J. DIFFERENTIAL GEOMETRY
24 (1986) 343–372

ISOTOPY OF 4-MANIFOLDS

FRANK QUINN

The principal result of this paper is that the group of homeomorphisms mod isotopy (the "homeotopy" group) of a closed simply-connected 4-manifold is equal to the automorphism group of the quadratic form on H_2.

There is an analogy between simply-connected closed 4-manifolds and connected surfaces, in that they are both classified by simple algebraic-topological data. Surfaces are classified up to homeomorphism by the isomorphism class of the fundamental group. The 4-manifolds are classified up to homeomorphism by the isomorphism class of the intersection form and the Kirby-Siebenmann invariant in $\mathbf{Z}/2$ [3]. The analogy now extends to automorphisms, in that in both cases homeomorphisms are classified by the induced automorphism of the algebraic structure.

Other results include a "uniqueness" for handlebody structures on simply-connected 5-manifolds, the determination of $\pi_4(\mathrm{TOP}(4)/\mathrm{O}(4))$, and a pseudo-isotopy theorem for simply connected 4-manifolds with boundary.

1. Statements of results

Suppose M is a closed oriented (topological) manifold of dimension 4. Then intersections define a symmetric nonsingular bilinear form, denoted by λ, on H_2M. A homeomorphism of manifolds induces an isometry of H_2, and isotopic homeomorphisms induce the same isometry. Therefore there is a natural homomorphism from $\pi_0\,\mathrm{TOP}(M)$ (= homeomorphisms mod isotopy) to $\mathrm{Aut}(H_2M, \lambda)$.

1.1 Theorem. *Suppose M is a closed 1-connected 4-manifold. Then the natural homomorphism $\pi_0\,\mathrm{TOP}(M) \to \mathrm{Aut}(H_2M, \lambda)$ is an isomorphism.*

Freedman [3] has shown this to be onto. For injectivity we show that homeomorphisms which are equal on homology are isotopic, or equivalently that a homeomorphism inducing the identity on homology is isotopic to the

Received October 9, 1985 and, in revised form, June 2, 1986. The author was partially supported by the National Science Foundation.

identity. We note the curious fact that there are homotopy equivalences which are the identity on homology, but are not homotopic to the identity.

To prove injectivity we first show by a homotopy calculation that homotopy equivalences inducing the identity on homology are detected by normal invariants. This means a homeomorphism must be homotopic to the identity. Second a standard 5-dimensional surgery argument shows that homotopy to the identity implies pseudoisotopy to the identity. Finally pseudoisotopy is shown to imply isotopy. The technically difficult step is the pseudoisotopy theorem, which in this setting (1-connected closed) is due to B. Perron [12]. The easier part (identity on homology implies pseudoisotopy) has also been shown by M. Kreck [8] for the special case of stability smoothable homeomorphisms.

The next result concerns the classifying spaces $\mathrm{TOP}(n)/\mathrm{PL}(n)$. These classify PL structures on a given n-dimensional topological microbundle, and for manifolds which are open or have dimension $\geqslant 5$, maps to this space classify sliced concordance classes of PL structures. There is a stabilization map $\mathrm{TOP}(n)/\mathrm{PL}(n) \to \mathrm{TOP}/\mathrm{PL}$. One of the main results (the high-dimensional "annulus conjecture") of [7] is that the only nonvanishing homotopy group of TOP/PL is $\mathbf{Z}/2$ in dimension 3, and that the stabilization induces an isomorphism of π_j in the "geometrically significant" range $j \leqslant n$, provided $n \geqslant 5$. When $n \leqslant 3$ the spaces are contractible (see [7, p. 253]). When $n = 4$ the stabilization was shown to be an isomorphism for $j = 0, 1, 2$ in [13, part III], and $j = 2, 3$ in [10]. The final "geometrically significant" case is settled here, and as in the other cases it is the same as the stable group:

1.2 Theorem. $\pi_4 \mathrm{TOP}(4)/\mathrm{PL}(4) = 0$.

Lashof and Shaneson [9] have defined "S-smoothings" and "S-isotopies" to be smoothings and isotopies of connected sums $M \# k(S^2 \times S^2)$ for some k, satisfying a regularity condition on the $k(S^2 \times S^2)$. They showed that S-smoothings are classified up to S-isotopy by lifts of the tangent microbundle of M from $B_{\mathrm{TOP}(4)}$ to $B_{\mathrm{O}(4)}$. The calculation of the homotopy of the fiber makes this result much more explicit. The calculation of the lower homotopy identified the obstruction to the existence of an S-smoothing to be the Kirby-Siebenmann obstruction in $H^4(M;\ \mathbf{Z}/2)$. Theorem 1.2 implies (and indeed is equivalent to) the only obstruction to S-isotopy of two structures is the relative obstruction in $H^3(M;\ \mathbf{Z}/2)$.

