Annals of Mathematics, **116** (1982), 621–659

Embedded minimal surfaces, exotic spheres, and manifolds with positive Ricci curvature

By William Meeks III, Leon Simon and Shing-Tung Yau

Let N be a three dimensional Riemannian manifold. Let Σ be a closed embedded surface in N. Then it is a question of basic interest to see whether one can deform Σ in its isotopy class to some "canonical" embedded surface. From the point of view of geometry, a natural "canonical" surface will be the extremal surface of some functional defined on the space of embedded surfaces. The simplest functional is the area functional. The extremal surface of the area functional is called the minimal surface. Such minimal surfaces were used extensively by Meeks-Yau [MY2] in studying group actions on three dimensional manifolds.

In [MY2], the theory of minimal surfaces was used to simplify and strengthen the classical Dehn's lemma, loop theorem and the sphere theorem. In the setting there, one minimizes area among all immersed surfaces and proves that the extremal object is embedded. In this paper, we minimize area among all embedded surfaces isotopic to a fixed embedded surface. In the category of these surfaces, we prove a general existence theorem (Theorem 1). A particular consequence of this theorem is that for irreducible manifolds an embedded incompressible surface is isotopic to an embedded incompressible surface with minimal area. We also prove that there exists an embedded sphere of least area enclosing a fake cell, provided the complementary volume is not a standard ball, and provided there exists no embedded one-sided $\mathbf{R}P^2$.

By making use of the last result, and a cutting and pasting argument, we are able to settle a well-known problem in the theory of three dimensional manifolds. We prove that the covering space of any irreducible orientable three dimensional manifold is irreducible. It is possible to exploit our existence theorem to study finite group actions on three dimensional manifolds as in [MY2].

In the second part of the paper, we apply our existence theorem to study the topology of compact three dimensional manifolds with non-negative Ricci curva-

003–486X/82/0116-3/0621/039$03.90/1

ture. We classify these manifolds except in the case when the manifold is covered by an irreducible homotopy sphere. As a consequence, if one can prove the existence of a metric with positive Ricci curvature on any compact simply connected three dimensional manifold, then the Poincaré conjecture is valid. It should be mentioned that our existence theorem was used by Schoen-Yau [SY2] to prove that the only complete non-compact three dimensional manifold with positive Ricci curvature is diffeomorphic to $\mathbf{R}^3$. In this paper, we also classify the topology of compact three dimensional manifolds whose boundary has non-negative mean curvature with respect to the outward normal.

In the above process, we study the topology of compact embedded orientable minimal surfaces in a three dimensional manifold diffeomorphic to $\mathbf{S}^3 \#_{i=1}^n \mathbf{S}^2 \times \mathbf{S}^1$ which is equipped with a metric with non-negative scalar curvature. We find the condition for which two compact embedded orientable surfaces are conjugate to each other under a diffeomorphism of the ambient space. If the manifold is diffeomorphic to the three dimensional sphere, then the minimal surface is unique topologically. This generalizes a previous theorem of Lawson [LH] where the metric has positive Ricci curvature and a theorem of Meeks [MW2] where the metric has non-negative Ricci curvature.

In the last section, we study complete manifolds (non-compact) with positive Ricci curvature whose boundary has non-negative mean curvature with respect to the outward normal. We prove that the boundary is connected unless it is a Riemannian product or is a handlebody. As in the paper of Frankel [FT], this gives some information about the fundamental group of the boundary.

Finally, we should mention that the regularity of the extremal embedded minimal surface in the main existence theorem depends on the theory of Almgren-Simon [AS], where they deal with minimal surfaces in $\mathbf{R}^3$. It should also be mentioned that very recently, Freedman-Hass-Scott were able to improve one aspect of our theorem and prove that if a compact incompressible minimal surface minimizes area in its homotopy class and if it is homotopic to an embedded surface, then it is embedded.

1. Terminology and statement of main results

$\mathbf{B}_\rho$ will denote the closed 3-ball of radius ρ and center 0 in $\mathbf{R}^3$,

$$\mathbf{B} = \mathbf{B}_1, \qquad \mathbf{S}^2 = \partial \mathbf{B}.$$

$\mathbf{D}$ will denote the closed unit disc with center 0 in $\mathbf{R}^2$.

N will denote a complete (not necessarily orientable) Riemannian 3-manifold. If $\Sigma \subset \Lambda$ is in a smooth surface, we let $|\Sigma|$ denote the area (two dimensional Hausdorff measure) of Σ.

N will always be supposed to have the following "homogeneous regularity" property for some $\rho_0 > 0$:

For each $x_0 \in N$ there is an open geodesic ball $G_{\rho_0}(x_0)$ with center x_0 and radius ρ_0 such that the exponential map $\exp_{x_0}$ provides a diffeomorphism φ of $\mathbf{B}_{\rho_0}$ onto $\overline{G}_{\rho_0}(x_0)$, satisfying

$$\|d_y\varphi\|, \qquad \|d_z\varphi^{-1}\| \leq 2, \qquad y \in \mathbf{B}_{\rho_0}, \qquad z \in G_{\rho_0}(x_0). \tag{1.1}$$

We also require that there be a constant μ independent of x_0 such that

$$\sup_{\mathbf{B}_{\rho_0}}\left|\frac{\partial g_{ij}}{\partial x^k}\right| \leq \mu/\rho_0, \qquad \sup_{\mathbf{B}_{\rho_0}}\left|\frac{\partial^2 g_{ij}}{\partial x^k \partial x^l}\right| \leq \mu/\rho_0^2 \tag{1.2}$$

for $i, j, k, l = 1, 2, 3$, where $g_{ij}\,dx^i\,dx^j$ is the metric relative to normal coordinates for $G_{\rho_0}(x_0)$.

Of course it is trivial that such a ρ_0 and such a μ exist in case N is compact. By using comparison theorems in differential geometry, we can prove that a manifold is homogeneous regular if and only if it is a complete manifold whose injectivity radius is bounded from below and whose sectional curvature is bounded.

$\mathcal{C}$ will denote the collection of all connected compact (not necessarily orientable) smooth 2-dimensional surfaces-without-boundary embedded in N. $\mathcal{C}_1$ will denote the collection of compact embedded surfaces Σ such that each component of Σ is an element of $\mathcal{C}$.

Given $\Sigma \in \mathcal{C}_1$, we let $\mathcal{I}(\Sigma)$ denote the isotopy class of Σ; that is $\mathcal{I}(\Sigma)$ is the collection of all $\tilde{\Sigma} \in \mathcal{C}_1$ such that $\tilde{\Sigma}$ is isotopic to Σ via a smooth isotopy φ: $[0,1] \times N \to N$, where $\varphi_0 = \mathbf{1}_N$ and each φ_t is a diffeomorphism of N onto N. Here φ_t is defined by $\varphi_t(x) = \varphi(t, x)$, $(t, x) \in [0,1] \times N$; we shall often write $\varphi = \{\varphi_t\}_{0 \leq t \leq 1}$. In case N is non-compact, we also require that there be a fixed compact $K \subset N$ such that $\varphi_t|_{N \sim K} = \mathbf{1}_{N \sim K}$ for each $t \in [0,1]$.

Now suppose $\Sigma \in \mathcal{C}$ is given. If $\inf_{\tilde{\Sigma} \in \mathcal{I}(\Sigma)} |\tilde{\Sigma}| \neq 0$, then we may select a sequence $\{\Sigma_k\} \subset \mathcal{I}(\Sigma)$ with $\lim |\Sigma_k| = \inf_{\tilde{\Sigma} \in \mathcal{I}(\Sigma)} |\tilde{\Sigma}|$. We call such a sequence a minimizing sequence for $\mathcal{I}(\Sigma)$. More generally, $\{\Sigma_k\} \subset \mathcal{I}$ is called a minimizing sequence if $|\Sigma_k| \leq \inf_{\Sigma \in \mathcal{I}(\Sigma_k)} |\Sigma| + \varepsilon_k$ with $\varepsilon_k \to 0$ as $k \to \infty$ and if $\limsup_{k \to \infty}(|\Sigma_k| + \operatorname{genus}(\Sigma_k)) < \infty$.

By a standard compactness theorem for Borel measures (applied to the measures μ_k given by $\mu_k(f) = \int_{\Sigma_k} f$, $f \in C_0(N)$), we know that there is a subsequence $\{\Sigma_{k'}\} \subset \{\Sigma_k\}$ and a Borel measure μ with

$$\mu(f) = \lim_{k' \to \infty} \int_{\Sigma_{k'}} f, \qquad f \in C_0(N). \tag{1.3}$$

Our main existence theorem is then given as follows:

THEOREM 1. *Suppose N is compact, and $\{\Sigma_k\} \subset \mathcal{C}$ is a minimizing sequence; let $\{\Sigma_{k'}\}$, μ be as in* (1.3) *above, and suppose* $\lim |\Sigma_{k'}| > 0$.

Then there are positive integers $R, n_1, \ldots, n_R$ and pairwise disjoint minimal surfaces $\Sigma^{(1)}, \ldots, \Sigma^{(R)} \in \mathcal{C}$ such that

$$\Sigma_{k'} \to n_1\Sigma^{(1)} + n_2\Sigma^{(2)} + \cdots + n_R\Sigma^{(R)}$$

(*in the sense that* $\mu(f) = \sum_{j=1}^{R} n_j \int_{\Sigma^{(j)}} f$, $f \in C(N)$, μ *as in* (1.3)).

Furthermore, if $g_j = \text{genus}(\Sigma^{(j)})$, *then*

$$\sum_{j \in \mathcal{U}} \tfrac{1}{2} n_j (g_j - 1) + \sum_{j \in \mathcal{O}} n_j g_j \leq \text{genus}(\Sigma_{k'}) \tag{1.4}$$

for all sufficiently large k', where $\mathcal{U} = \{j: \Sigma^{(j)}$ is one-sided in $N\}$ and $\mathcal{O} = \{j: \Sigma^{(j)}$ is two-sided in $N\}$. (Notice that $g_j \geq 1$ for all $j \in \mathcal{U}$; hence all terms in (1.4) *are non-negative.)*

If each $\Sigma_{k'}$ is two-sided in N, then each $\Sigma^{(j)}$ satisfies the stability

$$\int_{\Sigma^{(j)}} \left(\zeta^2\left(|A|^2 + \text{Ric}(\nu, \nu)\right) - |\nabla\zeta|^2\right) \leq 0, \qquad \zeta \in C^1(\Sigma^{(j)}), \tag{1.5}$$

where ν is any unit normal for $\Sigma^{(j)}$ ($\Sigma^{(j)}$ need not necessarily be two-sided, so that here ν is not assumed to be continuous.) In any case, even when $\Sigma_{k'}$ is one-sided, (1.5) *holds for any $\Sigma^{(j)}$ which is two-sided in N.*

In case N merely satisfies the homogeneous regularity condition described above, the hypothesis $\lim |\Sigma_{k'}| > 0$ *must be replaced by the hypothesis that* $\liminf |\Sigma_{k'} \cap K| > 0$ *for some compact $K \subset N$; then the above conclusions continue to hold. In fact there are currents $S_{k'}$ which tend to infinity and which can be written as finite sums of embedded closed surfaces with uniformly bounded diameter and with area bounded from below (by a fixed positive constant) so that*

$$\Sigma_{k'} - S_{k'} \to n_1\Sigma^{(1)} + n_2\Sigma^{(2)} + \cdots + n_k\Sigma^{(k)},$$

and

$$\lim |S_{k'}| = \lim |\Sigma_{k'}| - \left(n_1|\Sigma^{(1)}| + \cdots + n_k|\Sigma^{(k)}|\right).$$

Furthermore (1.4) *and* (1.5) *hold, where in* (1.4) *we can add the corresponding sum associated to the genus of the surfaces in $S_{k'}$.*

In particular, if N satisfies the additional condition that for each $c > 0$ there exists a compact set of N so that a geodesic ball of radius c in the *complement of this compact set is a subset of some open domain diffeomorphic to the ball, then we may take* $S_{k'} = \varnothing$.

(1.6) *Remarks.* 1. We shall give a more precise statement concerning the relation (up to isotopy) of the $\Sigma^{(1)}, \ldots, \Sigma^{(R)}$, $n_1, \ldots, n_R$ and the sequence $\{\Sigma_{k'}\}$ at a later stage. (See Remark (3.27).)

2. We shall also show that if each $\Sigma_{k'}$ is two-sided, then all the $\Sigma^{(i)}$ such that n_i is *odd* are also two-sided. However (see Remark (3.27)), $\Sigma^{(i)}$ may be one-sided in case n_i is even.

3. Corresponding to each Σ_k we have a varifold $\mathbf{v}(\Sigma_k)$ (see [AW, § 3.5]); since $|\Sigma_k|$ is bounded, a subsequence $\mathbf{v}(\Sigma_{\hat{k}})$ of $\mathbf{v}(\Sigma_{k'})$ will converge to a stationary varifold V such that $\|V\| = \mu$ (μ as in (1.3)). In view of the constancy theorem ([AW]) the content of the theorem is then

$$V = n_1\mathbf{v}(\Sigma^{(1)}) + \cdots + n_R\mathbf{v}(\Sigma^{(R)}), \tag{1.7}$$

with n_i, $\Sigma^{(i)}$ as described.

The fact that V is stationary, and in fact stable, is readily seen as follows. If $\varphi = \{\varphi_t\}_{0\le t\le 1}$ is any smooth isotopy as above, then

$$\mathbf{M}(\varphi_{t\#}\mathbf{v}(\Sigma_k)) = \mathbf{M}(\mathbf{v}(\varphi_t(\Sigma_k))) \geq |\Sigma_k| - \varepsilon_k \qquad (\varepsilon_k \to 0)$$

for each $t \in [0,1]$ and each $k = 1, 2, \ldots$, by the assumption that $\{\Sigma_k\}$ is a minimizing sequence. Thus, taking limits as $k \to \infty$, we get

$$\mathbf{M}(\varphi_{t\#}V) \geq \mathbf{M}(V) \tag{1.8}$$

for every such isotopy. Notice that this is in fact a stronger condition than stability, because here φ is any isotopy as described above.

Since V is stationary in N we have that there are constants $\eta \in (0,1)$ and $c_1 > 0$ (depending only on μ) such that $c_1\eta \le \frac{1}{2}$ and $(1 - c_1\rho/\rho_0)\rho^{-2}\|V\|(G_\rho(y))$ is increasing in ρ for $\rho \le \eta\rho_0$. In particular it follows that

$$\sigma^{-2}\|V\|(G_\sigma(y)) \le c_2\rho^{-2}\|V\|(G_\rho(y)) \tag{1.9}$$

for any $0 < \sigma < \rho \le \rho_0$, with c_2 depending only on μ_0.

For a proof of this (which uses (1.1), (1.2)), see the example [SS, § 5]; the proof is a straightforward modification of standard monotonicity arguments. (See e.g. [AW], [MS].)

2. Preliminary lemmas

LEMMA 1. *Let* ρ_0 *be as in* (1.1) *and* (1.2). *There is a number* $\delta \in (0,1)$ (*independent of* N, ρ_0) *such that if* $\Sigma \in \mathcal{C}$ *satisfies*

$$|\Sigma \cap G_{\rho_0}(x_0)| < \delta^2\rho_0^2 \tag{2.1}$$

for each $x_0 \in N$, *then there exists a unique compact* $K_\Sigma \subset N$ *with* $\partial K_\Sigma = \Sigma$ *and*

$$\text{(2.2)} \qquad \operatorname{vol}\big(K_\Sigma \cap G_{\rho_0}(x_0)\big) \le \delta^2 \rho_0^3, \qquad x_0 \in N.$$

This K_Σ *also satisfies*

$$\text{(2.3)} \qquad \operatorname{vol}(K_\Sigma) \le c\,|\,\Sigma\,|^{3/2}, \qquad c = c(\mu).$$

Also, if $\Sigma \approx \mathbf{S}^2$, *then* $K_\Sigma \approx \mathbf{B}$.

(2.4) *Remark.* Evidently, if the hypotheses of the above lemma hold and if $\tilde{\Sigma}$ is isotopic to Σ via an isotopy $\varphi = \{\varphi_t\}_{0 \le t \le 1}$ as in Section 1, then $N \sim \tilde{\Sigma}$ has two components U, V such that $\bar{U}$ is diffeomorphic to K_Σ.

