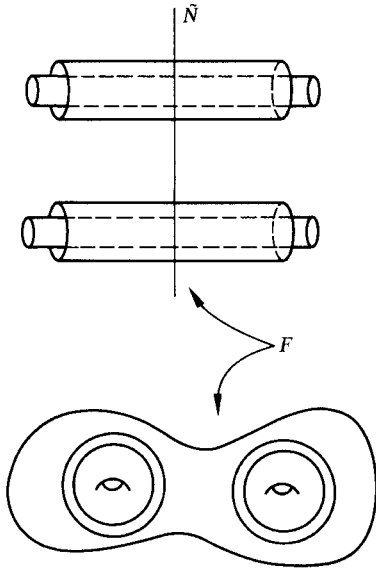


and the components of  $p^{-1}(jF)$  meeting  $\tilde{j}(F)$ . This is provided by Diagram 8.3a. Since all such components meeting  $\tilde{j}(F)$  are cylinders, Lemma 6.5 shows the intersections of  $g$  are transverse, so that the lift  $\tilde{g}$  of  $g$  meets each cylinder in an essential circle. Furthermore the disjointness of these four cylinders in  $p^{-1}(j(F))$  translates, by an exchange argument, into disjointness of the four cylinders of  $p^{-1}(g(F))$  which meet  $\tilde{g}(F)$ . The result is that  $g$  is a general position immersion without triple points with two double curves,  $d(g)=2$ . Similarly if we supposed that the least area  $g$  ( $g \simeq f$ ) lay entirely in  $N'$  we would arrive at the same conclusion except we would find  $d(g)=1$ .

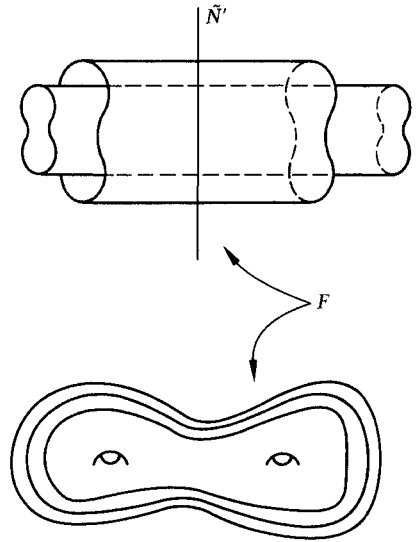
Observe that if we write  $M = N \cup H$  ( $M = N' \cup H'$ ) the preceding paragraph applies equally to least area immersions  $g: F \rightarrow N \cup H - l$  ( $g: F \rightarrow N' \cup H' - l'$ ) where  $l$  ( $l'$ ) is a closed arc which is a co-core of the 2-handle  $H$  ( $H'$ ). Thus it remains to construct a metric  $\langle, \rangle$  ( $\langle, \rangle'$ ) with mean curvature  $\partial M = 0$  and so that any least area  $g$  ( $g \simeq f$ ),  $g: F \rightarrow M$ , will necessarily lie in  $N \cup H - l$  ( $N' \cup H' - l'$ ) for some co-core  $l$  ( $l'$ ). We construct  $\langle, \rangle$ . The construction of  $\langle, \rangle'$  is parallel.

The 2-handle  $H$  has the product structure  $H \stackrel{\text{diff eo}}{=} D^2 \times D^1$ . Let  $1/2D^2 \subset D^2$  be the disk of radius  $= 1/2$ ,  $\{x, y | x^2 + y^2 \leq 1/2\}$ . Any arc  $x \times D^1$ , for  $x \in 1/2D^2$  will be a *co-core*. We may establish a one parameter family of Riemannian metrics  $\langle, \rangle_r$  on  $M$  such that:





$F \cap \text{cylinders} = \text{these 4 circles}$   
Diagram 8.3a



$F \cap \text{cylinders} = \text{these 2 circles}$   
Diagram 8.3b

$F$  meets two disjoint "pairs of pants" in either one or two circles each depending on whether one lifts  $j'$  or  $j$  respectively.

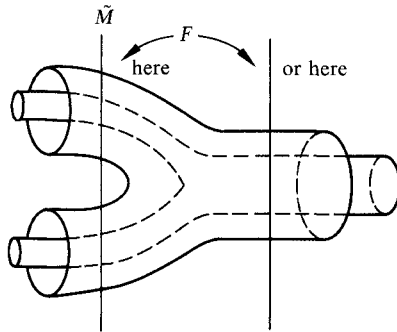


Diagram 8.3c

(1)  $\partial M$  has a product collar. Thus the second fundamental form and a fortiori the mean curvature of  $\partial M$  vanish.