Next we consider the uniqueness of handlebody structures. A handlebody structure on a manifold pair $(M, \partial_0 M)$ is an identification of M with a collar on $\partial_0 M$ with locally finite collections of handles attached. In a 1-parameter family of structures the following changes are allowed: change of attaching embeddings by ambient isotopy, introduction or deletion of a cancelling pair of handles, change of order of attachment, and change of the identification with M by isotopy. We also allow insertion or deletion of collars between layers of handles. (This is essentially the same as changing attaching maps by isotopy.)

It is a fundamental fact that not only do handlebody structures almost always exist, but usually any two can be joined by a 1-parameter family (i.e. the structure is unique up to such deformations). For topological manifolds the existence result is that $(M, \partial_0 M)$ has a handlebody structure if $\dim M \neq 4$, and when $\dim M = 4$ if and only if M has a smooth structure. The final case of this to be settled was $\dim M = 5$ in [13, part III]. The uniqueness known to date is that if $\dim M \neq 4, 5$ any two handlebody structures can be joined by a 1-parameter family, and when $\dim M = 4$ this is the case if and only if the corresponding smooth structures on M are sliced concordant. When $\dim M > 5$ this follows from [7], and in dimensions $\leqslant 4$ it is a consequence of the hauptvermutung for 3-manifolds. The 5-dimensional case is still unsettled, but we make a start on it here:

1.3 Theorem. *Any two (topological) handlebody structures on a compact 1-connected 5-manifold pair can be joined by a 1-parameter family of structures.*

The principal technical ingredient in these results is a study of 4-dimensional pseudoisotopy. The main conclusion is:

1.4 Theorem. *Suppose M is a compact simply-connected 4-manifold. Then a pseudoisotopy of M which is the identity on the boundary is topologically isotopic rel boundary to an isotopy. A smooth pseudoisotopy is smoothly isotopic to the identity after connected sum with some number of copies of the identity on $S^2 \times S^2 \times I$.*

The topological case is an extension of a result of Perron [12], which applies if M is closed or has a handlebody structure with no 1-handles. For smooth compact higher dimensions manifolds the simply-connected case is due to Cerf [2], and the general compact case to Hatcher and Wagoner [6]. The modifications necessary for the high-dimensional topological case are given by Pedersen in the appendix to [1]. A high-dimensional noncompact version is given in [13, part IV].

The final result we state here is a weak version of the assertion "1-connected 4-manifolds have handlebody structures with no 1-handles." The assertion itself is false. There are theorems asserting that j-connected manifolds need not have $\leqslant j$ handles, but they do not apply in codimension 3. Worse, there are 1-connected 4-manifolds with no handlebody structure at all. However we can weaken the conclusion a little to obtain a true statement. Note if a manifold has no 1-handles then it is obtained from a collar on its boundary by adding handles of index $\leqslant 2$, and then a 4-cell attached on S^3 in the boundary.

1.5 Proposition. *Suppose M is a compact 1-connected 4-manifold. Then $M = N \cup D$, where N is a handlebody on ∂M with handles of index $\leqslant 2$ and D is a flat 4-cell.*

This differs from "no 1-handles" in that the intersection $N \cap D$ is not specified to be S^3. When M is closed this is due to P. Vogel, and later Siebenmann (see [12]).

The paper is organized as follows: the main theorems are reduced to pseudoisotopy results, Theorem 1.1 in §2, and Theorems 1.2, 1.3 in §3. §4 contains most of the technical material on pseudoisotopy. We follow the traditional plan of considering a pseudoisotopy as two collar structures on $M \times I$, suppose they are joined by a 1-parameter family of handle structures, and eliminate the handles from the family. This gives the stable smooth version. It also shows that the topological version holds if the existence of 1-parameter families, as in 1.3, is assumed. Unfortunately the existence of 1-parameter families is deduced from the pseudoisotopy theorem, in §3. To avoid a logical circularity, the material of §4 is applied in §5 to infinite handlebodies, to obtain a codimension 2 result. Theorem 1.4 is deduced from this and Proposition 1.5. The codimension 2 result is similar to the main technical result of Perron [12], and it may be that the topological part of 1.4 can be deduced from Perron's result and 1.5. Finally 1.5 is proved in §6.

I would like to thank M. Freedman and the referee for their comments on an earlier version.

2. Automorphisms

In this section we prove Theorem 1.1 by showing that homologous homeomorphisms are isotopic. They are first shown to be homotopic (Proposition 2.1), then pseudoisotopic (Proposition 2.2). Theorem 1.4 then implies they are isotopic. Throughout we assume M is a simply-connected closed topological 4-manifold.