Proof. We first note that by (1.1), (1.2) we can find a triangulation $\mathcal{K}^0$ for N such that each 3-cell K of $\mathcal{K}^0$ has $\operatorname{diam} K \le \rho_0$ and such that there is a diffeomorphism φ_K of $K_{\rho_0} \equiv \{x \in N\colon \operatorname{dist}(x, N) < \rho_0\}$ onto s_{ρ_0}, $s = \{(x^1, x^2, x^3) \in \mathbf{R}^3\colon\ 0 \le x^i,\ \Sigma_{i=1}^3 x^i \le \rho_0\}$, $s_{\rho_0} = \{x \in \mathbf{R}^3\colon\ \operatorname{dist}(x, s) < \rho_0\}$, with $\varphi_K(K) = s$ and

$$\text{(2.5)} \quad \sup_{y \in K_{\rho_0}} \|d_y \varphi_K\| \le c_1, \qquad \sup_{z \in s_{\rho_0}} \|d_z \varphi_K^{-1}\| \le c_1, \qquad c_1 = c_1(\mu).$$

For δ small enough we can perturb $\mathcal{K}^0$ slightly to give a new triangulation $\mathcal{K}$ such that Σ *does not intersect the* 1-*skeleton of* $\mathcal{K}$,* and such that for each 3-cell K of $\mathcal{K}$ there is a diffeomorphism ψ_K of K onto s with

$$\text{(2.5)}' \qquad \sup_{y \in K} \|d_y \psi_K\| \le {}'c_1', \sup_{z \in s} \|d_z \psi_K^{-1}\| \le c_1', c_1' = c_1'(\mu).$$

(Of course $\mathcal{K}$ then depends on Σ, but we do have (2.5)′ with c_1' depending only on μ.) We note that this $\mathcal{K}$ can be selected so that (by virtue of (2.1)) there is at least one 3-cell $K^0 \in \mathcal{K}$ such that Σ intersects ∂K^0 transversally and

$$\text{(2.6)} \qquad \operatorname{length}(\Sigma \cap \partial K^0) \le c_2 \delta \rho_0, \qquad c_2 = c_2(\mu).$$

Since Σ does not intersect the 1-skeleton of $\mathcal{K}$, Σ must be contained in a regular neighborhood of the dual triangulation $\mathcal{K}'$. Therefore Σ is contained in a handlebody and it is then standard that there is a compact K_Σ, contained in this handlebody, which is bounded by Σ; thus

$$\text{(2.7)} \qquad \partial K_\Sigma = \Sigma, \qquad K_\Sigma \cap E = \varnothing,$$

where E denotes the 1-skeleton of $\mathcal{K}$. It is also standard that this K_Σ is then diffeomorphic to $\mathbf{B}$ in case $\Sigma \approx \mathbf{S}^2$.

---

*To see this, let σ be any edge of s, let $D(\sigma)$ be the disc normal to σ with radius $\rho_0/2$ and centre at the mid-point of σ, and for each $\xi \in D(\sigma)$, let σ_ξ be the segment parallel to σ with mid-point ξ and length equal to length $(\sigma) + \rho_0$. By (2.1) and (2.5) we must have $\sigma_\xi \cap \varphi_K(\Sigma \cap K_{\rho_0}) = \varnothing$ except for a set of $\xi \in D(\sigma)$ of area $\le c\delta^2\rho_0^2$. For δ small enough it will then evidently be possible to choose $\mathcal{K}$ so that each edge of $\mathcal{K}$ is contained in one of the curves in the family $\{\varphi_K^{-1}(\sigma_\xi)\colon K \in \mathcal{K}^0,\ \xi \in D(\sigma),\ \sigma \text{ an edge of } s\}$.

It thus remains only to prove (2.2) and (2.3). First take K^0 as in (2.6). By virtue of (2.5)′, (2.6), and (2.7), we must have (by the isoperimetric inequality in $\mathbf{R}^2$ that

$$|K_\Sigma \cap \partial K^0| \le c_3\delta^2\rho_0^2, \qquad c_3 = c_3(\mu).$$

Then by (2.1), (2.5)′ and the isoperimetric inequality in $\mathbf{R}^3$ we have

$$\operatorname{vol}\left(K_\Sigma \cap K^0\right) \le c_4\delta^3\rho_0^3, \qquad c_4 = c_4(\mu). \tag{2.8}$$

On the other hand, for any 3-cells K, $K' \in \mathcal{K}$ *such that* K, K' *share a common two-dimensional face*, we have by (2.1) and the Poincaré inequality in $\mathbf{R}^3$, that

$$\min\{\operatorname{vol}((K \cup K') \cap K_\Sigma), \operatorname{vol}((K \cup K') \sim K_\Sigma)\} \le c_5\delta^3\rho_0^3.$$

Thus if δ is small enough we must have (since $\operatorname{vol}(K)$, $\operatorname{vol}(K') \le c_6\rho_0^3$ by (2.5)),

$$\begin{array}{ll} \textit{either} & \max\{\operatorname{vol}(K \cap K_\Sigma), \operatorname{vol}(K' \cap K_\Sigma)\} \le c_6\delta^3\rho_0^3 \\ \textit{or} & \min\{\operatorname{vol}(K \cap K_\Sigma), \operatorname{vol}(K' \cap K_\Sigma)\} \ge c_6\rho_0^3, \end{array}$$

where $c_6 = c_6(\mu)$. Since this is true for any K, $K' \in \mathcal{K}$ sharing a common face, we then have from (2.8) that (for δ small enough)

$$\operatorname{vol}(K \cap K_\Sigma) \le c_7\delta^3\rho_0^3 \tag{2.9}$$

for every 3-cell K in the triangulation $\mathcal{K}$. By (2.5) and the Poincaré inequality, this implies, again for small enough δ, that

$$\operatorname{vol}(K \cap K_\Sigma) \le c_8|\Sigma \cap K|^{3/2} \text{ for all } K \in \mathcal{K}; \tag{2.10}$$

(2.2) and (2.3) now evidently follow from (2.9), (2.10), provided δ is sufficiently small.

LEMMA 2. *Suppose $M_1,\ldots,M_R$ are diffeomorphic to $\mathbf{D}$, suppose $M_j \sim \partial M_j \subset A \sim \partial A$, $\partial M_j \subset \partial A$, $j = 1,\ldots,R$, where $A \subset N$ is diffeomorphic to $\mathbf{B}$, and suppose that $\partial M_i \cap \partial M_j = \varnothing$ and that either $M_i \cap M_j = \varnothing$ or M_i intersects M_j transversally for all $i \neq j$.*

Then there exist pairwise disjoint $\tilde{M}_1,\ldots,\tilde{M}_R$ with $\tilde{M}_j \sim \partial\tilde{M}_j \subset A \sim \partial A$, $\partial\tilde{M}_j = \partial M_j$ and $|\tilde{M}_j| \le |M_j|$, $j = 1,\ldots,R$.

Proof. We can evidently assume $R \ge 2$ and that $M_1,\ldots,M_{R-1}$ are already pairwise disjoint. (If we can prove the required result in this case, then the general result clearly follows by induction on R.)

Let $\Gamma_1,\ldots,\Gamma_q$ be pairwise disjoint Jordan curves such that

$$M_R \cap \left(\bigcup_{j=1}^{R-1} M_j\right) = \bigcup_{j=1}^{q} \Gamma_j, \tag{2.11}$$

and, as an inductive hypothesis, assume the theorem true whenever (2.11) holds with $r \leq q - 1$ in place of q ($M_1, \ldots, M_{R-1}$ still being assumed pairwise disjoint).

For each $j = 1, \ldots, q$, let E_j be the disc contained in M_R such that $\partial E_j = \Gamma_j$, let F_j be the corresponding disc in $\bigcup_{i=1}^{R-1} M_i$ with $\partial F_j = \Gamma_j$, and let $K \subset \bigcup_{i=1}^{R} M_i$ be a disc such that $\partial K = \Gamma_{j_0}$ for some j_0 and such that

$$|K| \leq |F_j|, |E_j| \qquad \text{for all } j = 1, \ldots, q.$$

Let $J \neq K$ be the other disc in $\bigcup_{j=1}^{R} M_j$ such that $\partial J = \partial K (= \Gamma_{j_0})$. Evidently we must then have

$$(K \sim \partial K) \cap \left(\bigcup_{i \neq i_0} M_i \right) = \varnothing, \tag{2.12}$$

where i_0 is such that $K \subset M_{i_0}$. Let i_1 ($\neq i_0$) be such that $J \subseteq M_{i_1}$ (note that then one of i_0, i_1 is equal to R), and define $\hat{M}_j = M_j$ if $j \neq i_1$ and $\hat{M}_{i_1} = (M_{i_1} \sim J) \cup K$. By (2.12) we have that each $\hat{M}_j$ is an embedded disc, and clearly $\partial \hat{M}_j = \partial M_j$, $|\hat{M}_j| \leq |M_j|$, $\hat{M}_1, \ldots, \hat{M}_{R-1}$ are disjoint and

$$\hat{M}_R \cap \left(\bigcup_{j=1}^{R-1} \hat{M}_j \right) = K \cup \left(\bigcup_{j \neq j_0} \Gamma_j \right). \tag{2.13}$$

By smoothing $\hat{M}_{i_1}$ near Γ_{j_0} and making a slight perturbation near K, we then obtain discs $\hat{M}_1^*, \ldots, \hat{M}_R^*$ with $\partial \hat{M}_j^* = \partial M_j$, $|\hat{M}_j^*| \leq |M_j|$, $\hat{M}_1^*, \ldots, \hat{M}_{R-1}^*$ pairwise disjoint, and (with (2.13)),

$$\hat{M}_R^* \cap \left(\bigcup_{j=1}^{R-1} \hat{M}_j^* \right) = \bigcup_{j \neq j_0} \Gamma_j.$$

Hence we can apply the inductive hypothesis to the collection $\{\hat{M}_j^*\}$, thus obtaining the required collection $\tilde{M}_1, \ldots, \tilde{M}_R$.

3. γ-reduction

Here and subsequently $\delta > 0$ is a fixed number such that the conclusion of Lemma 1 holds, and $0 < \gamma < \delta^2/9$. We shall also assume for convenience of notation that $\rho_0 = 1$ throughout this section. (This can of course be arranged by changing scale in N.)

Given $\Sigma_1, \Sigma_2 \in \mathcal{C}_1$, we write

$$\Sigma_2 \underset{\gamma}{\ll} \Sigma_1,$$

and we say Σ is a γ-reduction of Σ_1, if the following conditions are satisfied:

(i) $\Sigma_1 \sim \Sigma_2$ has closure A diffeomorphic to the standard closed annulus $\{x \in \mathbf{R}^2, \frac{1}{2} \leq |x| \leq 1\}$;

(ii) $\Sigma_2 \sim \Sigma_1$ has closure consisting of two components D_1, D_2, each diffeomorphic to $\mathbf{D}$;

$$\text{(iii)} \quad \begin{cases} \partial A = \partial D_1 \cup \partial D_2, & |A| + |D_1| + |D_2| < 2\gamma \\ & Y \text{ homeomorphic to } \mathbf{B} \text{ and} \\ A \cup D_1 \cup D_2 = \partial Y, & (Y \sim \partial Y) \cap (\Sigma_1 \cup \Sigma_2) = \varnothing; \end{cases}$$

(iv) In case $\Sigma_1^* \sim A$ is not connected, each component is either not simply connected or else has area $\geq \delta^2/2$; here Σ_1^* denotes the component of Σ_1 containing A.

(3.1) *Remark*. Notice that in case $\Sigma_2 \underset{\gamma}{\ll} \Sigma_1$, then each component of Σ_2 is two-sided in N if each component of Σ_1 is two-sided in N.

Notice also that

$$\text{genus}\,\Sigma_2 \leq \text{genus}\,\Sigma_1, \tag{3.2}$$

with strict inequality if Σ_2 has the same number of components as Σ_1. (Here, genus $\Sigma = \sum_{j=1}^{k}$ genus $\Sigma^{(j)}$, $\Sigma^{(1)}, \ldots, \Sigma^{(k)}$ denoting the components of Σ; in case Σ is connected and orientable, genus Σ is the number of handles and in case Σ is connected but not orientable, genus Σ is the number of cross-caps.) Indeed if $\Sigma_2 \underset{\gamma}{\ll} \Sigma_1$ then genus $\Sigma_2 <$ genus Σ_1 or else one of the components Σ_1^* (as in (iv) above) forms two new components Λ_1, Λ_2 ($\Lambda_1 \cup \Lambda_2 = (\Sigma_1^* \sim A) \cup D_1 \cup D_2$) in the notation of (i)–(iv) above), where $|\Lambda_1| + |\Lambda_2| \leq |\Sigma_1^*| + 2\gamma \leq |\Sigma_1^*| + \delta^2/3$ (since $2\gamma < \delta^2/3$) and for each $i = 1, 2$ either Λ_i is not simply connected or else $|\Lambda_i| \geq \delta^2/2$. It is thus clear that, given any sequence

$$\Sigma_k \underset{\gamma}{\ll} \Sigma_{k-1} \underset{\gamma}{\ll} \cdots \underset{\gamma}{\ll} \Sigma_1, \tag{3.3}$$

there must be a bound on k depending only on genus Σ_1 and $|\Sigma_1|/\delta^2$. Thus (provided $2\gamma < \delta^2/3$) we have

$$k \leq c, \tag{3.4}$$

where c depends only on δ and any upper bound for genus(Σ_1) and $|\Sigma_1|/\delta^2$.

We say that Σ is γ-irreducible if there is no $\tilde{\Sigma} \in \mathcal{C}_1$ with $\tilde{\Sigma} \underset{\gamma}{\ll} \Sigma$. Evidently by (3.4) we know that for any $\Sigma_1 \in \mathcal{C}_1$, either Σ_1 is γ-irreducible or there is a sequence as in (3.3) such that Σ_k is γ-irreducible.

Furthermore, since $|(\tilde{\Sigma} \sim \Sigma) \cup (\Sigma \sim \tilde{\Sigma})| \leq 2\gamma$ whenever $\tilde{\Sigma} \underset{\gamma}{\ll} \Sigma$, we have that if Σ_j are as in (3.3), then

$$|(\Sigma_k \sim \Sigma_1) \cup (\Sigma_1 \sim \Sigma_k)| \leq 2c\gamma \qquad (c \text{ as in } (3.4)). \tag{3.5}$$

We also note that if $\Sigma \in \mathcal{C}$ is an incompressible surface ([HJ, Ch. 6]) and if (3.3) holds, then Σ_k is homeomorphic to disjoint union of Σ and $(k-1)$-

diffeomorphic copies of $\mathbf{S}^2$, each diffeomorphic copy of $\mathbf{S}^2$ having area $\geq \delta^2/2$. (There is thus a fixed bound on the *number* of diffeomorphic copies of $\mathbf{S}^2$ involved.)

(3.6) *Remark*. One readily checks that $\Sigma \in \mathcal{C}_1$ is γ-irreducible if and only if the following holds:

Whenever Δ is a disc with $\partial\Delta = \Delta \cap \Sigma$ and $|\Delta| < \gamma$, then there is a disc $\tilde{\Delta} \subset \Sigma$ with $\partial\tilde{\Delta} = \partial\Delta$ and $|\tilde{\Delta}| < \delta^2/2$.

It will be convenient to consider a slightly weaker relation than $\underset{\gamma}{\ll}$: viz. for $\Sigma_1, \Sigma_2 \in \mathcal{C}_1$ we write

$$\Sigma_2 \underset{\gamma}{<} \Sigma_1 \tag{3.7}$$

if there is a $\tilde{\Sigma}_1 \in \mathcal{G}(\Sigma_1)$ with $|(\tilde{\Sigma}_1 \sim \Sigma_1) \cup (\Sigma_1 \sim \tilde{\Sigma}_1)| < \gamma$ and $\Sigma_2 \underset{\gamma}{\ll} \tilde{\Sigma}_1$.