(2) The metric remains constant in a small neighborhood of  $j(F)$ .

(3) The metric on  $1/2D^2 \times D^1$  is  $rdx^2 + rdy^2 + dz^2$  for  $0 < r < \infty$ .

By the existence theorems [18] and [15], for each  $r$  there is at least one least area immersion  $g: F \rightarrow M$  in the homotopy class of  $f$ . By choosing  $r$  so that  $\frac{\pi r^2}{4} \geq \text{area}(f)$  we obtain the following chain of inequalities:

$$\text{area}(g(F) \cap 1/2D^2 \times D^1) < \text{area}(g) \leq \text{area}(f) < \text{area } 1/2D^2 \times 0 = \frac{\pi r^2}{4}.$$

Since projection  $1/2D^2 \times D^1 \xrightarrow{\tau} 1/2D^2 \times 0$  is area nonincreasing we have area  $(\tau(g(F) \cap 1/2D^2 \times D^1)) < \frac{\pi r^2}{4}$ . Thus  $\tau(g(F) \cap 1/2D^2 \times D^1)$  is a proper subset of  $1/2D^2 \times 0$  and some cocore  $l$  must be disjoint from  $g(F)$ . This completes the verification that our examples satisfy the desired properties.

## References

1. Bers, L.: Local behaviour of solutions of general linear elliptic equations. *Comm. Pure Appl. Math.* **8**, 473–496 (1955)
2. Cheng, S.-Y.: Eigenfunctions and Nodal Sets. *Comment. Math. Helv.* **51**, 43–55 (1976)
3. Feustel, C.D.: Some applications of Waldhausen's results on irreducible 3-manifolds. *Trans. Am. Math. Soc.* **149**, 575–583 (1970)
4. Freedman, M.: A conservative Dehn's lemma
5. Freedman, M., Hass, J., Scott, P.: Closed geodesics on surfaces. *Bull. London Math. Soc.* **14**, 385–391 (1982)
6. Gulliver, R.: Regularity of minimising surfaces of prescribed mean curvature. *Annals of Math.* **97**, 275–305 (1973)
7. Hass, J.: Embedded minimal surfaces in three and four dimensional manifolds. Thesis, University of California, Berkeley 1981
8. Hempel, J.: 3-manifolds, *Annals of Math. Studies No. 86*, Princeton University Press, Princeton, N.J. 1976
9. Jaco, W.: Finitely presented subgroups of 3-manifold groups. *Invent. Math.* **13**, 335–346 (1971)
10. Lemaire, L.: Boundary value problems for harmonic and minimal maps of surfaces into manifolds. *Annali della Scuola Normale Superiore di Pisa* **9**, 91–103 (1982)
11. Lu, Y.-C.: Singularity theory and an introduction to catastrophe theory. Berlin-Heidelberg-New York: Springer 1976
12. Meeks, W.H., III: Lectures on Plateau's Problem, I.M.P.A., Rio de Janeiro, Brazil, 1978
13. Meeks, W.H., III, Yau, S.T.: Topology of three dimensional manifolds and the embedding problems in minimal surface theory. *Annals of Math.* **112**, 441–485 (1980)
14. Meeks, W.H., III, Yau, S.T.: The classical plateau problem and the topology of three dimensional manifolds. *Topology* **21**, 409–442 (1982)
15. Meeks, W.H., Yau, S.T.: The existence of embedded minimal surfaces and the problem of uniqueness. *Math. Z.* **179**, 151–168 (1982)
16. Osserman, R.: A survey of minimal surfaces. *Rheinhold Mathematical Studies Vol. 25*. Van Nostrand 1969
17. Sacks, J., Uhlenbeck, K.: Minimal immersions of closed riemann surfaces
18. Schoen, R., Yau, S.T.: Existence of incompressible minimal surfaces and the topology of three dimensional manifolds with non-negative scalar curvature. *Annals of Math.* **110**, 127–142 (1979)
19. Scott, P.: On sufficiently large 3-manifolds. *Q. J. Math., Oxford (2)* **23**, 159–172 (1972)
20. Scott, P.: Subgroups of surface groups are almost geometric. *J. London Math. Soc. (2)* **17**, 555–565 (1978)
21. Stallings, J.R.: On fibering certain 3-manifolds (*Proc. The Univ. of Georgia Inst.*, 1961) Prentice Hall 1962, 120–167
22. Uhlenbeck, K.: Minimal embeddings of surfaces in hyperbolic 3-manifolds
23. Waldhausen, F.: On irreducible 3-manifolds which are sufficiently large. *Annals of Math.* **87**, 56–88 (1968)