Suppose $f\colon M \to M$ is a homotopy equivalence. Then there is a normal invariant $n(f)$ defined in $[M; G/\mathrm{TOP}] \approx H^2(M; \mathbf{Z}/2) \oplus H^4(M; \mathbf{Z})$.

2.1 Proposition. *n defines a bijection, from homotopy classes of maps inducing the identity on homology, to the subgroup of $H^2(M; \mathbf{Z}/2)$ orthogonal to ω_2.*

Proof. Wall [16, p. 237] shows n is onto by identifying the subgroup of H^2 with the torsion part of $[M; G/\mathrm{PL}] \subset [M; G/\mathrm{TOP}]$, and showing this can be realized. His construction will be described in more detail later. The image of n

can be no larger than $[M; G/\mathrm{PL}]$ because the remaining $\mathbf{Z}/2$ is the difference of the Kirby-Siebenmann invariants of the domain and range, hence zero. It remains to be seen that n is injective.

Let c: $M \to M \vee S^4$ be the map which pinches off a top cell, and let τ: $S^4 \to M$ be a map. Define $\langle \tau \rangle$ to be the composition $(1, \tau)c$: $M \to M \vee S^4 \to M$. This is a map of M which is the identity on homology, and the construction $\langle - \rangle$ defines a homomorphism from $\pi_4(M)$ to such maps. The first step is to show $\langle - \rangle$ is onto.

M is homotopy equivalent to a CW complex of the form $(\vee^m S^2) \cup_\alpha D^4$, where m is the rank of $H_2(M)$ and α: $S^3 \to \vee^m S^2$ is the attaching map of the top cell [11]. Suppose f: $M \to M$ is the identity on homology. We may suppose it is the identity on the 2-skeleton, since a map $\vee^m S^2 \to \vee^m S^2$ is determined by homology. Since the identity preserves the attaching map, we may assume f is the identity near ∂D^4. Think of S^4 as $D^4_+ \cup D^4_-$ with the boundaries identified, and define τ: $S^4 \to M$ to be f on D^4_+ and the identity on D^4_-. Then $f = \langle \tau \rangle$, so the construction is onto.

The next step is to determine which maps $\langle \tau \rangle$ are homotopic to the identity, and for this we need an explicit description of $\pi_4(M)$. The homomorphism $\pi_4(\vee^m S^2) \to \pi_4(M)$ is onto if $M \neq S^4$. To see this first note any τ: $S^4 \to M$ must be trivial on H_4, since otherwise in homology with coefficients prime to the degree it would be degree 1, hence onto H_2. But a map trivial on H_4 deforms off the 4-cell into the 2-skeleton.

Denote the inclusion of the ith summand by x_i: $S^2 \to \vee^m S^2$, then according to the Hilton-Milnor theorem [17, p. 533], $\pi_4(\vee^m S^2)$ is generated by

(1) Whitehead products $[[x_i, x_j], x_k]$, with $i > j \leqslant k$,

(2) compositions $[x_i, x_j]s\eta$ with $i > j$, where $s\eta$ is the suspension of the Hopf map, and

(3) compositions $x_j \eta s \eta$, which we denote by η_j^2.

In $\pi_4(\vee^m S^2)$ these elements are independent, and those in (1) are of infinite order, those in (2) and (3) have order 2 since $s\eta$ has order 2. In $\pi_4(M)$ there are additional relations from the attaching map of the top cell.

Identify the subgroup generated by the η_j^2 with $H_2(M; \mathbf{Z}/2)$, generated by x_j. Then the argument of Wall [16, p. 237] shows that $n\langle - \rangle$ (normal invariant of the associated map) is an isomorphism when restricted to the kernel of ω_2: $H_2(M; \mathbf{Z}/2) \to \mathbf{Z}/2$. So to show the normal invariant is injective on homotopy classes it is sufficient to show that maps corresponding to elements of type (1) or (2) are homotopic to the identity, and that a sum of η^2 elements not in the kernel of ω_2 (presuming there is one) gives a map homotopic to the identity.

As above represent M as $\bigvee^m S^2 \cup_\alpha D^4$, and define a homotopy H_k: $M \times I \to M \vee S^3$ by: retract $M \times I$ to $(\bigvee^m S^2) \times I \cup_\alpha D^4 \times \{0\}$. Then map this to M by projection on all but the kth S^2 component of the wedge. Map the kth $S^2 \times I$ by the projection plus the degree 1 map to the 3-sphere. The map $H_k(1)$ is the identity on the 2-skeleton of M. As above this means it is obtained from the inclusion $M \to M \vee S^3$ by changing the top cell by some element $\beta_k \in \pi_4(M \vee S^3)$. Denote the inclusion by y: $S^3 \to M \vee S^3$; then

Claim. $\beta_k = \Sigma_i \lambda(x_i, x_k)[y, x_i]$.