In view of the above discussion of the relation $\underset{\gamma}{\ll}$, it is evident that if $3\gamma < \delta^2/3$, which we subsequently assume, and if

$$\Sigma_k \underset{\gamma}{<} \Sigma_{k-1} \underset{\gamma}{<} \cdots \underset{\gamma}{<} \Sigma_1 \tag{3.8}$$

then (cf. (3.4), (3.5))

$$k \leq c, \tag{3.9}$$

$$|(\Sigma_k \sim \Sigma_1) \cup (\Sigma_1 \sim \Sigma_k)| \leq 3c\gamma. \tag{3.10}$$

We shall say that Σ is *strongly γ-irreducible* if there is no $\tilde{\Sigma} \in \mathcal{C}_1$ with $\tilde{\Sigma} \underset{\gamma}{<} \Sigma$. Of course strongly γ-irreducible implies γ-irreducible. Also, given $\Sigma_1 \in \mathcal{C}_1$ we have either that Σ_1 is strongly γ-irreducible or else there exist $\Sigma_2, \ldots, \Sigma_k \in \mathcal{C}_1$ such that Σ_k is strongly γ-irreducible and such that (3.8), (3.9), (3.10) hold.

(3.11) *Remark*. If $\Sigma_1 \in \mathcal{G}(\Sigma)$ with $|(\Sigma \sim \Sigma_1) \cup (\Sigma_1 \sim \Sigma)| < \theta < \gamma$, and if Σ is strongly γ-irreducible, then Σ_1 is strongly $(\gamma - \theta)$-irreducible (because $|(\Sigma_1 \sim \tilde{\Sigma}) \cup (\tilde{\Sigma} \sim \Sigma_1)| < \gamma - \theta$ implies $|(\Sigma \sim \tilde{\Sigma}) \cup (\tilde{\Sigma} \sim \Sigma)| < \gamma$, by virtue of the fact that $|(\Sigma \sim \Sigma_1) \cup (\Sigma_1 \sim \Sigma)| < \theta$).

The following theorem gives our main result for strongly irreducible $\Sigma \in \mathcal{C}_1$. In this theorem we use the notation that

$$E(\Sigma) = |\Sigma| - \inf_{\tilde{\Sigma} \in \mathcal{G}(\Sigma)} |\tilde{\Sigma}|.$$

THEOREM 2. *Suppose $A \subset N$ is diffeomorphic to $\mathbf{B}$, suppose $\Sigma \in \mathcal{C}_1$, $E(\Sigma) \leq \gamma/4$, Σ is strongly γ-irreducible, Σ intersects ∂A transversally, and, for each component Γ of $\Sigma \cap \partial A$, let F_Γ be a disc in ∂A with $\partial F_\Gamma = \Gamma$ and $|F_\Gamma| = \min\{|F_\Gamma|, |\partial A \sim F_\Gamma|\}$. Suppose furthermore that $\sum_{j=1}^{q} |F_j| \leq \gamma/8$, where $F_j = F_{\Gamma_j}$ and $\Gamma_1, \ldots, \Gamma_q$ denote the components of $\Sigma \cap \partial A$, and let Σ_0 denote the* *union of all components Λ of Σ such that $\Lambda \subset K_\Lambda$ and $\partial K_\Lambda \cap \Sigma = \varnothing$ for some $K_\Lambda \subset N$ with K_Λ diffeomorphic to* $\mathbf{B}$.

Then $|\Sigma_0| \leq E(\Sigma)$ and there exist pairwise disjoint closed discs $D_1, \ldots, D_p$ with

$$(3.12)\quad D_j \subset \Sigma \sim \Sigma_0, \qquad \partial D_j \subset \partial A, \qquad \sum_{j=1}^{p} |D_j| \leq \sum_{j=1}^{q} |F_j| + E(\Sigma),$$

$$\left(\bigcup_{j=1}^{p} D_j\right) \cap A = (\Sigma \sim \Sigma_0) \cap A$$

and with

$$(3.13)\qquad \bigcup_{j=1}^{p} \left(\varphi_1(D_j) \sim \partial D_j\right) \subset A \sim \partial A$$

for some isotopy $\varphi = \{\varphi_t\}_{0 \leq t \leq 1}$ of N such that $\varphi_t(x) = x$ for all $(t, x) \in [0,1] \times W$, W some neighborhood of $(\Sigma \sim \Sigma_0) \sim \bigcup_{j=1}^{p}(D_j \sim \partial D_j)$.

(3.14) *Remark*. Let $K_1, \ldots, K_S$ be diffeomorphs of $\mathbf{B}$ such that $\Sigma_0 \subset \bigcup_{j=1}^{s} K_j$ and $\partial K_j \cap \Sigma = \varnothing$, for all $j = 1, \ldots, s$.

Since each K_j is diffeomorphic to $\mathbf{B}$ (and since $(\bigcup_{j=1}^{s} K_j) \cap (\Sigma \sim \Sigma_0) = \varnothing$ by definition of Σ_0), we can show that for each $\varepsilon > 0$ there is an open $U \supset \bigcup_{j=1}^{s} K_j$ and an isotopy $\psi = \{\psi_t\}_{0 \leq t \leq 1}$ with $U \cap (\Sigma \sim \Sigma_0) = \varnothing$ and

(3.15)

$$\psi_t(x) = x \text{ for all } (x, t) \in (N \sim U) \times [0,1], \qquad \psi_t(U) \subset U, \qquad |\psi_1(\Sigma_0)| < \varepsilon,$$

and such that each component of $\psi_1(\Sigma_0)$ has diameter less than ε.

It follows, by definition of $E(\Sigma)$, that

$$(3.16)\quad |\Sigma_0| \leq E(\Sigma) \qquad \text{and} \qquad |\Sigma \sim \Sigma_0| \leq \inf_{\tilde{\Sigma} \in \mathcal{J}(\Sigma \sim \Sigma_0)} |\tilde{\Sigma}| + E(\Sigma).$$

Because of this, and because of (3.15) and Lemma 1, it is not difficult to see that the general case of Theorem 1 follows directly from the special case where $\Sigma_0 = \varnothing$.

We shall therefore assume, in the proof below, that

$$(3.17)\qquad \Sigma_0 = \varnothing.$$

Proof of Theorem 2. We proceed by induction on q. Assume that Σ satisfies the hypotheses (3.17) and (in the notation of Theorem 2)

$$(3.18)\quad \sum_{j=1}^{q} |F_j| \leq \gamma/8,\ E(\Sigma) \leq \gamma/2 - \sum_{j=1}^{q} |F_j|,\ \Sigma \text{ is strongly } \tilde{\gamma}\text{-irreducible},$$

where $\tilde{\gamma} = \gamma/4 + 4\sum_{j=1}^{q} |F_j| + E(\Sigma)$, and, as an inductive hypothesis, assume the theorem true, with $\hat{\Sigma}_0 = \varnothing$ and $\Sigma_{j=1}^{p} |D_j| \le \Sigma_{j=1}^{q} |F_j| + E(\hat{\Sigma})$, whenever $\hat{\Sigma} \in \mathcal{G}(\Sigma)$ satisfies the hypotheses (3.17), (3.18) with $q-1$ in place of q and $\hat{\Sigma}$ in place of Σ.

Relabeling, if necessary, we may assume that $F_q \cap \Gamma_j = \varnothing$ for all $j \neq q$. Since Σ is strongly $\tilde{\gamma}$-irreducible and $|F_q| < \tilde{\gamma}$, we know by Remark (3.6) that there is a disc $D \subset \Sigma$ such that $\partial D = \Gamma_q$ and $|D| < \delta^2/2$. Then $D \cup F_q$ is homeomorphic to S^2 and $|D \cup F_q| < \delta^2/2 + \delta^2/2 = \delta^2$, and hence by Lemma 1 we know that $D \cup F_q = \partial U$, U open, with $\bar{U}$ homeomorphic to $\mathbf{B}$. Let Λ be the component of Σ containing D, and consider the possibility that $\Lambda \sim D \subset U$. Since $\bar{U} \approx \mathbf{B}$, it would follow that $\Lambda \subset \Sigma_0$. Since we assume (3.17), this is impossible. Hence we must have (again using (3.17))

$$(\Sigma \sim D) \cap U = \varnothing. \tag{3.19}$$

Writing $\Sigma_* = (\Sigma \sim D) \cup F_q$ and $F_{q,\varepsilon} = \{x \in N\colon \operatorname{dist}(x, F_q) < \varepsilon\}$, for each $\varepsilon > 0$, we can select a continuous isotopy $\gamma = \{\gamma_t\}_{0 \le t \le 1}$ such that $\gamma_t(F_{q,\varepsilon}) \subset F_{q,\varepsilon}$, $\gamma_t(x) \equiv x$, $x \notin F_{q,\varepsilon}$, $|\Sigma_* \cap F_{q,\varepsilon}| \le (\gamma_1(\Sigma_* \cap F_{q,\varepsilon})| \le |\Sigma_* \cap F_{q,\varepsilon}| + \varepsilon$, $\gamma_1(\Sigma_* \cap F_{q,\varepsilon}) \cap \partial A = \varnothing$, and $\hat{\Sigma}_* \equiv \gamma_1(\Sigma_*) \in \mathcal{C}_1$. Then we have, for ε small enough,

(i) $\hat{\Sigma}_* \in \mathcal{G}(\Sigma)$, $\hat{\Sigma}_* \cap \partial A = \bigcup_{j=1}^{q-1} \Gamma_j$,

(ii) $|(\hat{\Sigma}_* \sim \Sigma) \cup (\Sigma \sim \hat{\Sigma}_*)| < |D| + |F_q| + \varepsilon$,

(iii) $|\hat{\Sigma}_*| < |\Sigma| + |F_q| - |D| + \varepsilon$.

Notice that (iii) is equivalent to

(iii)′ $E(\hat{\Sigma}_*) < E(\Sigma) + |F_q| - |D| + \varepsilon$.

Taking $\varepsilon \le |F_q|$, we deduce from (i), (ii) and Remark (3.11) that $\hat{\Sigma}_*$ is strongly γ^*-irreducible, where

$$\begin{aligned}\gamma^* &= \gamma/4 + E(\Sigma) + 4\sum_{j=1}^{q} |F_j| - \big(|D| + 2|F_q|\big)\\ &= \gamma/4 + E(\Sigma) + 4\sum_{j=1}^{q-1} |F_j| + 2|F_q| - |D|.\end{aligned}$$

By (iii)′ we thus have $\gamma^* \ge \gamma/4 + E(\hat{\Sigma}_*) + 4\Sigma_{j=1}^{q-1} |F_j|$. Furthermore, by (iii)′,

$$E\big(\hat{\Sigma}_*\big) \le E(\Sigma) + 2|F_q| \le \gamma/2 - 2\sum_{j=1}^{q} |F_j| + 2|F_q| = \gamma/2 - 2\sum_{j=1}^{q-1} |F_j|.$$

Thus $\hat{\Sigma}_*$ satisfies the inductive hypotheses with $q-1$ in place of q. Hence there must be pairwise disjoint discs $\tilde{\Delta}_1, \dots, \tilde{\Delta}_p$ contained in $\hat{\Sigma}_*$ with $\partial\tilde{\Delta}_j \subset \partial A$,

(3.20)

$$\left(\bigcup_{j=1}^{p} \tilde{\Delta}_j\right) \cap (A \sim \partial A) = \hat{\Sigma}_* \cap (A \sim \partial A), \ \sum_{j=1}^{p} |\tilde{\Delta}_j| \leq \sum_{j=1}^{q-1} |F_j| + E(\hat{\Sigma}_*)$$

and with $\tilde{\psi}_1(\tilde{\Delta}_j) \sim \partial\tilde{\Delta}_j \subset A \sim \partial A$ for some isotopy $\tilde{\psi} = \{\tilde{\psi}_t\}_{0 \leq t \leq 1}$ which holds a neighborhood of the set $\hat{\Sigma}_* \sim \bigcup_{j=1}^{p}(\tilde{\Delta}_j \sim \partial\tilde{\Delta}_j)$ fixed. It follows that there are pairwise disjoint discs $\Delta_1, \dots, \Delta_p \subset \Sigma_* \equiv (\Sigma \sim D) \cup F_q$ with

$$\sum_{j=1}^{p} |\Delta_j| \leq \sum_{j=1}^{q-1} |F_j| + E(\Sigma) + |F_q| - |D|, \tag{3.21}$$

$$\left(\bigcup_{j=1}^{p} \Delta_j\right) \cap (A \sim \partial A) = \Sigma_* \cap (A \sim \partial A), \partial\Delta_j = \partial\tilde{\Delta}_j$$

and $\psi_1(\Delta_j) \sim \partial\Delta_j \subset A \sim \partial A$ for some isotopy $\psi = \{\psi_i\}_{0 \leq t \leq 1}$ which holds a neighborhood of the set $\Sigma_* \sim \bigcup_{j=1}^{p}(\Delta_j \sim \partial\Delta_j) \sim F_{q,\varepsilon}$ fixed. (In fact we can take $\Delta_j = \gamma_1^{-1}(\tilde{\Delta}_j)$ and $\psi = \tilde{\psi}_* \gamma$.)

Now, to proceed, let U (with $\partial U = D \cup F_q$) be as above, and let $\beta = \{\beta_t\}_{0 \leq t \leq 1}$ be a continuous isotopy such that $\beta_t(U) \subset U$, $\beta_t|_{\Sigma \sim D} \equiv \mathbf{1}_{\Sigma \sim D}$ and $\beta_1(D) = F_q$. (Notice that such an isotopy exists because $\bar{U} \approx \mathbf{B}$ and because $(\Sigma \sim D) \cap U = \varnothing$ by virtue of (3.19).) Consider the following two cases:

Case (i): $F_q \subset \bigcup_{j=1}^{p} \Delta_j$

Case (ii): $F_q \not\subset \bigcup_{j=1}^{p} \Delta_j$.

In Case (i) we select the discs $D_1, \dots, D_p$ by taking $D_{j_0} = (\Delta_{j_0} \sim F_q) \cup D$ for the (unique) j_0 such that $F_q \subset \Delta_{j_0}$, and we select $D_j = \Delta_j$ for all $j \neq j_0$. Also, we define a continuous isotopy $\tilde{\varphi} = \{\tilde{\varphi}_t\}_{0 \leq t \leq 1}$ by $\tilde{\varphi} = \psi * \beta$; by smoothing $\tilde{\varphi}$ we obtain an isotopy φ satisfying the required conditions. (Here $\psi * \beta$ is defined by $\psi * \beta(t, x) = \beta(2t, x)$ if $0 \leq t \leq \frac{1}{2}$, and $\psi * \beta(t, x) = \psi(2t - 1, \beta(1, x))$ if $\frac{1}{2} < t \leq 1$.)

In Case (ii) we define discs $D_1, \dots, D_{p+1}$ by setting $D_j = \Delta_j$, $j = 1, \dots, p$, and $D_{p+1} = D$. In this case we define a continuous isotopy $\tilde{\varphi}$ by setting $\tilde{\varphi} = \hat{\beta} * (\psi * \beta)$, where $\hat{\beta} = \{\hat{\beta}_t\}_{0 \leq t \leq 1}$ is a smooth isotopy such that $\hat{\beta}_t(x) = x$ for all $(x, t) \in (\Sigma \sim D) \times [0, 1]$ and such that $\hat{\beta}_1(F_q)$ is a disc $\tilde{D} \subset A$ with $\partial\tilde{D} = \partial D$, $\tilde{D} \cap \partial A = \partial\tilde{D}$, and $\tilde{D} \cap \psi_t(\Sigma_*) = \Gamma_q$ for all $t \in [0, 1]$. Now we claim that in Case (ii) there is a neighborhood W of $\partial D (= \Gamma_q)$ such that $W \cap D \subset A$. Otherwise we would have W with $W \supset \partial D$ and $W \cap (\Sigma \sim D) \subset A \sim \partial A$, and this would imply $F_q \subset \bigcup_{j=1}^{p} \Delta_j$ by virtue of (3.21), thus contradicting the fact that we are in Case (ii). By smoothing $\tilde{\varphi}$ we then again obtain the required isotopy φ.

In each of the above cases we have, by (3.21), that

$$\sum_{j=1}^{p} |\Delta_j| \leq \sum_{j=1}^{q-1} |F_j| + E(\Sigma) + |F_q| - |D|$$

and hence

$$\begin{aligned} \sum_{j} |D_j| &\leq \sum_{j=1}^{q-1} |F_j| + E(\Sigma) + |F_q| - |D| + |D| \\ &= \sum_{j=1}^{q} |F_j| + E(\Sigma). \end{aligned}$$

This completes the proof by induction.

We would of course like to apply Theorem 2 to the minimizing sequence $\{\Sigma_k\}$ of Section 1. However, the theorem is not directly applicable because Σ_k is not necessarily strongly γ-irreducible for any γ. However, we now show that there is a $\gamma_0 > 0$ such that, after γ_0-reduction, Σ_k yields a strongly γ_0-irreducible $\tilde{\Sigma}_k$ with $\lim \mathrm{v}(\tilde{\Sigma}_k) = \lim \mathrm{v}(\Sigma_k)$. (Notation as in § 1.)

To see this, consider $0 < \gamma < \delta^2/9$. For $q = 1, 2, \ldots,$ let $k_q(\gamma)$ be the *largest* integer such that there exists $\Sigma_q^{(j)} \in \mathcal{C}_1$, $j = 1, \ldots, k_q(\gamma)$, with

$$\Sigma_q^{(k_q(\gamma))} \underset{\gamma}{<} \cdots \underset{\gamma}{<} \Sigma_q^{(2)} \underset{\gamma}{<} \Sigma_q^{(1)} = \Sigma_q; \tag{3.22}$$

for convenience set $k_q(\gamma) = 1$ if Σ_q is strongly γ-irreducible. Then (noting that $k_q(\gamma)$ is bounded independent of q and γ by (3.9)), we let $l(\gamma)$ be the non-negative integer defined by

$$l(\gamma) = \limsup_{q \to \infty} k_q(\gamma). \tag{3.23}$$

Evidently l is an increasing function of γ; hence since l is integer-valued, there is $\gamma_0 \in (0, \delta^2/9)$ such that

$$l(\gamma) \equiv l(\gamma_0) \text{ for all } \gamma \in (0, \gamma_0].$$

Now for each $n \geq \gamma_0^{-1}$ there is a $q_n \geq n$ such that $k_{q_n}(1/n) = l(1/n) (= l(\gamma_0))$. We set

$$\tilde{\Sigma}_n = \Sigma_{q_n}^{k_{q_n}(1/n)}.$$

Evidently $\tilde{\Sigma}_n$ is then strongly γ_0-irreducible for all sufficiently large n, because otherwise we would have, for infinitely many n, that

$$\Sigma_{q_n} = \Sigma_{q_n}^{(1)} \underset{\frac{1}{n}}{>} \Sigma_{q_n}^{(2)} \underset{\frac{1}{n}}{>} \cdots \underset{\frac{1}{n}}{>} \Sigma_{q_n}^{k_{q_n}\left(\frac{1}{n}\right)} \underset{\gamma_0}{>} \hat{\Sigma}_n$$

for some $\hat{\Sigma}_n \in \mathcal{C}_1$, thus implying that

$$k_{q_n}(\gamma_0) \geq l(\gamma_0) + 1$$

for infinitely many n, which contradicts the definition of $l(\gamma_0)$ in (3.23).

Thus, relabeling if necessary, we can assert that $\tilde{\Sigma}_n$ is strongly γ_0-irreducible for all n, and that (by (3.10))

$$|\left(\tilde{\Sigma}_k \sim \Sigma_{q_k}\right) \cup \left(\Sigma_{q_k} \sim \tilde{\Sigma}_k\right)| \leq c/k \tag{3.24}$$

where c is independent of k. Thus

$$\lim \mathrm{v}\left(\tilde{\Sigma}_k\right) = \lim \mathrm{v}(\Sigma_k). \tag{3.25}$$

Further, since Σ_{q_k} can be recovered, up to isotopy, by cutting out discs of $\tilde{\Sigma}_k$ and adding arbitrarily thin tubes, we evidently have

$$|\tilde{\Sigma}_k| \leq \inf_{\Sigma \in \mathfrak{I}(\tilde{\Sigma}_k)} |\Sigma| + \varepsilon_k \tag{3.26}$$

(where $\varepsilon_k \to 0$ as $k \to \infty$), by virtue of (3.24) and the fact that the original sequence $\{\Sigma_k\}$ is minimizing.

(3.27) *Remark*. We are now in a position to describe more precisely the relation between the minimizing sequence $\{\Sigma_k\}$, the surfaces $\Sigma^{(1)}, \dots, \Sigma^{(R)}$, and the integers $n_1, \dots, n_R$ of Theorem 1. We shall in fact prove that, if $\tilde{\Sigma}_k$ is obtained by γ_0-reduction of Σ_{q_k} as described above, then, for all sufficiently large k,

$$\tilde{\Sigma}_k \in \mathfrak{I}(S_k), \qquad S_k = \bigcup_{j=0}^{R} S_k^{(j)}, \tag{3.28}$$

where $\lim |S_k^{(0)} \cap K| = 0$ for each compact $K \subset N$, $S_k^{(0)} \cap (\bigcup_{j=1}^{R} S_k^{(j)}) = \varnothing$, and, for k sufficiently large and $j = 1, \dots, R$, $S_k^{(j)}$ is defined by

(3.29)

$$S_k^{(j)} = \begin{cases} \bigcup_{r=1}^{m_j} \{x \in N\colon \mathrm{dist}(x, \Sigma^{(j)}) = r/k\} & \text{if } n_j = 2m_j \text{ is even} \\ \Sigma^{(j)} \cup \left(\bigcup_{r=1}^{m_j} \{x \in N\colon \mathrm{dist}(x, \Sigma^{(j)}) = r/k\} \right) & \text{if } n_j = 2m_j + 1 \text{ is odd.} \end{cases}$$

It is important to note here that in case $\Sigma^{(j)}$ is two-sided (i.e. has a smooth unit normal), the set $\{x \in N\colon \mathrm{dist}(x, \Sigma^{(j)}) = \varepsilon\}$ (for ε sufficiently small) has two components, each isotopic to $\Sigma^{(j)}$; on the other hand, if $\Sigma^{(j)}$ is *one-sided* then this set is a smooth connected two-sided surface which gives a double cover for $\Sigma^{(j)}$ via the nearest-point projection.

Because of this last point, and because of Remark (3.1), we see that if n_i is *odd* and if the Σ_k are all two-sided in N, then $\Sigma^{(i)}$ is two-sided in N. Nevertheless, it can of course happen that some of the $\Sigma^{(i)}$ are one-sided in N, even if each Σ_k is two-sided and N is orientable.

4. Minimizing sequences of discs in N

Here we wish to explain the straightforward technical modifications needed to extend the interior regularity results of [AS] to the present case when N is a homogeneously regular Riemannian 3-manifold as in Section 1. (Only the case $N = \mathbf{R}^3$ was specifically considered in [AS].)

It is emphasized that here we shall not need any boundary regularity theory; in fact only Sections 3–6 of [AS] are needed. (Lemma 2 above will be used to replace Lemma 2 and Corollary 1 of § 2 of [AS].)

We shall need the following technical lemma. In this lemma d_U denotes the distance function of U defined by

$$d_U(x) = \operatorname{dist}(x, U), \qquad x \in N.$$

We also let

$$U(s) = \{x \in N: d_U(x) < s\} \qquad \text{for each } s > 0.$$

LEMMA 3. *Suppose $U \subset G_{\rho_0/2}(x_0)$, $\overline{U} \approx \mathbf{B}$, and U is convex in the strong sense that d_U is a convex function on $\{x \in N\colon\ d_U(x) < \theta\rho_0\}$*(1) *for some $\theta \in (0, 1/2)$, and let $\beta \geq 1$ be a constant such that, for each $s \in (\theta\rho_0/2, \theta\rho_0)$,*

$$\min\{|E|, |\partial U(s) \sim E|\} \leq \beta(\operatorname{length}(\partial E))^2 \tag{4.3}$$

whenever E is a disc contained in $\partial U(s)$.

If $\delta_1 = \min\{\delta, (1 + 64c)^{-1}\theta\beta^{-1/2}\}$ (c, δ as in Lemma 1), if M is any smooth disc with $\partial M \subset N \sim U$ and with M intersecting ∂U transversally, if

$$|\partial U| + |M| \leq \delta_1^2\rho_0^2/16 \tag{4.4}$$

and if Λ is any component of $M \sim U$ with $\partial M \cap \Lambda = \varnothing$, then there is a unique $K_\Lambda \subset N \sim U$ such that

$$\operatorname{vol}(K_\Lambda) \leq c\delta_1^3\rho_0^3, \qquad \partial K_\Lambda = \Lambda \cup F, \tag{4.5}$$

where $F \subset \partial U$ is a compact (not necessarily connected) surface with $\partial F = \partial\Lambda$ and

$$|F| < |\Lambda \cap U(\theta\rho_0)|. \tag{4.6}$$

(1) By this we mean that if φ is a geodesic in $U(\theta\rho_0)$ which is parametrized by arc length s, $a \leq s \leq b$, then $d_U(\varphi(s))$ is a convex function of s, $a \leq s \leq b$.

Proof. First find a closed $F_0 \subset \partial U$ with $\partial F_0 = \partial \Lambda$. (Evidently such an F_0 exists by virtue of the fact that $\partial U \approx \mathbf{S}^2$, although of course F_0 may not be connected.) Then (4.4) implies that $|F_0| + |\Lambda| \le \delta_1^2 \rho_0^2 \le \delta^2 \rho_0^2$ and hence by Lemma 1 there is a compact W with

$$\operatorname{vol}(W) \le c\delta_1^3 \rho_0^3, \qquad \partial W = F_0 \cup \Lambda.$$

Then we set $K_\Lambda = W \sim U$. Evidently then (4.5) holds with either $F = F_0$ or $F = \partial U \sim F_0$. The uniqueness of K_Λ is evident in view of (4.5) and (1.1).

To proceed, define, for $t \ge 0$,

$$F_t = K_\Lambda \cap \{x : d_U(x) = t\}, \qquad E_t = \{x \in \Lambda : d_U(x) < t\}.$$

Now by the convexity of d_U on $U(\theta\rho_0)$, we know that $\Delta d_U \ge 0$ in $U(\theta\rho_0)$. Hence for $0 \le t_1 < t_2 \le \theta\rho_0$ we can use the divergence theorem (applied to the vector field $\operatorname{grad} d_U$ on $\{x \in K_\Lambda : t_1 \le d_U(x) \le t_2\}$) to give

$$|F_{t_1}| - |F_{t_2}| \le \int_{\bar{E}_{t_2} \sim E_{t_1}} |\langle \nu, \operatorname{grad} d_U \rangle|,$$

where ν is the unit normal of Λ pointing out of K_Λ. Since $|\langle \nu, \operatorname{grad} d_U\rangle| \le 1$ and since strict inequality must hold on a set of positive 2-dimensional measures in $E_{t_2} \sim E_{t_1}$ unless $E_{t_2} = E_{t_1}$, we deduce that

$$|F_{t_1}| - |F_{t_2}| \le |\bar{E}_{t_2}| - |E_{t_1}|, \tag{4.7}$$

for all $0 \le t_1 < t_2 \le \theta\rho_0$, and that for such t_1, t_2

$$|F_{t_1}| - |F_{t_2}| < |\bar{E}_{t_2}| - |E_{t_1}| \tag{4.7$'$}$$

in case $E_{t_2} \ne E_{t_1}$.

We also have by (4.5) and the co-area formula that

$$\int_0^{\theta\rho_0} |F_s|\, ds \le \operatorname{vol}(K_\Lambda \cap U(\theta\rho_0)) \le c\delta_1^3\rho_0^3,$$

and hence

$$|F_t| \le 4c\theta^{-1}\delta_1^3\rho_0^2$$

for a set of $t \in [0, \theta\rho_0]$ of Lebesgue measure $\ge \frac{3}{4}\theta\rho_0$. But then, since $\delta_1 < (64c)^{-1}\theta$, we have from (4.4) and (4.7) that

$$|F_t| \le \delta_1^2\rho_0^2/16 \quad \text{for all } t \in [0, 3\theta\rho_0/4]. \tag{4.8}$$

Also, setting $t_1 = 0$ and $t_2 = t \in (0, \theta\rho_0]$ in (4.7)$'$, we have

$$|F| - |\bar{E}_t| < |F_t|, \qquad t \in (0, \theta\rho_0]. \tag{4.9}$$

If $|F_{\theta\rho_0}| = 0$ we then have (4.6) by setting $t = \theta\rho_0$. If $|F_{\theta\rho_0}| \ne 0$, we argue as follows. By (1.1) together with the fact that $U(\theta\rho_0/2)$ contains a geodesic ball of radius $\theta\rho_0/2$, we may use the isoperimetric inequality (in $\mathbf{R}^3$) to give

$$|\partial U(t)| \geq (\mathrm{vol}(U(t)))^{2/3}, \qquad |\partial U(t)| \geq (\mathrm{vol}(U(t)))^{2/3} \geq \theta^2\rho_0^2/8,$$

$t \in [\theta\rho_0/2, \theta\rho_0]$. Then by (4.8) we have

$$|F_t| \leq \tfrac{1}{2}|\partial U(t)| \quad \text{for all } t \in [\theta\rho_0/2, 3\theta\rho_0/4],$$

and it follows from (4.3) and the co-area formula that, almost everywhere $t \in [\theta\rho_0/2, 3\theta\rho_0/4]$,

$$|F_t| \leq \beta|\partial F_t|^2 = \beta|\partial E_t|^2 \leq \beta\left(\frac{d}{dt}|E_t|\right)^2. \tag{4.10}$$

Thus (4.9) implies

(4.11)

$$|F| - |E_t| \leq \beta\left(\frac{d}{dt}(|F| - |E_t|)\right)^2 \text{ almost everywhere,} \qquad t \in [\theta\rho_0/2, \theta\rho_0].$$

By integration over $[\theta\rho_0/2, 3\theta\rho_0/4]$ (using the fact that $\sqrt{|F| - |E_t|}$ is a decreasing function), we then have

$$\sqrt{|F| - |E_{\theta\rho_0/2}|} - \sqrt{|F| - |E_{3\theta\rho_0/4}|} \geq \beta^{-1/2}\theta\rho_0/4$$

provided that $|F| > |E_{3\theta\rho_0/4}|$. However, since $\sqrt{|F|} \leq \delta_1\rho_0$ (by (4.4)), this is impossible by the choice of δ_1. Thus $|F| \leq |E_{3\theta\rho_0/4}| < |E_{\theta\rho_0}|$ since $|F_{\theta\rho_0}| \neq 0$.

With Lemma 3 proved, it is now elementary to modify the proof of the Replacement Theorem (Theorem 1 of [AS]) to the present manifold setting:

Firstly the hypothesis in [AS: Theorem 1] that U is convex is replaced by the hypothesis that U is as in Lemma 3 above. We also need the hypothesis that U and the disc M under consideration satisfies (4.4), and the hypothesis (iii) of [AS, Theorem 1] is replaced by the hypothesis

$$\partial M \sim \partial U \text{ is not contained in any } K_\Lambda, \tag{4.12}$$

where Λ is any component of $M \sim U$ with $\partial M \cap \Lambda = \varnothing$ and K_Λ is as in Lemma 3 above.

We would also point out a misprint in the statement of Theorem 1 of [AS]: the equality $(N \sim M) \cap \partial U = \varnothing$ in (iii) should read $(N \sim M) \cap U = \varnothing$ (so that the inward pointing co-normal of ∂M points *into* U at points of $\partial M \cap \partial U$).

With these modified hypotheses, Theorem 1 of [AS] carries over directly to the present setting. The proof is essentially unchanged except that (4.6) of Lemma 3 is used in place of inequality (3.1) of [AS].

The filigree lemma (Lemma 3 of [AS]) also directly generalizes to the following:

In the present setting we take the same hypothesis, except that each Y_t is required to satisfy the hypotheses of the set U of Lemma 3 above (with constants β, θ independent of t). Also, (4.12) is required to hold with $U = Y_t$ for every $t \in (0, 1)$ and the hypothesis that ∂M is to be contained in the unbounded component of $\mathbf{R}^3 \sim (\overline{Y}_t \cup (M \sim \partial M))$ is replaced by the requirement that (4.12) holds for $U = Y_t$ for every $t \in (0, 1)$. Then the proof of the filigree lemma in the present manifold setting is essentially unchanged.