Here λ denotes the pairing obtained by cup product of the corresponding cohomology classes, evaluated on the fundamental class: $\lambda(x_i, x_j) = (x_i^* \cup x_j^*)[M]$. This is the inverse of the intersection form used in Theorem 1.1. I would like to thank T. Cochran and N. Habegger for pointing out this discrepancy.

We first show that the proposition follows from the claim. Since λ is nonsingular, the matrix $(\lambda(x_i, x_k))$ has an inverse, $(r_{k,j})$. Define homotopies $h_j = \Sigma_k r_{k,j} H_k$; then the map $h_j(1)$ is obtained by changing the top cell by $\Sigma_k r_{k,j}\beta_k$. Substituting for β_k from the claim we get $h_j(1) = \langle [y, x_j] \rangle$. Now suppose $\tau \in \pi_3(M)$. The composition $(1, \tau)h_j$ is a homotopy from the identity of M to the map $\langle [\tau, x_j] \rangle$. First let $\tau = [x_i, x_k]$; then this gives homotopies of the generators of type (1) ($\langle [[x_i, x_k], x_j] \rangle$) to the identity. Second let $\tau = x_i\eta$; then we get a homotopy of $\langle [x_i\eta, x_j] \rangle$ to the identity. Since $[x_i\eta, x_j] = [x_i, x_j]s\eta + [[x_i, x_j], x_i]$, and $\langle [[x_i, x_j], x_i] \rangle$ is homotopic to the identity, this gives homotopies from type (2) generators to the identity.

Finally recall that the attaching map of the top cell is given by $\alpha = \Sigma_{i<j} \lambda(x_i, x_j)[x_i, x_j] + \Sigma_i \lambda(x_i, x_i)x_i\eta$ [11]. Composing with $s\eta$ gives a homotopy from $\Sigma_{i<j}\lambda(x_i, x_j)[x_i, x_j]s\eta$ to $\Sigma_i \omega_2(x_i)\eta_i^2$. Here we have used the fact that η_i^2 is of order 2 to substitute $\omega_2(x_i)$ for $\lambda(x_i, x_i)$. The first expression involves generators of type (2), so gives a map homotopic to the identity. Therefore $\langle \Sigma_i \omega_2(x_i)\eta_i^2 \rangle$ is homotopic to the identity. If $\omega_2 \neq 0$, this is a nonzero element not in the kernel of ω_2. As noted above this is sufficient to show that the normal invariant is injective.

Proof of the claim. We describe β_k explicitly as a homotopy $S^3 \times I \to (\bigvee^m S^2 \cup_\alpha D^4) \vee S^3$ which takes $S^3 \times \{0,1\}$ to the basepoint (the center of the 4-cell). On $S^3 \times [0, 1/3]$ go radially from the center to ∂D^4. On $S^3 \times [1/3, 2/3]$ it is the composition of the attaching map α with the homotopy $\bigvee^m S^2 \times I \to \bigvee^m S^2 \vee S^3$ which is constant on all but the kth component, and there is the identity plus the degree 1 map on S^3. Finally on $S^3 \times [2/3, 1]$ go radially back to the center.

The attaching map is given by $\alpha = \Sigma_{i<j}\lambda(x_i, x_j)[x_i, x_j] + \Sigma_i \lambda(x_i, x_i)x_i\eta$. This means there are a lot of little cells in S^3 whose complement is mapped to

the basepoint, and whose interiors are mapped to some $\pm[x_i, x_j]$ or $\pm x_i\eta$. On the complement the homotopy β_k goes radially out to the boundary, sits there at the basepoint a while, and then comes back. Such a map is canonically nullhomotopic. Therefore β_k as a homotopy class decomposes as a sum of pieces, one for each little cell. Denote by A_i the contribution from a cell where the attaching map is $[x_i, x_k]$ if $i \neq k$, and is $x_k\eta$ if $i = k$. Note cells where the attaching map is $\pm[x_i, x_j]$ or $\pm x_i\eta$, give nullhomotopic elements of π_4 unless one index is equal to k. Therefore $\beta_k = \Sigma_i \lambda(x_i, x_k)A_i$. The final step is to evaluate A_i.

The homotopy A_i associated to a little cell in ∂D^4 takes place in the image of the cone on the cell. By compressing into the boundary we get a homotopy in the 2-skeleton, which can be described as follows: Choose $S^3 \times I \to D^3$ so it takes $S^3 \times \{0,1\} \cup \{\text{pt}\} \times I$ to a point in ∂D^3, and for $t \in [1/3, 2/3]$ maps each hemisphere of S^3 by the identity to D^3. Compose this with a_i: $D^3 \to \bigvee^m S^2$, where $a_i = [x_i, x_k]$ if $i \neq k$, and $x_k\eta$ if $i = k$. Finally on one hemisphere modify this near $1/2$ by pushing the kth S^2 over the S^3.