We can now apply all the arguments of Sections 5, 6 of [AS] (using the modified replacement and filigree lemmas as described above, and using Lemma 2 above in place of [AS, Lemma 2, Corollary 1]) in order to establish interior regularity for varifold limits of minimizing sequences of discs in N.

The reader may wish to note that in those parts of the argument relating to homothetic expansion, it is convenient to assume that N is (at least locally) isometrically embedded in some Euclidean space. We also remark that it is not necessary to use any analogue of the convex hull property (Appendix A of [AS]) because here we shall be concerned only with interior regularity. We do need the fact that if M_1, M_2 are C^2 minimal surfaces in N which in terms of suitable local coordinates x^1, x^2, x^3 for N are expressed as

$$x^3 = u_1(x^1, x^2), \qquad x^3 = u_2(x^1, x^2)$$

for $(x^1, x^2) \in \Omega \subset \mathbf{R}^2$ (Ω open and connected), and if $u_1 \geq u_2$ in Ω, then $u_1 = u_2$ at some point of Ω implies that $u_1 \equiv u_2$ in Ω. This is readily seen from the fact that the difference $\varphi = u_1 - u_2$ satisfies a uniformly elliptic equation of the form $D_i(a_{ij}D_j\varphi) + c\varphi = 0$, and hence the required identity follows from the Harnack inequality for such equations.

5. Convergence of the minimizing sequence $\{\tilde{\Sigma}_k\}$

In this section we let $\{\tilde{\Sigma}_k\} \subset \mathcal{C}_1$ be any strongly γ_0-irreducible sequence such that $|\tilde{\Sigma}_k| + \operatorname{genus}(\tilde{\Sigma}_k)$ is bounded, such that (3.26) holds, and such that $\lim \mathrm{v}(\tilde{\Sigma}_k)$ exists. (Thus the discussion here is certainly applicable to the sequence $\{\tilde{\Sigma}_k\}$ constructed in § 3.) Let $\hat{\Sigma}_k$ be obtained by deleting all components Λ of $\tilde{\Sigma}_k$ such that there is a $K \approx \mathbf{B}$ with $\Lambda \subset K$ and $\tilde{\Sigma}_k \cap \partial K = \varnothing$. By virtue of Remark (3.14) we know that (3.26) continues to hold for $\hat{\Sigma}_k$ and that $\mathrm{v}(\hat{\Sigma}_k)$ has the same limit as $\mathrm{v}(\tilde{\Sigma}_k)$. Furthermore it is easy to check (again using Remark (3.14)) that $\hat{\Sigma}_k$ is strongly $(3\gamma_0/4)$-irreducible for all sufficiently large k.

Let $V = \lim \mathrm{v}(\hat{\Sigma}_k)$ $(= \lim \mathrm{v}(\tilde{\Sigma}_k))$ and let $x_0 \in \operatorname{spt}\|V\|$. By virtue of (3.26) we can apply the reasoning of Section 1 to deduce that (1.9) holds for V. Now by the co-area formula we have

$$\int_{\rho-\sigma}^{\rho} \operatorname{length}\big(\tilde{\Sigma}_k \cap \partial G_s(x_0)\big)\, ds \le \big| \tilde{\Sigma}_k \cap \big(\bar{G}_\rho(x_0) \sim G_{\rho-\sigma}(x_0)\big) \big| \tag{5.1}$$

which holds almost everwhere, $\rho \in (0, \rho_0)$, $\sigma \in (0, \rho)$. Taking $\sigma = \rho/2$, (1.9) then gives

$$\rho^{-1} \int_{\rho/2}^{\rho} \operatorname{length}\big(\tilde{\Sigma}_k \cap \partial G_s(x_0)\big)\, ds \le c\rho \tag{5.2}$$

for all sufficiently large k, where c depends only on μ and any upper bound for $\rho^{-2} \| V \|(G_{\rho_0}(x_0))$. Hence we can find a sequence $\{\rho_k\} \subset (3\rho/4, \rho)$ such that $\tilde{\Sigma}_k$ intersects $\partial G_{\rho_k}(x_0)$ transversally and such that

$$\operatorname{length}\big(\tilde{\Sigma}_k \cap \partial G_{\rho_k}(x_0)\big) \le c\rho \le c\eta\rho_0, \tag{5.3}$$

for all sufficiently large k, provided $\rho \le \eta\rho_0$, where for the moment $\eta \in (0,1)$ is arbitrary. If η is sufficiently small (depending only on γ_0 and $\|V\|(N)$) we see from (5.3) that Theorem 2 is applicable with $\tilde{\Sigma}_k$ in place of Σ (with $\gamma_0/2$ in place of γ and with $G_{\rho_k}(x_0)$ in place of A). Then there are discs $D_k^{(1)}, \dots, D_k^{(q_k)} \subset \tilde{\Sigma}_k$ and isotopies $\varphi^{(k)} = \{\varphi_t^{(k)}\}_{0 \le t \le 1}$ of N such that

$$\begin{cases} \partial D_k^{(j)} \subset \partial G_{\rho_k}(x_0), \quad \hat{\Sigma}_k \cap G_{\rho_k}(x_0) = \left(\dot{\bigcup\limits_{j=1}^{q_k}} D_k^{(j)} \right) \cap G_{\rho_k}(x_0) \\ \varphi_1^{(k)}\big(D_k^{(j)}\big) \sim \partial D_k^{(j)} \subset G_{\rho_k}(x_0), \quad \sum\limits_{j=1}^{q_k} | D_k^{(j)} | \le c\rho^2 \le c\eta^2\rho_0^2, \end{cases} \tag{5.4}$$

where c is independent of k, η, ρ. Since $| D_k^{(j)} | \le c\eta^2\rho^2$, we know that (for η sufficiently small), by the modified replacement lemma described in Section 4, there are disks $\tilde{D}_k^{(j)}$ with

$$\partial \tilde{D}_k^{(j)} = \partial D_k^{(j)}, \qquad \tilde{D}_k^{(j)} \sim \partial D_k^{(j)} \subset G_{\rho_k}(x_0) \tag{5.5}$$

$$| \tilde{D}_k^{(j)} | \le | D_k^{(j)} | . \tag{5.6}$$

Now by (5.4), Lemma 2, (5.5), (5.6) and (3.26) (with $\hat{\Sigma}_k$ in place of $\tilde{\Sigma}_k$), we deduce

$$| D_k^{(j)} | \le | \tilde{D}_k^{(j)} | + \varepsilon_{k,j} \le \inf\nolimits_{\mathfrak{D}_{j,k}} | \Delta | + 2\varepsilon_{k,j}, \qquad j = 1, \dots, q_k, \tag{5.7}$$

where $\mathfrak{D}_{k,j}$ denotes the set of all discs in N with boundary $\partial D_k^{(j)}$, and where $\sum_{j=1}^{q_k} \varepsilon_{k,j} \to 0$ as $k \to \infty$.

By (5.4) and (5.7) we may use the modified filigree lemma, as described in Section 4, to infer that there is an l, independent of k, such that for all sufficiently large k

$$| D_k^{(j)} \cap G_{\rho/2}(x_0) | \ge 4\varepsilon_{k,j} \tag{5.8}$$

for at most l integers $j \in \{1,\ldots,q_k\}$. Thus, relabeling if necessary, we know that

$$(5.9)\qquad \begin{cases} \left(\lim \mathrm{v}\left(\sum_{j=1}^{l} D_k^{(j)}\right)\right) \llcorner G_{\rho/2}(x_0) = V \llcorner G_{\rho/2}(x_0) \\ \left(\lim \mathrm{v}\left(\sum_{j=l+1}^{q_k} D_k^{(j)}\right)\right) \llcorner G_{\rho/2}(x_0) = 0. \end{cases}$$

Now let $\{k'\} \subset \{k\}$ be a subsequence such that $\lim \mathrm{v}(D_{k'}^{(j)})$ exists for each $j = 1,\ldots,l$. In view of (5.9) we can now use the theory of Sections 5, 6 of [AS] (modified as described in § 4 above and applied separately to each of the sequences $\{D_{k'}^{(j)}\}$, $j = 1,\ldots,l$) in order to deduce that there are integers $m_1,\ldots,m_P$ and stable properly embedded minimal surfaces $M^{(1)},\ldots,M^{(P)}$ in $G_{\rho/4}(x_0)$ such that

$$(5.10)\qquad V \llcorner G_{\rho/4}(x_0) = \sum_{j=1}^{P} m_j \mathrm{v}(M^{(j)}),$$

where $V = \lim \mathrm{v}(\hat{\Sigma}_k) = \lim \mathrm{v}(\tilde{\Sigma}_k)$. Since $x_0 \in \operatorname{spt}\|V\|$ is arbitrary, we thus deduce that there are positive integers R, $n_1,\ldots,n_R$, and complete stable properly embedded surfaces $\Sigma^{(1)},\ldots,\Sigma^{(R)}$ such that

$$(5.11)\qquad V = n_1 \mathrm{v}(\Sigma^{(1)}) + \cdots + n_R \mathrm{v}(\Sigma^{(R)}).$$

Notice that by letting $\sigma \downarrow 0$ in (1.9), we have that

$$\|V\| G_{\rho_0}(x_0) \geq c\rho_0^2 \quad \text{for any } x_0 \in \bigcup_{j=1}^{R} \Sigma^{(j)},$$

where c is independent of x_0. Since $\|V\|(N) < \infty$ it follows that each $\Sigma^{(j)}$ is compact (and $R = \infty$ is not possible) even when N is not compact (although of course we have used very strongly the fact that N is at least homogeneously regular). In particular we have proved Theorem 1, except for the assertions (1.4), (1.5).

We now want to show that the assertions of Remark 3.27 apply ((1.4), (1.5) of Theorem 1 will then follow directly).

Firstly, by (5.11) we note that there is a positive sequence $h_k \downarrow 0$ such that $\tilde{\Sigma}_k$ intersects each $\{x \in N\colon \operatorname{dist}(x, \Sigma^{(j)}) = h_k\}$ transversally and such that

$$(5.12)\qquad \operatorname{length}\left(\tilde{\Sigma}_k \cap \left\{x\colon \operatorname{dist}\left(x, \bigcup_{j=1}^{R} \Sigma^{(j)}\right) = h_k\right\}\right) + \left|\tilde{\Sigma}_k \cap \left\{x\colon h_k^{-1} > \operatorname{dist}\left(x, \bigcup_{j=1}^{R} \Sigma^{(j)}\right) > h_k\right\}\right| \to 0$$

as $k \to \infty$. Furthermore, by (5.11) we know that if $\rho \le \eta\rho_0$, where η depends only on γ_0, V, then $\|V\|(G_\rho(x_0) \sim G_{\rho-\sigma}(x_0)) \le c\rho\sigma$ for each $\sigma \in (0, \rho)$, and hence by (5.1) we have in place of (5.2) the more general inequality $\sigma^{-1}\int_{\rho-\sigma}^{\rho} \operatorname{length}(\tilde{\Sigma}_k \cap \partial G_s(x_0))\, ds \le c\rho$ for all sufficiently large k (depending on σ). It follows that (5.3), (5.4) hold for some sequence $\{\rho_n\}$ such that

$$\rho_k \uparrow \rho \quad \text{as } k \to \infty,$$

(5.13)

$$\operatorname{length}\left(\tilde{\Sigma}_k \cap \partial G_{\rho_k}(x_0) \cap \left\{x\colon \operatorname{dist}\left(x, \bigcup_{j=1}^{R} \Sigma^{(j)}\right) \ge h_k\right\}\right) \to 0 \quad \text{as } k \to \infty,$$

$$\tilde{\Sigma}_k \cap \partial G_{\rho_k}(x_0) \cap \left\{x\colon \operatorname{dist}\left(x, \bigcup_{j=1}^{R} \Sigma^{(j)}\right) = h_k\right\} = \varnothing.$$

By (5.13), (5.5) and an obvious area comparison argument, we know that for each $i = 1,\dots,l$ either $|D_k^{(i)}| \to 0$ as $k \to \infty$ or else, for all sufficiently large k,

(5.14) $\partial D_k^{(i)}$ *is not null-homotopic in* $\partial G_{\rho_k}(x_0) \cap \{x\colon \operatorname{dist}(x, \Sigma^{(i)}) < h_k\}$

and

(5.15)

$$(\pi - \varepsilon(\rho))\sigma^2 \le |D_k^{(i)} \cap G_\sigma(x_0)| \le (\pi + \varepsilon(\rho))\sigma^2, \quad \text{for any } \sigma \in [\rho/2, \rho],$$

where $\varepsilon(\rho) \downarrow 0$ as $\rho \downarrow 0$. ($\varepsilon(\rho)$ is independent of x_0, k, and the sequence ρ_k used in the construction of $D_k^{(i)}$.) Again relabeling if necessary, we see (from (5.9) and (5.15)) that we can take $l = n_j$ in the above discussion (so that (5.9) holds with $l = n_j$) and that (5.14), (5.15) are valid precisely for $i = 1,\dots,n_j$; this is subject to the stipulation that $x_0 \in \Sigma^{(j)}$ and that ρ is sufficiently small (in fact $\rho \le \eta\rho_0$, where η depends only on γ_0 and V).

On the other hand, if we select $\{k'\} \subset \{k\}$ so that $\lim \mathrm{v}(D_{k'}^{(i)})$ exists for each $i = 1,\dots,n_j$, then (5.11), together with (5.15) and the interior regularity theory of [AS, §§ 5, 6] (modified as in § 4 above and applied separately to each of the sequences $\{D_{k'}^{(i)}\}$), imply that

$$\lim \mathrm{v}\big(D_{k'}^{(i)}\big) = \mathrm{v}\big(\Sigma^{(j)} \cap G_\rho(x_0)\big), \qquad i = 1,\dots,n_j, \tag{5.16}$$

again provided $x_0 \in \Sigma^{(j)}$ and provided ρ is sufficiently small. By (5.11) this also gives

$$\lim \sum_{i=n_j+1}^{q_{k'}} |D_{k'}^{(i)}| = 0. \tag{5.17}$$

A simple argument using (5.16), (5.17), (5.14), (5.12), (5.11) and (5.1) then demonstrates that the sequence $\rho_k \uparrow \rho$ can in fact be selected so that all of the above continue to hold, together with the additional property that

$$(5.18) \qquad \lim \operatorname{length}\left(\partial D_{k'}^{(i)}\right) = \operatorname{length}\left(\Sigma^{(j)} \cap \partial G_\rho(x_0)\right), \qquad i = 1,\ldots,n_j$$

and

$$(5.19) \qquad \lim \operatorname{length}\left(\Sigma_{k'} \cap \partial G_{\rho_{k'}}(x_0) \sim \left(\bigcup_{i=1}^{n_j} \partial D_{k'}^{(k)}\right)\right) = 0.$$

Since this can be done for any $x_0 \in \Sigma^{(j)}$ ($j \in \{1,\ldots,R\}$), and any ρ sufficiently small depending only on γ_0 and V, it is now not difficult to conclude the proof of the assertions of Remark 3.27.

Specifically, we first note that for each sufficiently small σ, there are points $y_1,\ldots,y_P \in \bigcup_{j=1}^R \Sigma^{(j)}$ such that

$$(5.20) \qquad \overline{G}_{\sigma/8}(y_i) \cap \overline{G}_{\sigma/8}(y_j) = \varnothing, \qquad i \neq j,$$

$$(5.21) \qquad \bigcup_{j=1}^{R} \Sigma^{(j)} \subset \bigcup_{i=1}^{P} G_{\sigma/2}(y_i).$$

Let us agree that σ is small enough to ensure that $G_\sigma(y) \cap (\bigcup_{j=1}^R \Sigma^{(j)})$ is a single smooth two dimensional disc for each $y \in \bigcup_{j=1}^R \Sigma^{(j)}$, and also small enough to ensure that (5.12)–(5.19) hold with any $\rho < \sigma$, any $x_0 \in \Sigma^{(j)}$, and with any $(\gamma_0/2)$-irreducible minimizing sequence $\hat{\Sigma}_k^{(1)} \in \mathcal{I}(\Sigma_k)$ in place of $\hat{\Sigma}_k$. (Of course such a replacement of $\hat{\Sigma}_k$ by a new sequence $\hat{\Sigma}_k^{(1)}$ requires that we select a new sequence $\rho_k \uparrow \rho$; nevertheless, all of the above discussion applies.) Now select $\sigma_i \in (\sigma/2, \sigma)$ such that, with the notation $\hat{G}_i = G_{\sigma_i}(y_i) \cap (\bigcup_{j=1}^R \Sigma^{(j)})$,

$$(5.22) \qquad \begin{cases} \partial\hat{G}_i, \partial\hat{G}_j \text{ are either disjoint or intersect transversally}, i \neq j, \\ \partial\hat{G}_i \cap \partial\hat{G}_j \cap \partial\hat{G}_k = \varnothing \text{ for all triples of distinct integers } \{i, j, k \in 1,\ldots,P\}. \end{cases}$$

Thus each component of $(\bigcup_{j=1}^R \Sigma^{(j)}) \sim (\bigcup_{i=1}^P \partial\hat{G}_i)$ is a polygonal region with boundary consisting of closed arcs contained in the $\partial\hat{G}_i$, and each vertex on this polygonal boundary is contained in precisely two of the $\partial\hat{G}_i$. Suppose the totality of all these vertices is a collection $p_1,\ldots,p_Q$, and let $\tau < \sigma/2$ be such that $G_\tau(p_j) \cap G_\tau(p_k) = \varnothing$ for all $j \neq k$. By applying Remark 3.11 and the above discussion ((5.12)–(5.19)), with τ in place of ρ and p_m in place of x_0, it is quite easy to see that, since $\tau < \sigma$, we can modify $\hat{\Sigma}_k$ to give a $(\gamma_0/2)$-irreducible sequence $\hat{\Sigma}_k^{(1)} \in \mathcal{I}(\hat{\Sigma}_k)$ with $\lim \mathrm{v}(\hat{\Sigma}_k^{(1)}) = \lim \mathrm{v}(\tilde{\Sigma}_k)$ and with

$$(5.23) \qquad \hat{\Sigma}_k^{(1)} \cap G_{\tau/2}(p_m) = \left(\bigcup_{j=1}^{R} S_k^{(j)}\right) \cap G_{\tau/2}(p_m), \qquad m = 1,\ldots,Q,$$

where $S_k^{(j)}$ are as in Remark 3.27. Then applying (5.12)–(5.19) with $\hat{\Sigma}_k^{(1)}$ in place of $\hat{\Sigma}_k$, with $y_m \in \Sigma^{(j)}$ in place of x_0 and with σ_m in place of ρ, we obtain $\rho_k^m \uparrow \sigma_m$ and discs $\{D_k^{(i,m)}\}$, $m = 1,\ldots,Q$, $i = 1,\ldots,q_k$, such that, for each $m \in \{1,\ldots,P\}$ with $y_m \in \Sigma^{(j)}$,

$$(5.24)\quad \begin{cases} \hat{\Sigma}_k^{(1)} \cap G_{\rho_k^m}(y_m) = \bigcup_{m=1}^{q_k} D_k^{(i,m)} \cap G_{\rho_k^m}(y_m), \qquad \partial D_k^{(i,m)} \subset \partial G_{\rho_k^m}(y) \\ \mathbf{v}\big(D_k^{(i,m)}\big) \to \mathbf{v}\big(\Sigma^{(j)} \cap G_{\sigma_m}(y_m)\big), \quad i = 1,\ldots,n_j, \qquad \sum_{i=n_j+1}^{q_k} |D_k^{(i,m)}| \to 0 \\ \text{length}\big(\partial D_k^{(i,m)}\big) \to \text{length}\big(\Sigma^{(j)} \cap \partial G_{\sigma_m}(y_m)\big), \quad m = 1,\ldots,n_j, \\ \text{length}\left(\Sigma_k \cap \partial G_{\rho_k^m}(y_m) \sim \left(\bigcup_{i=1}^{n_j} \partial D_k^{(i,m)}\right)\right) \to 0, \end{cases}$$

and such that (5.14) holds with $D_k^{(i,m)}$ in place of $D_k^{(i)}$, $i = 1,\ldots,n_j$. In view of (5.23), (5.24), we can modify $\hat{\Sigma}_k^{(1)}$ slightly to give $\hat{\Sigma}_k^{(2)} \in \mathcal{I}(\hat{\Sigma}_k^{(1)})$ $(= \mathcal{I}(\hat{\Sigma}_k))$ such that $\lim \mathbf{v}(\hat{\Sigma}_k^{(2)}) = \lim \mathbf{v}(\tilde{\Sigma}_k)$ and such that (5.23), (5.24) hold with $\hat{\Sigma}^{(2)}$ in place of $\hat{\Sigma}^{(1)}$ and with $\rho_k^m \equiv \sigma_m$ for all sufficiently large k and for all $j = 1,\ldots,R$. It is then quite straightforward to show that $\hat{\Sigma}_k^{(2)} \in \mathcal{I}(\bigcup_{j=0}^R S_k^{(j)})$, with $S_k^{(j)}$ as in Remark 3.27.

Having established the assertions of Remark 3.27, it is now straightforward to check that (1.4) and (1.5) of Theorem 1 hold. Indeed (1.4) is a direct consequence of (3.27), and (1.5) is established as follows:

First, from the fact that (3.26) holds, it follows directly from (3.27) that each $\{S_k^{(j)}\}$ (for any $j \in \{1,\ldots,R\}$) satisfies

$$(5.25)\qquad |S_k^{(j)}| \le \inf_{\substack{\Sigma \in \mathcal{I}(S_k^{(j)}) \\ \Sigma \subset \{x:\, \text{dist}(x,\Sigma^{(j)}) < d/2\}}} |\Sigma| + \varepsilon_k, \qquad \varepsilon_k \to 0 \text{ as } k \to \infty,$$

where $d = \inf\{\text{dist}(x, y)\colon x, y$ lie in distinct components of $\bigcup_{i=1}^R \Sigma^{(i)}\}$ (and $d = \infty$ when $R = 1$). In case $\Sigma^{(j)}$ is two-sided in N it follows directly from (5.25) that

$$|\Sigma^{(j)}| = \inf_{\substack{\Sigma \in \mathcal{I}(\Sigma^{(j)}) \\ \Sigma \subset \{x:\, \text{dist}(x,\Sigma^{(j)}) < d/2\}}} |\Sigma|$$

and it is then standard that (1.5) holds in this case. In case $\Sigma^{(j)}$ is one-sided in N but each Σ_k is two-sided in N, we have clearly that each component of $\tilde{\Sigma}_k$ is two-sided in N, and hence each component of $S_k^{(j)}$ is two-sided in N (by (3.27)) (and hence, incidentally, n_j must be *even* by virtue of the definition of $S_k^{(j)}$ and the fact that $\Sigma^{(j)}$ is one-sided in N). Then if $T_k^{(j)}$ is any one of the components of

$S_k^{(i)}$, we have that $T_k^{(i)}$ is a double cover for $\Sigma^{(i)}$ with $\lim \mathrm{v}(T_k^{(i)}) = 2\mathrm{v}(\Sigma^{(i)})$ and (by (5.25))

$$(5.26) \qquad |T_k^{(i)}| \le \inf_{\substack{\Sigma \in \mathcal{I}(T_k^{(i)}) \\ \Sigma \subset \{x:\, \mathrm{dist}(x,\, \Sigma^{(i)}) < d/2\}}} |\Sigma| + \varepsilon_k', \qquad \varepsilon_k' \to 0 \text{ as } k \to \infty .$$

It is then again standard that (1.5) holds.

As a matter of fact, the above argument shows more generally that (1.5) holds whenever either $\Sigma^{(i)}$ is two-sided in N or when $\Sigma^{(i)}$ is one-sided in N and $n_i > 1$, because if $n_i > 1$ then $S_k^{(i)}$ must have at least one two-sided component $T_k^{(i)}$ as in (5.24).

6. Existence theorem for homogeneous regular manifolds with boundary

In this section, we generalize the main existence theorem to complete homogeneous regular manifolds N with $\partial N \neq \varnothing$. (A manifold with boundary is called homogeneous regular if it is a subdomain of some homogeneous regular manifold without boundary.) We assume that ∂N has non-negative mean curvature with respect to the outward normal. We may allow ∂N to be piecewise smooth if the intersecting angles between the boundary pieces are "convex." (See [MY2], Section 1.)

THEOREM 1′. *If N is a homogeneous regular manifold whose boundary has non-negative mean curvature with respect to the outward normal, then the same statement as in Theorem* 1 *holds.*

Proof. Let us first assume that ∂N is compact and smooth. As in the proof of Theorem 1 of [MY3], we can isometrically embed N into a complete homogeneous regular manifold $\tilde{N}$ (without boundary) as a subdomain. The manifold $\tilde{N}$ can be chosen to have the following properties:

(i) Any compact minimal surface (with boundary in N) of $\tilde{N}$ is a subset of N.

(ii) There is a diffeomorphism isotopic to identity which maps $\tilde{N}$ onto N and is equal to identity in the complement of a regular neighborhood of ∂N in N.

(iii) For any constant $c > 0$, there exists a compact set of $\tilde{N}$ so that any geodesic ball of radius c in the complement of this compact set is included in an open subset diffeomorphic to the ball.

It is clear that Theorem 1′ follows from Theorem 1 and these properties of $\tilde{N}$.

When ∂N is compact but piecewise smooth (see [MY3], Section 1), the approximation procedure used in [MY3] can be applied to the present situation.

Let us now study the case when ∂N is not compact. Let Ω_i be an increasing union of compact subdomains of N so that the distance between $\partial\Omega_i \setminus \partial N$ and $\partial\Omega_{i-1} \setminus \partial N$ tends to infinity as $i \to \infty$. By changing the metric in a neighborhood of $\partial\Omega_i \setminus \partial N$ suitably, we can find a sequence of metrics dS_i^2 on each Ω_i so that $dS_j^2 \mid \Omega_i = dS_{i+1}^2 \mid \Omega_i$ for $j \geq i + 1$ and each $\partial\Omega_i$ has non-negative mean curvature with respect to the outward normal with respect to dS_i^2.

If for some compact set $K \subset \Omega_i$, $\liminf |\Sigma_k \cap K| > 0$ and if $\Sigma_k \subset \Omega_i$ for some k, then we can apply our previous statement to find the minimal embedded surfaces $\Sigma_i^{(1)}, \Sigma_i^{(2)}, \ldots,$ and $\Sigma_i^{(k)}$ in Ω_i which satisfy all the conditions stated in Theorem 1.

Since the total area of these minimal surfaces is bounded by a constant independent of i, one can prove that on each compact subset of N, the minimal surfaces $\Sigma_i^{(j)}$ converge as $i \to \infty$. The rest of the arguments are the same as in the proof of Theorem 1.

THEOREM 1″. *If N is a complete manifold whose boundary has non-negative mean curvature with respect to the outward normal, then the results in Theorem 1 hold except that $S_{k'}$ need not have bounded diameter and $\Sigma^{(k)}$ need not be compact (but it has to have finite volume).*

Proof. It is basically the same as the one given in the proof of Theorem 1′.

7. A covering space of an irreducible orientable three-dimensional manifold is irreducible

In this section we will see how the Fundamental Existence Theorem 1 can be applied to get new information on the topology of three-dimensional manifolds. It has been conjectured that if N is a compact irreducible orientable three-dimensional manifold with infinite fundamental group, then the universal covering space $\tilde{N}$ of N is diffeomorphic to $\mathbf{R}^3$. This conjecture would be established if it could be shown that $\tilde{N}$ is irreducible and simply connected at infinity. In this section we shall prove that $\tilde{N}$ is irreducible. In order to prove this result, we shall need the following lemma.

LEMMA 4. *Let N be a three-dimensional Riemannian manifold. Let S_1 and S_2 be two distinct embedded two spheres in N which do not bound balls and which have least area among spheres with this property. Then S_1 and S_2 are disjoint.*

Proof. First suppose that S_1 and S_2 intersect transversely. In this case $\Gamma = S_1 \cap S_2$ consists of a collection of Jordan curves each of which is the boundary of exactly one disc on S_1 and another disc on S_2. Let $\mathfrak{D}$ be the collection of all such discs. Choose a $D \in \mathfrak{D}$ such that

$$|D| \leq |D'| \quad \text{for all } D' \in \mathfrak{D}.$$

Let $\gamma = \partial D$. Then after a possible change of indices, we may assume that $D \subset S_1$ and that the interior of D is disjoint from S_2. The curve γ bounds discs D_1 and D_2 on S_2. Now form

$$S_3 = D \cup D_1 \qquad \text{and} \qquad S_4 = D \cup D_2.$$

Note that if both S_3 and S_4 are boundaries of balls, then S_2 is also the boundary of a ball. So we may assume that one of them, say S_3, is not the boundary of a ball. Then

$$|S_3| = |D| + |D_1| \leq |D_1| + |D_2| = |S_2|,$$

and after a small perturbation of S_3 along γ we get an S_3^* which does not bound a ball and $|S_3^*| < |S_2|$. The existence of such an S_3^* contradicts the least area property of S_2. Therefore, in the case of transverse intersection, we may assume that $S_1 \cap S_2$ is empty.

In [MY1] it was shown that S_1 intersects S_2 transversely except in a finite number of points and that the intersection at the nontransverse points looks like a finite even number of rays intersecting at a point. The type of approximation procedure used in the proof of Theorems 5 and 6 in [MY1] shows that we may assume that S_1 and S_2 intersect transversely which reduces us to the case considered above. This completes the proof of the lemma.

THEOREM 3. *A covering space of an irreducible orientable three-dimensional manifold is irreducible.*

Proof. Since in the above theorem we shall be dealing with manifolds N which are not necessarily compact and may have boundary, we recall that a three-dimensional manifold N is called irreducible if every two sphere S^2 in N bounds a ball in N. Clearly, if some cover of N is not irreducible then the universal cover of N is also not irreducible. Hence we need only show that the universal cover of N is irreducible.

Suppose now that N is irreducible but its universal cover $\pi\colon \tilde{N} \to N$ is not irreducible. Let S^2 be a two sphere in $\tilde{N}$ which does not bound a ball. If S^2 does not bound a compact region of $\tilde{N}$, then it follows that S^2 is homotopically nontrivial in $\tilde{N}$ and hence that the second homotopy group of N is nontrivial. By the sphere theorem, there exists an embedded sphere S in N which does not bound a ball. Thus we may assume that such an S^2 is the boundary of a compact region R of $\tilde{N}$. Van Kampen's theorem actually shows that R is simply connected and hence R is a fake ball.

Let R be as above and π: $\tilde{N} \to N$ be the universal cover of N. Consider the compact set $X \subset N$ which is the union of the compact set $\pi(R)$ and all the balls B in N such that $\partial B \subset \pi(R)$. By a simple general position argument we may assume that both $\pi(R)$ and X are manifolds (possibly) with boundary. Since N is irreducible every two sphere in N bounds a ball. It follows from the definition of X and the sphere theorem [HJ] that the second homotopy group of X is zero and that every two-sphere in X is the boundary of a ball in X.

Let P: $\tilde{X} \to X$ be the universal cover of X. Since $\pi \mid \pi^{-1}(X)$: $\pi^{-1}(X) \to X$ is a covering space containing the fake ball R, the universal cover $\tilde{X}$ also contains a fake ball. Now choose a metric on X so that the boundary of X is convex and lift this metric to $\tilde{X}$.

Next consider a sequence of two spheres $C = \{S_k\}_{k=1}^{\infty}$ such that S_k is not the boundary of a standard ball and such that the areas of the spheres in C converge to the infimum of the areas of such spheres. (Note that such spheres exist or otherwise $\tilde{X}$ and hence $\tilde{N}$ are irreducible homotopy spheres.)

By Lemma 1 and an application of the covering transformations, we may assume that there is a fixed compact set C in $\tilde{X}$ which is the closure of a fundamental region of the covering P: $\tilde{X} \to X$ and such that $|S_k \cap C| > \varepsilon$ for some fixed $\varepsilon > 0$. Therefore we may assume from the Existence Theorem 1 and the fact that $\tilde{X}$ contains no projective planes that there exists a two sphere S_1 which does not bound a ball and which has least area with this property.