Note this description takes place in $S^2 \vee S^2 \vee S^3$ if $i \neq k$, and $S^2 \vee S^3$ if $i = k$. As an aid to visualization note that the attaching maps are the ones which occur in the manifolds $S^2 \times S^2$, and $\mathbf{C}P^2$ respectively. By drawing a picture of the homotopy described above it can be seen directly that $A_i = [x_i, y]$. Substituting this in the formula for β_k yields the claim, and completes the proof of 2.1.

2.2 Proposition. *Suppose M is a 1-connected 4-manifold with connected boundary. Then homeomorphisms of M which agree on ∂M and are homotopic rel ∂M are pseudoisotopic.*

Proof. The homotopy given is a map F: $M \times I \to M \times I$, which is a homeomorphism on $M \times \{0,1\} \cup \partial M \times I$. This defines an element of the structure set $S_{\text{TOP}}(M \times I, \partial)$. A pseudoisotopy is such a homotopy which is itself a homeomorphism, so to prove the proposition we show F is homotopic rel ∂M to a homeomorphism. For this it is sufficient to show the structure set consists of a single element.

Since M is 1-connected the normal invariant defines a bijection $S_{\text{TOP}}(M \times I, \partial) \to [(M \times I)_0, \partial;\ G/\text{TOP}, \text{pt}]$. Here $(-)_0$ denotes $(-)$ with an interior point removed. $[(M \times I)_0, \partial;\ G/\text{TOP}, \text{pt}] \approx [M_0, \partial;\ \Omega(G/\text{TOP}), \text{pt}]$, and there is an exact sequence

$$\to H^3(M, \partial; \mathbf{Z}) \to \left[M_0, \partial;\ \Omega(G/\text{TOP}), \text{pt}\right] \to H^1(M, \partial; \mathbf{Z}/2) \to .$$

The end groups are trivial since M is 1-connected and ∂M is connected. Thus the middle group, and the structure set, is trivial as claimed.

Proposition 2.1 and 2.2 imply that a self-homeomorphism which induces the identity on homology is pseudoisotopic to the identity. Theorem 1.4 then implies that it is isotopic to the identity, as required to complete 1.1. We note that 2.2 and 1.4 both apply if M has boundary, which is held fixed. Therefore extensions of 1.1 to the bounded case will follow from working out what happens in the homotopy theory in 2.1. For example understanding automorphisms of manifolds with boundary an S^1 bundle over S^2 would be useful in determining the uniqueness of embeddings $S^2 \subset M^4$ with 1-connected complement.

3. Bundles and handle uniqueness

Theorems 1.2, on the homotopy of TOP(4)/O(4), and 1.3, on handlebody structures on 5-manifolds, are deduced from the pseudoisotopy results in this section.

Proof of Theorem 1.2. We show that any map $f\colon S^4 \to \mathrm{TOP}(4)/\mathrm{O}(4)$ is nullhomotopic by studying the stable uniqueness of smooth structures on 4-manifolds.

A smooth structure on M^4 determines a classifying map $M \to B_{\mathrm{O}(4)}$. A second smooth structure on the underlying topological manifold determines a second map with the same image in $B_{\mathrm{TOP}(4)}$. The difference is classified by a map $M \to \mathrm{TOP}(4)/\mathrm{O}(4)$. If the smooth structures are sliced concordant, then the map to TOP(4)/O(4) is nullhomotopic. (A sliced concordance is a smooth structure on $M \times I$ with the given ones at the ends, and such that the projection to I is a smooth submersion; see [7].)

In higher dimensions any map $M^n \to \mathrm{TOP}(n)/\mathrm{O}(n)$ (M smooth) can be realized as a difference of tangent bundles of smooth structures [7], so $\pi_n(\mathrm{TOP}(n)/\mathrm{O}(n))$ can be determined by studying the uniqueness of smooth structures on S^n. In dimension 4 we follow the same plan, but have to stabilize by connected sum with $S^2 \times S^2$.

If a map $f\colon S^4 \to \mathrm{TOP}(4)/\mathrm{O}(4)$ is stabilized by connected sum with the point map on copies of $S^2 \times S^2$, then for some k it can be realized as a smooth structure on $\#^k(S^2 \times S^2)$ [10]. Since $\pi_4(\mathrm{TOP}(5)/\mathrm{O}(5)) = 0$, when crossed with $\mathbf{R}$ this structure is isotopic to the standard one. This means there is a smooth structure on $(\#^k(S^2 \times S^2)) \times I$ which is standard on one end, and the structure corresponding to f on the other. This structure will define a sliced concordance if we can arrange the projection to I to be a submersion.