Now let τ: $\tilde{X} \to \tilde{X}$ be any nontrivial covering transformation. If $\tau(S_1)$ intersects S_1 nontrivially, then Lemma 3 shows that $\tau(S_1) = S_1$. Since the subgroup G of covering transformations which leave S_1 invariant acts freely as a group of isometries, G must be isomorphic to the trivial group or to Z_2. If G is isomorphic to Z_2, then S_1 descends to an embedded projective plane in N. As N is orientable and irreducible, N is then actually diffeomorphic to the three-dimensional projective plane and hence $\tilde{N} = \mathbf{S}^3$ which is irreducible. Therefore, we may assume that when τ is a nontrivial covering transformation, $\tau(S_1) \cap S_1 = \varnothing$.

Since $\tau(S_1)$ is disjoint from S_1 for all nontrivial covering transformations of the covering $\tilde{X} \to X$, the sphere S_1 descends to an embedded sphere S in X. Since N is irreducible, S bounds a ball B in N and by the definition of X this ball is contained in X. The ball B lifts to a ball $\tilde{B}$ in $\tilde{X}$ such that the boundary of $\tilde{B}$ is S_1. This contradicts the defining property of S as being a sphere in $\tilde{X}$ of least area with the property that it does not bound a ball. This contradiction shows that $\tilde{X} = \tilde{N}$ is an irreducible homotopy sphere.

Remark. When M is irreducible and is not orientable, the universal cover need not be irreducible as was shown by W. H. Row [RW]. However, as is clear from the proof of Theorem 3, this can only happen if the nonorientable M contains an embedded two-sided projective plane.

Lemma 3 can also be used to generalize the equivariant group action theorem given in [MY2] to the case where the prime factors are allowed to be homotopy three spheres. An interesting application of this generalized equivariant group action is the following:

THEOREM 4. *Let N be a compact orientable nonprime three-dimensional manifold whose prime decomposition contains exactly one copy of some prime nontrivial homotopy sphere $\hat{\mathbf{S}}^3$. Then every orientation preserving diffeomorphism of M has infinite order.*

Proof. The equivariant group action theorem in [MY2] as generalized in the above remark implies that a periodic diffeomorphism f: $N \to N$ induces a periodic diffeomorphism of $\hat{\mathbf{S}}^3$ which has fixed points. The proof of the generalized Smith conjecture in [SC] states that $\hat{\mathbf{S}}^3$ does not have an orientation preserving periodic diffeomorphism with fixed points. This contradiction proves the theorem.

8. Topology of complete three-dimensional manifolds with positive Ricci curvature and a geometric characterization of a three dimensional handlebody

The existence of topological obstructions to the existence of certain metrics with positive curvature, positive Ricci curvature or positive scalar curvature have been found. In dimension three, one expects to be able to give more complete information. In fact, recently R. Schoen and S. T. Yau [SY2] have proved that a complete noncompact three dimensional manifold N whose Ricci curvature is positive is diffeomorphic to $\mathbf{R}^3$. Their result depends in part on the results of this paper and in particular they need the following theorem.

THEOREM 5. *Let N be a compact orientable three-dimensional Riemannian manifold with non-negative Ricci curvature whose boundary, possibly empty, has non-negative mean curvature with respect to the outward normal. Then either*

(1) *N is covered by an irreducible homotopy sphere, or*

(2) *N is diffeomorphic to a three dimensional handlebody, or*

(3) *N is covered by a Riemannian product $\mathbf{S}^2 \times \mathbf{S}^1$ or $\mathbf{S}^2 \times [0, \mathbf{C}]$ with a metric of non-negative Gaussian curvature on $\mathbf{S}^2$, or*

(4) *N is flat and is covered by T^3 or $\mathbf{S}^1 \times \mathbf{S}^1 \times [0, \mathbf{C}]$.*

Before we prove Theorem 5, we shall need the following geometric characterization of a three-dimensional handlebody.

PROPOSITION 1. *Let N be a compact three-dimensional Riemannian manifold with non-empty boundary. N is a handlebody if and only if for every compact surface Σ in the interior of N and for every positive number ε there exists a surface Σ' isotopic to Σ such that $|\Sigma'| < \varepsilon$. Actually Σ will be a handlebody if and only if the isotopy class of a surface parallel to a boundary component contains surfaces of arbitrarily small area.*

Proof of Proposition 1. First note that if N is a handlebody, then the area of any compact surface can be shrunk down in a standard way to a one-dimensional complex in such a way that the area of the surface approaches zero with further and further shrinkings. Hence we shall only be concerned with the proof of the converse of the above statement.

More generally, if M is a fixed compact Riemannian manifold, possibly with boundary and $f: \Sigma \to M$ is a mapping of a compact orientable surface in M such that the homotopy class of f contains surfaces of arbitrarily small area, then if Σ has genus g, there will exist g homologically independent pairwise disjoint curves on Σ whose images are very short for some surface f' homotopic to f. In particular, in the above case there exist g homologically independent pairwise disjoint curves on Σ whose images under f are homotopically trivial in M (since the curves are homotopic to very short curves).

We shall now apply the result referred to in the previous paragraph to a particular surface ∂_1 in N. Let $\tilde{\partial}_1$ be a surface parallel to some component ∂_1 of ∂N. In other words, $\tilde{\partial}_1$ is just ∂_1 slightly pushed into the interior of N. If the area of $\tilde{\partial}_1$ gets arbitrarily small in its isotopy class, then there exists a collection $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$ of pairwise disjoint homologically independent Jordan curves on $\tilde{\partial}_1$ which are homotopically trivial in N. Since $\tilde{\partial}_1$ is isotopic to a surface $\tilde{\partial}_1^*$ of small area, the proof of Lemma 1 shows that $\tilde{\partial}_1^*$ is contained in a handlebody. As two surfaces that are isotopic in N differ by a diffeomorphism it also follows that $\tilde{\partial}_1$ is contained in a handlebody H in N.

By Dehn's lemma [HJ] the curves in Γ bound a pair-wise disjoint collection of disks $\mathfrak{D} = \{D_1, \ldots, D_n\}$ and clearly the interiors of these disks are, after a possible perturbation, contained in the component C of $N \sim \tilde{\partial}_1$ which is disjoint from ∂_1 and this C is contained in the handlebody H.

ASSERTION. *C is a handlebody.*

Proof of the assertion. First note that elementary surface topology shows that $\tilde{\partial}_1 - (\cup \Gamma)$ is a connected planar domain. Thus the surface $\Sigma \subset C$ obtained by surgery along the disks in $\mathfrak{D}$ is a two sphere. Since $\Sigma \subset C \subset H$, Alexander's theorem implies Σ bounds a ball B. From the construction of Σ it follows that C is a ball with one-handles attached. Thus C is by definition a handlebody which proves the assertion.

The proposition follows from the above assertion since the region between $\partial C = \tilde{\partial}_1$ and ∂_1 is diffeomorphic to $\partial C \times [0, 1]$ and hence C is diffeomorphic to N.

Proof of Theorem 5. We first prove that if N contains an embedded oriented two sphere S of least area in its isotopy class, then (3) holds. Let S_t, for small t, be the spheres of distance t from S where we use positive t for surfaces on the right-hand side of S and negative t for the surfaces on the left-hand side of S. Then from the fact that the mean curvature of S is zero and the Ricci curvature of N is non-negative, it is standard to use the second variational formula to show that $\Delta t \leq 0$. It is also known that $\Delta t = 0$ in an open neighborhood of S if and only if $\partial/\partial t$ defines a parallel vector field in that neighborhood.

Integrating $\Delta t \leq 0$ in a region bounded by S and S_ε, we find that the area of S_ε is not greater than the area of S and they are equal if and only if $\Delta t = 0$ in that region. Since S has least area in its isotopy class, the area of S_ε is not less than the area of S and hence $\Delta t = 0$ in a neighborhood of S and $\partial/\partial t$ defines a parallel vector there. Therefore, a neighborhood of S is isometric to a Riemannian product $S \times (-\varepsilon, \varepsilon)$.

Now let $\tilde{N}$ be the universal cover of N and $\tilde{S}$ be a component of the inverse image of S in $\tilde{N}$. Then the geodesics orthogonal to $\tilde{S}$ cannot intersect $\tilde{S}$ more than once; otherwise we can find a closed Jordan curve which has nonzero intersection number with $\tilde{S}$, in contradiction to the simply connectivity of $\tilde{N}$. If $\tilde{N}$ has no boundary, then from the conclusion of the above paragraph, $\tilde{N}$ is isometric to $\tilde{S} \times R$. If $\tilde{N}$ has boundary, then $\tilde{N}$ must be isometric to a product of $\tilde{S}$ with an interval. In fact, if c is the first number > 0 so that the region between $\tilde{S}_c$ and $\tilde{S}$ is isometric to $\tilde{S} \times [0, c]$ and $\tilde{S}_c \cap \partial\tilde{N} \neq \varnothing$, then by the maximum principle, one can prove that a component of $\partial\tilde{N}$ must be equal to $\tilde{S}_c$. In this way, we can prove that $\tilde{N}$ must be isometric to a product of $\tilde{S}$ with an interval.

Now let us assume that (3) does not hold. Then the above arguments show that $\tilde{N}$ admits no embedded oriented two sphere of least area in its isotopy class. Hence as in the proof of Theorem 3, we conclude that every sphere in N bounds a ball and N is irreducible.

If $\partial N \neq \varnothing$, we can apply Proposition 1 to conclude that either N is a handlebody (and (2) holds) or N admits a compact embedded minimal surface Σ with the following properties. Either Σ is orientable and has least area in its isotopy class or Σ is non-orientable and the area of the nearby level surfaces of the exponential map applied to the unit normal line field of Σ have area not less than twice the area of Σ. In the latter case, we can take a double cover of N so that we can assume Σ is orientable. By the second variational formula (see (1.5)), we can prove that Σ is either a sphere or a flat torus. By considering the covering of N associated to the subgroup $\pi_1(\Sigma)$ of $\pi_1(N)$, we can apply our previous argument to prove that N is covered isometrically by the product of Σ with an interval. Hence either (3) or (4) holds.

If N has no boundary, then the splitting theorem of Cheeger-Gromoll [CG] shows that either the universal cover $\tilde{N}$ of N is compact or isometric to $\mathbf{R}^3$, or isometric to the product of a compact surface of genus zero with the straight line. If $\tilde{N}$ is compact, then it is a homotopy sphere. Now $\tilde{N}$ must be irreducible because otherwise it contains a sphere with least area in its isotopy class and the above arguments show that this is a contradiction. This finishes the proof of Theorem 5.

COROLLARY. *Suppose N is a compact three dimensional orientable Riemannian manifold with non-negative Ricci curvature whose boundary, possibly empty, has non-negative mean curvature with respect to the outward normal. If Σ is a compact embedded orientable minimal surface in N, then one of the following holds*:

(1) *Σ is a Heegard surface.*

(2) *N is flat and Σ is totally geodesic.*

(3) *N is isometric to $\mathbf{S}^2 \times \mathbf{S}^1$ or $\mathbf{S}^2 \times I$ with a product metric and Σ is one of the sphere factors.*

(4) *N is diffeomorphic to $\mathbf{P}^3$ minus a ball or is diffeomorphic to $\mathbf{P}^3 \# \mathbf{P}^3$ where $\mathbf{P}^3$ is a three dimensional projective space. In this case, Σ is a totally geodesic sphere in N such that each component of $N \sim \Sigma$ is isometric to the nontrivial interval bundle over $\mathbf{P}^2$ induced as a Z_2-quotient of $\mathbf{S}^2 \times [0,1]$ with a product metric by an isometry $(x, t) \to (-x, -t)$.*

Proof. Since Σ is orientable, $N \setminus \Sigma$ consists of one or two components. We deform Σ isotopically in each of these components to obtain surfaces with least area. If, for each of these components, the minimal surface of least area does not exist, then according to Theorem 1 and Proposition 1, each of these components is a handlebody and Σ is a Heegard surface. Hence we can assume that for at least one component of $N \sim \Sigma$, Σ can be deformed isotopically to a surface of least area. As in the proof of Theorem 5, we can take a covering $\tilde{N}$ of N so that $\tilde{N}$ is either isometric to the product of a surface of genus zero with an interval or isometric to the product of the flat torus with an interval. From the way that we constructed this isometry, we know that a lift of Σ in $\tilde{N}$ is disjoint from a spherical factor or a torus factor. By moving these factors along the interval and applying the maximum principle, we conclude that Σ must be one of these factors. Hence either (2), (3) or (4) holds.

Corollary. *Let N denote a compact three-dimensional manifold with positive Ricci curvature. If N admits an embedded minimal sphere, then N is diffeomorphic to the three-dimensional sphere.*

Proof. Suppose that there exists an embedded minimal sphere $\mathbf{S}^2$ in N. Then since the fundamental group of N is finite and $\mathbf{S}^2$ is two-sided in N, this sphere must disconnect N into two regions R_1 and R_2. It follows from Theorem 5 that R_1 and R_2 are both balls and hence that N is diffeomorphic to the three-sphere.

Remark. It is conjectured that if N is any compact simply connected three-dimensional manifold, then it admits an embedded minimal sphere.

9. Compact minimal surfaces in complete manifolds with non-negative Ricci curvature

In this section, we generalize some of the theorems in the previous section to non-compact manifolds.

Theorem 6. *Let Σ be a compact minimal (embedded) surface in a complete non-compact orientable three-dimensional manifold N with non-negative Ricci curvature. Then N is isometric to a product of Σ with a straight line.*

Proof. If N has two ends, then by the splitting theorem, N is isometric to a product of a surface with a straight line. By looking at the level surface which touches Σ, one sees that the image of Σ is equal to one of these level surfaces.

Hence we can assume that N has only one end. Let N_1 be the component of $N \setminus \Sigma$ which contains infinity. Then ∂N_1 has non-negative mean curvature with respect to the outward normal. By using an argument in Schoen-Yau [SY2], we can minimize the closed surfaces in N_1 which are homologous to ∂N_1. In this way we obtain a stable minimal surface (possibly non-compact) with finite area. However, by the theorem of Fisher-Colbrie and Schoen [FS], this stable minimal surface is totally geodesic and has non-negative curvature. Since a non-compact complete surface with non-negative curvature has infinite area, this stable surface must be compact. The arguments in the previous section show the validity of the theorem.

10. The topological uniqueness of certain minimal surfaces in compact three-dimensional manifolds with non-negative scalar curvature

Under certain geometric restrictions on the metric and on the topology of N, some global topological information can be found concerning the placement of an embedded minimal surface Σ in N. An outstanding example of such a global relationship was given by H. B. Lawson [LH] who proved that if N is a three-sphere with a metric of positive Ricci curvature, then two embedded compact minimal surfaces of the same genus in N are ambiently isotopic in N. The same type of uniqueness result continues to hold if the Ricci curvature is non-negative and N is diffeomorphic to $\mathbf{S}^3$ or $\mathbf{S}^2 \times \mathbf{S}^1$ as was shown by W. Meeks in [MW1].

We shall now generalize the above stated uniqueness theorems to certain manifolds whose scalar curvature is non-negative.

THEOREM 7. *Let $X = \mathbf{S}^3 \mathop{\#}_{i=1}^{n} \mathbf{S}^2 \times \mathbf{S}^1$ be equipped with a Riemannian metric of non-negative scalar curvature. Let Σ be a compact embedded orientable minimal surface in X. Let N_j, $j = 1$ and/or 2, be a component of $X \sim \Sigma$ and define $I(N_j)$ to be the rank of the image of the induced map $i_*: \pi_2(N_j) \to H_2(N_j, Z_2)$. If Σ' is another embedded orientable minimal surface of the same genus in X, then there exists a diffeomorphism of X taking Σ to Σ' if and only if the number of components of $X \sim \Sigma$ is the same and if the corresponding indices $I(N_j)$ are the same.*

Proof of Theorem 7. The result is well-known for Σ of genus 0; hence assume genus of $\Sigma > 0$. We shall first consider the case when Σ disconnects X into two components. Let N_1 be one of these components. By the geometric sphere theorem in [MY2] there exist pairwise disjoint minimal spheres $S_1, \ldots, S_{I(N_1)}$ which generate the image of the induced map $i_*: \pi_2(N_1) \to H_2(N_1, Z_2)$. Let Y_1 be the geodesic closure of $N_1 \sim \bigcup_{i=1}^{I(N_1)} S_i$. Similarly we define N_2, the spheres $S_1^*, \ldots, S_{I(N_2)}^*$ and then define Y_2 to be the geodesic closure of $N_2 \sim \bigcup_{i=1}^{I(N_2)} S_i^*$.

ASSERTION 1. *Y_1, and similarly Y_2, is a handlebody with a finite number of balls removed.*

Proof of Assertion 1. Since the boundary of Y_1 has zero mean curvature the generalized loop theorem in [MY2] with the more general boundary value properties given in [MY3] yields a pairwise disjoint collection of minimal disks $\mathfrak{D} = \{D_1, \ldots, D_k\}$ with $\partial D_i \subset \partial Y_1$ and such that the normal subgroup of $\pi_1(\partial Y_1)$, generated by the curves $\Gamma = \{\partial D_1, \ldots, \partial D_k\}$, is the kernel of $i_*: \pi_1(\partial Y_1) \to \pi_1(Y_1)$.

Let W be the geodesic closure of $Y_1 \sim \bigcup_i D_i$. By construction, the boundary surfaces of each component of W are incompressible in W. Since ∂W has non-negative mean curvature on its smooth parts and its interior angles are less than $180°$, each surface in ∂W can be minimized in its isotopy class by pinching off a finite number of spheres (see the discussion of the minimizing process in the previous sections). Thus if some component of ∂W has positive genus, then W admits a positive genus stable minimal surface which is impossible if X has positive scalar curvature. In general, an argument due to Bourguignon (see [SY1]) says that we can approximate the metric of X by metrics with positive scalar curvature. Thus we may assume that each component of ∂W is a sphere.

Since ∂W consists of spheres and each of these spheres disconnects Y_1, a straightforward topological analysis using Alexander's theorem shows that each component of ∂W is a two-sphere which bounds a ball with a finite number of open balls removed. Therefore, Y_1, which is obtained from W by attaching one-handles, is a handlebody with a finite number of open balls removed which then proves Assertion 1.

Now consider the manifold $K = X \sim (\bigcup_i S_i \cup \bigcup_j S_j^*)$. This manifold is diffeomorphic to a manifold Z which is the connected sum of $n - I(N_1) - I(N_2)$ copies of $\mathbf{S}^2 \times \mathbf{S}^1$ and then with $2(I(N_1) + I(N_2))$ closed balls removed. After filling in the missing balls to get Z^*, Assertion 1 shows that Σ is a Heegard surface in Z^*. Waldhausen's uniqueness theorem [W] for such Heegard splittings implies that Σ is unique up to isotopy in Z modulo the condition stated in the theorem concerning the number of ends of Z in each component of $Z \sim \Sigma$. This fact is evidently equivalent to the uniqueness statement in the theorem in the case that Σ disconnects X.

Suppose now that Σ does not disconnect X and let $N = N_1$ and $S_1, \ldots, S_{I(N_1)}$ be as before and let $I = I(N_1)$.

ASSERTION 2. *Consider the spheres $S_1, \ldots, S_I$ in X under the natural projection $\pi\colon N_1 \to X$. If T denotes the geodesic closure of $X \sim \bigcup_i S_i$, then Σ disconnects T.*

Proof of Assertion 2. Let Σ be considered under the natural inclusion as one of the two components of ∂N. If the assertion were to fail, then by intersection theory with Z_2-coefficients there would exist an embedded arc $\gamma\colon [0,1] \to Y_1$ such that $\gamma(0) \in \Sigma$ and $\gamma(1)$ lies on the other component of ∂Y_1 and γ is disjoint from the spheres $S_1, \ldots, S_I$. Let W and $\mathfrak{D} = \{D_1, \ldots, D_R\}$ be as in the proof of Assertion 1. There, each component of ∂W must be a two-sphere.

The sum of the boundary spheres $S_1^*, S_2^*, \ldots, S_n^*$ of W arising from surgery on Σ are under the natural projection $\pi\colon W \to Y_1$ homologous to Σ in Y_1. In fact, after simplicial subdivision, Σ is equal to the Z_2-cycle $\sum_{i=1}^N \pi(S_i^*)$. Therefore Σ is homologous as a Z_2-cycle to a sum of some number of the boundary spheres of T which generate the image of $I_*\colon \pi_2(T) \to H_\alpha(T, Z_2)$. However this last fact is impossible since $[\Sigma] \wedge [\gamma] \neq 0$ where

$$\wedge\colon H_2(T, Z_2) \times H_1(T, \partial T; Z_2) \to Z_2$$

is the intersection pairing. This contradiction proves Assertion 2.

Now consider $\Sigma \subset T$. Since T is diffeomorphic to the connected sum of $n - I$ copies of $\mathbf{S}^2 \times \mathbf{S}^1$ with a total of $2I$ open balls removed and Σ disconnects

T into two pieces, Assertion 1 shows that these pieces are handlebodies each with I open balls removed. Thus, by the previously considered case where Σ separates X, Σ is uniquely defined up to isotopy in T. Therefore, the number I determines Σ up to a diffeomorphism of X. The proof of this last case completes the proof of the theorem.

COROLLARY 1. *If X is the three-sphere with a metric of non-negative scalar curvature, then two embedded compact minimal surfaces of the same genus are isotopic.*

Proof. Let Σ be a compact embedded minimal surface in X. Since $H_2(X, Z_2)$ is zero, Σ separates X into two regions N_1 and N_2 and hence Σ is orientable. The usual application of the sphere theorem and Alexander's theorem shows that $\pi_2(N, Z_2)$ is zero. The corollary now follows from the previous theorem.

COROLLARY 2. *If X is diffeomorphic to $\mathbf{S}^2 \times \mathbf{S}^1$ and has a metric of non-negative scalar curvature, then there are exactly three distinct isotopy classes possible for an embedded compact orientable minimal surface Σ in X of a fixed genus. The three distinct classes are as follows:*

(1) *Σ does not separate X and Σ is isotopic to $\mathbf{S}^2 \times \{1\}$ with trivial handles attached.*

(2) *Σ is a Heegard surface in X.*

(3) *Σ separates X into two regions, one of which is a handlebody and the second a handlebody connected sum with $\mathbf{S}^2 \times \mathbf{S}^1$.*

Proof. The proof of the corollary consists of just tracing the proof and applying the statement of Theorem 7.

Remark. There exist metrics on $\mathbf{S}^2 \times \mathbf{S}^1$ with positive scalar curvature which have all the distinct possibilities stated in Corollary 2. For example, let γ: $[0, 1] \to \mathbf{R}^4 \sim B^4$ be a smooth embedded arc which intersects the unit ball B^4 orthogonally in exactly two close points $p = \alpha(0)$ and $q = \alpha(1)$. Let N^1 be the union of B^4 and a very small ε-neighborhood Δ of γ. Then after a small C^0-perturbation N of N near p and q, ∂N has a metric of positive scalar curvature.

Now let Σ be an example of an embedded minimal surface in $\mathbf{S}^3$ and rotate Σ to get Σ_1 so that p and q both lie in the same component of $\mathbf{S}^3 \sim \Sigma_1$ (this is possible since p and q are close), and such that the metric near Σ_1 in X is the same as the metric in $\mathbf{S}^3$. (This is possible since the metric on N^1 is perturbed only in a very small neighborhood of each of the points p and q.) Such a surface Σ_1 is of type (3) given in Corollary 2. With the same metric on ∂N we can create an example of type (1) on ∂N. For this construction, just rotate Σ in $\mathbf{S}^3$ to get Σ_2

with p and q in different components of $\mathbf{S}^3 \sim \Sigma_2$, and such that the metric near Σ_2 in ∂N^1 is the same as the metric in $\mathbf{S}^3$. Note that by taking small rotations of Σ_1 and Σ_2 in ∂N^1, these minimal surfaces are each part of two six-dimensional families of minimal surfaces. An example of case 2 occurs as the totally geodesic torus in the standard product metric on $\mathbf{S}^2 \times \mathbf{S}^1$.

11. Complements of minimal surfaces

In this section, we study non-compact properly embedded minimal surfaces, in a complete manifold with non-negative Ricci curvature.

LEMMA 4. *Let N be a complete Riemannian manifold with boundary. If the dimension of N is less than eight, the boundary of N has non-negative mean curvature with respect to the outward normal and the boundary ∂N has at least two components, then there exists a properly embedded stable codimension-one minimal submanifold in N.*

Proof. If ∂N has at least two components, ∂_1 and ∂_2, then the component ∂_1 represents a nontrivial class in $\tilde{H}_{n-1}(N, Z_2)$ where $\tilde{H}_i(N, Z_2)$ denotes the i-th homology group of N with chains being locally finite singular chains in N.

Consider a collection $A_1 \subset A_2 \subset \dots A_n \subset \dots$ of compact submanifolds (with boundary) of ∂_1 such that the union $\bigcup_i A_i = \partial_1$. As the boundary of N has non-negative mean curvature with respect to the outward normal there exists (see [MY3]) for each ∂A_i an area minimizing regular current C_i in N with $\partial C_i = \partial A_i$ and the Z_2-cycle $C_i + A_i$ is a Z_2-boundary of a locally finite Z_2-singular chain. Since we are using Z_2-coefficients, it is clear that the area of C_i cannot build up locally in N. Thus the usual convergence theorems show that a subsequence of the C_i converge to an embedded minimal submanifold Σ of N with the property that Σ represents the same Z_2-homology class as ∂_1. By construction Σ is stable in the sense that if Δ is a compact $(n-1)$-dimensional submanifold of Σ, then Δ has least area among currents in N which have boundary $\partial\Delta$ and are Z_2-homologous (mod $\partial\Delta$) to the singular chain represented by Δ.

THEOREM 8. *Let N be a complete three-dimensional orientable Riemannian manifold with non-empty boundary. If the boundary of N has non-negative mean curvature with respect to the outward normal and the scalar curvature of N is non-negative, then either:*

(1) *N contains an embedded stable minimal sphere or a proper embedded stable minimal surface which is a plane, a flat cylinder or a flat torus, or*

(2) *The boundary of N is connected.*

Furthermore, in case (1), *a cylinder or a flat torus can occur only if N is flat.*

COROLLARY 1. *If Σ_1 and Σ_2 are complete, proper and immersed minimal surfaces in the three dimensional euclidean space $\mathbf{R}^3$, then one of the following holds*:

(1) *Either Σ_1 or Σ_2 is a linear plane.*

(2) *There exists a linear plane P in $\mathbf{R}^3$ disjoint from Σ_1 and Σ_2 such that Σ_1 and Σ_2 lie in different components of $\mathbf{R}^3 \sim P$.*

(3) *Σ_1 and Σ_2 intersect nontrivially.*

Proof of Corollary 1. Suppose that neither case 1 nor case 3 occurs. Let N be the geodesic closure of the component of $\mathbf{R}^3 \sim (\Sigma_1 \cup \Sigma_2)$ which has as boundary part of Σ_1 and part of Σ_2. Then N satisfies the hypothesis of Theorem 8 except that the boundary of N is non-smooth at some self-intersection curves of Σ_1 or Σ_2. However, the interior angles of the non-smooth portion of ∂N are less than or equal to 180° and the argument given in [MY3] continues to hold for N with such boundary. Since ∂N is not connected we can apply Theorem 8. Since the only stable minimal surface in $\mathbf{R}^3$ is linear (see [FS] and [CP]), (2) must hold.

COROLLARY 2. *Let N be as in the theorem except that the Ricci curvature of N is positive. Then either N is a handlebody or else ∂N is non-compact and connected.*

Proof of Corollary 2. This corollary follows from Theorem 5, Theorem 8 and the stability theorem of Fischer-Colbrie and Schoen [FS].

Proof of Theorem 8. From Lemma 4, if the boundary of N is disconnected, we can deform one of the components of N to the sum of some properly embedded minimal surfaces. We can assume that they are homologous to that component of ∂N in $\tilde{H}_2(N, Z_2)$ and that they have least area in this class. These surfaces are orientable because they form the boundary of an orientable manifold. However, stable orientable surfaces in complete three-dimensional manifolds with non-negative scalar curvature were studied by Fischer-Colbrie and Schoen [FS] and we have proved (1) stated in the theorem.

UNIVERSITY OF CALIFORNIA, BERKELEY, CALIFORNIA
AUSTRALIAN NATIONAL UNIVERSITY, CANBERRA, AUSTRALIA
INSTITUTE FOR ADVANCED STUDY, PRINCETON, NEW JERSEY

REFERENCES

[AS] F. J. ALMGREM, JR. AND L. SIMON, Existence of embedded solutions of Plateau's problem, Ann. Scoula Norm. Sup. Pisa, **6** (1979) 447–495.

[AW] W. K. ALLARD, On the first variation of a varifold, Ann. of Math. **95** (1972), 417–491.

[CG] J. CHEEGER AND D. GROMOLL, The splitting theorem for manifolds of nonnegative Ricci curvature, J. Diff. Geom. **6** (1971), 119–128.

[CP] M. DO CARMO AND C. K. PENG, Stable minimal surfaces in $\mathbf{R}^3$ are planes, Bull. AMS (1979), 903–906.

[FS] D. FISCHER-COLBRIE AND R. SCHOEN, The structure of complete stable minimal surfaces in 3-manifolds, Comm. Pure Appl. Math. **33** (1980), 199–211.

[FT] T. FRANKEL, On the fundamental group of a compact minimal submanifold, Ann. of Math. **83** (1966), 68–73.

[HJ] J. HEMPEL, *Three Manifolds*, Annals of Math. Studies, #86, Princeton University Press, 1976.

[LH] H. B. LAWSON, The unknottedness of minimal embeddings, Invent. Math. **11** (1970), 183–187.

[MW1] W. H. MEEKS III, The conformal structure and geometry of triply periodic minimal surfaces in $\mathbf{R}^3$, Thesis (revised), University of California, Berkeley, 1977.

[MW2] ______, The topological uniqueness of minimal surfaces in Euclidean space, Topology **20** (1981), 389–410.

[MY1] W. H. MEEKS III AND S. T. YAU, The classical Plateau problem and the topology of three-dimensional manifolds, to appear in Topology.

[MY2] ______, Topology of three dimensional manifolds and the embedding problems in minimal surface theory, Ann. of Math. **112** (1980), 441–485.

[MY3] ______, The existence of embedded minimal surfaces and the problem of uniqueness, to appear in Math. Zeit.

[MS] J. MICHAEL AND L. SIMON, Sobolev and mean-value inequalities on generalized submanifolds of $\mathbf{R}^n$, Comm. Pure Appl. Math. **26** (1973), 361–379.

[PJ] J. PITTS, Existence and regularity of minimal surfaces on Riemannian manifolds, preprint.

[RW] W. H. ROW, Irreducible 3-manifolds whose orientable covers are not prime, Proc. AMS **34** (1972), 541–545.

[SC] Conference on the positive solution of the Smith conjecture, Columbia University, New York, 1979, preprint (Eds. John Morgan and Hyman Bass).

[SS] R. SCHOEN AND L. SIMON, Regularity of stable minimal hypersurfaces, preprint, Inst. for Advanced Study, Princeton, NJ, 1980.

[SY1] R. SCHOEN AND S. T. YAU, Existence of incompressible minimal surfaces and the topology of three dimensional manifolds with non-negative scalar curvature, Ann. of Math. **110** (1979), 127–142.

[SY2] ______, The topology of complete three dimensional manifolds with positive Ricci curvature and scalar curvature, to appear in *Seminar in Differential Geometry*, Ann. of Math. Studies, Princeton Univ. Press, 1982.

[SJ] J. R. STALLINGS, *Group Theory and Three-Dimensional Manifolds*, Yale University Press, New Haven, 1971.

[W] F. WALDHAUSEN, Heegard-Zerlegungen der 3-Sphäre, Topology **7** (1968), 195–203.

(Received September 24, 1981)