The structure is a smooth h-cobordism, which according to [14] becomes a smooth product after stabilization by connected sum along arcs with copies of $(S^2 \times S^2) \times I$. The product structure defines a pseudoisotopy. The stable smooth part of 1.4 implies that after further stabilizations there is an isotopy of the pseudoisotopy to an isotopy. Since an isotopy followed by the projection to I is a submersion, this defines a sliced concordance of structures on $\#^m(S^2 \times S^2)$, some m. As explained above this means there is a nullhomotopy of $f\#mp$, where p: $S^2 \times S^2 \to \mathrm{TOP}(4)/\mathrm{O}(4)$ is the point map.

Adding point maps gives an extension of f over the mapping cylinder of the degree 1 map $\#^m(S^2 \times S^2) \to S^4$. Let X denote the union of this mapping cylinder with the cone on $\#^m(S^2 \times S^2)$. The argument above implies that the original map on S^4 extends over X, i.e. there is a factorization of f as $S^4 \to X \to \mathrm{TOP}(4)/\mathrm{O}(4)$. But the inclusion $S^4 \to X$ is nullhomotopic, so f is nullhomotopic, and $\pi_4(\mathrm{TOP}(4)/\mathrm{O}(4)) = 0$.

To see this last point, the nullhomotopy of $S^4 \to X$, write $\#^m(S^2 \times S^2)$ as $(\vee^{2m} S^2) \cup_\alpha D^4$, where α is a sum of Whitehead products of components. X is almost the suspension of this; $X = [S(\vee^{2m} S^2) \vee S^4] \cup_{s\alpha+1} D^5$. But suspensions of Whitehead products are trivial, so $s\alpha = 0$, and the 5-cell is attached to the S^4 by the identity. Therefore the inclusion of S^4 is nullhomotopic.

Proof of Theorem 1.3. Denote the 5-manifold pair by $(M, \partial_0 M)$. Decompose M as $W_1 \cup W_2$, where W_1 is a collar on $\partial_0 M$ union the 0-, 1-, and 2-handles of the first handlebody structure, and W_2 is a collar on $\partial_1 W$ union the 3-, 4-, and 5-handles. Decompose M similarly as $V_1 \cup V_2$, with respect to the second structure. The first step is to deform the second handlebody structure until W_1 is a subhandlebody of it. This is done by smoothing things as much as possible.

Let $S \subseteq \partial_0 M$ be a set with one point in each component; then $\partial_0 M - S$ has a smooth structure [13, part III]. This determines arcs $S \times I$ in the collar in W_1. The handles in W_1 are attached on neighborhoods of 0- and 1-spheres, so we may arrange by isotopy that they are disjoint from $S \times I$, and smooth [13, part III], [15]. This gives $W_1 - S \times I$ a smooth structure. V_1 is smoothed similarly.

Next, by isotopy rel $\partial_0 M$, we may assume the collars in V_1 and W_1 agree in a neighborhood of the arcs $S \times I$ in each. Isotope W_1 into V_1, rel $S \times I$, by shrinking the collar and pushing the handles off the duals of the handles in V_2. It is now possible to change the inclusion $W_1 \subseteq V_1$ by isotopy rel $\partial_0 M \cup S \times I$ to be smooth on $W_1 - S \times I$. This is because M is 5-dimensional, so [7] applies, and $H^3(M, \partial_0 M; \mathbf{Z}/2) = 0$.

Smooth manifolds have smooth handle structures which are unique up to deformation through 1-parameter families, so the structure on W_1 extends to one on V_1, which can be deformed to the original one. In other words, the second handlebody structure on M can be deformed to one which has V_1 as a subhandlebody. By fixing these handles we are reduced to consideration of V_2, so we may assume that one handle structure has no handles of index $\leqslant 2$. Repeat this argument to eliminate the dual handles of V_2; then the theorem is reduced to the case where one structure has no handles at all.

Suppose, then, that $M = N \times I$, and there is a second handle structure on $(M, N \times \{0\})$. According to the h-cobordism theorem of [3] the second structure can be deformed to one with no handles (recall we have assumed M 1-connected). The homeomorphism between these two product structures is a pseudoisotopy. According to Theorem 1.4 a pseudoisotopy is isotopic rel $N \times \{0\}$ to the identity. This isotopy completes the deformation of the second handle structure to the first.

We observe that a $(\delta, 1)$-connected controlled version of 1.4 (in the sense of [13]) would imply uniqueness of handlebody structures on all 5-manifolds. Proceed as above, but in the last paragraph note that the projection p: $M \to N$ is a $(\delta, 1)$-connected (δ, h)-cobordism over N for all $\delta > 0$. According to the controlled h-cobordism theorem [13, part III] there is a deformation of the second handlebody structure to a product structure of radius less than any given δ. This defines a δ pseudoisotopy [13, part IV]. Applying a controlled version of 1.4 would complete the deformation between the structures. The controlled version has the further advantage that it can be reduced to noncompact manifolds which can be smoothed, thereby avoiding the question of existence of 1-parameter families to get the argument started.

4. Cancelling handles

In this section we show how to cancel handles from a 1-parameter family of handlebody structures on a 1-connected manifold. If a pseudoisotopy can be joined to the identity by a finite 1-parameter family, then applying this repeatedly eliminates all the handles. This leaves an isotopy to the identity, proving Theorem 1.4 under the assumption that there is a 1-parameter family.

The problem is reduced to a study of isotopies of 2-spheres in a 4-manifold in §4.1. §§4.2–4.4 describe simplifications of the structure of such isotopies. In §4.5 a criterion for cancellation is given. Finally in §4.6 the criterion is verified in the topological case, and in the smooth case after sum with $S^2 \times S^2$.

4.1 Reduction to "eyes." Think of the product structure on $(M \times I, M \times \{0\})$ as a handlebody without handles, and a pseudoisotopy as a second such structure. Presume that the two structures are connected by a 1-parameter family of structures, with only finitely many handles. According to Hatcher and Wagoner [6, IV3, p. 214] there is a deformation of this family to one which has only 2- and 3-handles, no handle additions, and independent birth and death points (a nested family of "eyes." Beware that the reduction to a single eye requires dimension $\geqslant 5$, so does not apply here.) What this means is that at t_0 some number of cancelling 2, 3 handle pairs are introduced, then the 3-handle attaching maps are changed by ambient isotopy (all together), until at t_1 they again intersect only the corresponding 2-handles in single points. Then they are all cancelled.

We give a detailed view of such an isotopy. Let N denote the level between the 2- and 3-handles, A_i denote the dual boundary 2-spheres of the 2-handles, and B_i denote the boundary spheres of the 2-handles. The isotopy moves the B_i with respect to the A_i. Generically these will be transverse except at a finite number of values of the parameter. At these values intersection points appear or disappear in pairs, by a small "finger move" along an arc, or by a "Whitney move" across a 2-disk (see [4, Chapter 1]).

After a finger or Whitney move there is a small Whitney disk or arc respectively, which can be used to reverse the move. Reversing the time parameter in the isotopy interchanges the two types of moves.

In intervals where intersections do not change, the isotopy can be extended to an ambient isotopy of the union $A_i \cup B_i$. This can be used to compare the data for adjacent moves; suppose a move occurs at s, and no moves occur between s and t. Apply the ambient isotopy to the small arc or Whitney disk at s, to get a "large" arc or Whitney disk at time t. Doing this for the first move on each side of t gives two arcs, an arc and a disk, or two disks, at t.

The first observation is that if the data for these adjacent moves is disjoint at t, then they can be done simultaneously, or slid past each other to change the order of occurrence. This is because the moves take place in small neighborhoods of the data. Since two arcs, or an arc and a disk, are disjoint in a 4-manifold this means we can slide all the finger moves left to occur at one time t_f, and all Whitney moves right to a time t_w. The picture is then that the B_i are changed by finger moves all at once, move by joint ambient isotopy with the A_i, and then the extra intersections are removed by Whitney moves all at once. At some intermediate time we see two complete sets of Whitney disks for the extra intersections, and these essentially determine the isotopy. We will denote the disks coming from the finger moves by V_*, and the ones from the Whitney moves by W_*. The problem in dimension 4 is that the W and V disks may intersect. There can be interior intersection points, and since the boundaries are arcs on 2-spheres, there can be boundary intersections.

The cancellation problem can be reduced to a single pair of handles. Choose a pair, say A_1, B_1, and slide the finger and Whitney moves involving only these handles slightly to the inside, to occur at times $t_{f1} > t_f$ and $t_{w1} < t_w$. Then A_1,

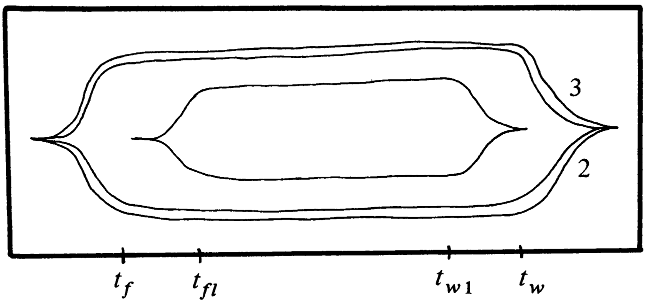


FIGURE 1

B_1 intersect in only one point from birth to just before t_{f1}, and from just after t_{w1} to death. They can therefore be cancelled on these intervals. The picture is now: all A_i, B_i except A_1, B_1 appear at once, and the B_i change by finger moves through the A_i at time t_f. Then A_1, B_1 appear, and B_1 changes by finger moves through A_1 at time t_{f1}. Next Whitney moves occur at t_{w1} to remove the extra intersections of A_1, B_1, and they are cancelled. After that the B_i are moved again, and are cancelled. The graphic is depicted in Figure 1. The family over the inner interval can be considered a 1-parameter family of structures on the intermediate level between A_i and B_i, $i \geqslant 2$. If this can be deformed to one in which A_1, B_1 always intersect in a single point, then they can be cancelled, thereby simplifying the original family. We caution that the A_1, B_1 birth is not "independent" in that the handles intersect the other handles when they appear. This means the inner "eye" in the graphic cannot be disengaged and moved outside the others.

4.2 Deformations of isotopies. To change a 1-parameter family of handlebodies we deform the isotopy of the middle level through isotopies. For this we describe changes in the families of disks which determine the isotopy. These deformations are smooth if the data used (disks, etc.) are smooth.

The simplest modification is isotopy of one family of disks, keeping A, B invariant. Whitney disks depending on a parameter yield isotopies depending on a parameter, hence a deformation. So for example we can do a finger move to push a V disk through a W, or a Whitney move to reduce VW intersections. We can also do "half" finger and Whitney moves of boundary curves of V, utilizing arcs and disks in A or B. We will find useful a move which involves pushing across a disk while pivoting one end around an AB intersection point. The data for this is a disk D in A (or B) bounded by one arc in ∂V, one arc in ∂W, and containing one AB intersection point and one VW boundary intersection. Push W across D holding the AB point fixed to remove the VW intersection (see Figure 2).

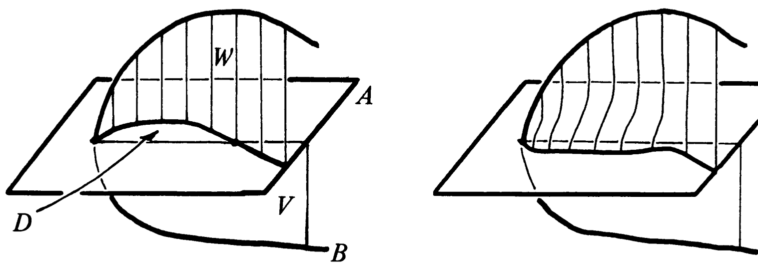


FIGURE 2

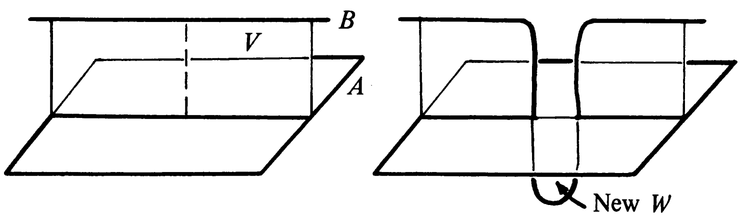


FIGURE 3

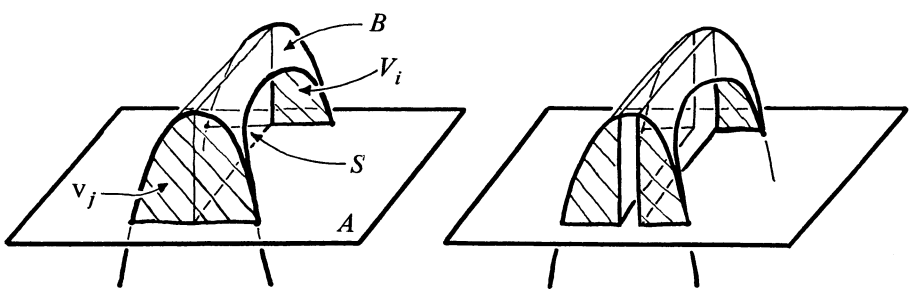


FIGURE 4

Splitting. Suppose an embedded arc is given on a V disk, going from the A side to the B side. Change B by a finger move along the arc through A. The V disk is split into two disks, and a new W disk is introduced (see Figure 3).

Sums. In this operation two V disks are cut apart and reassembled. The data is an embedded square S, with two edges on V disks, one on A, and one on B. Figure 4 provides a model. New V disks are obtained by cutting the given ones along the boundary edges of S, and glueing in two parallel copies of S. If the interior of S is disjoint from A, B and the V disks, then the new disks are embedded Whitney disks, so specify a new isotopy. To see a deformation between the two isotopies, imagine the original V disks pushed along S until they share an arc. When they touch they describe a "generalized" Whitney

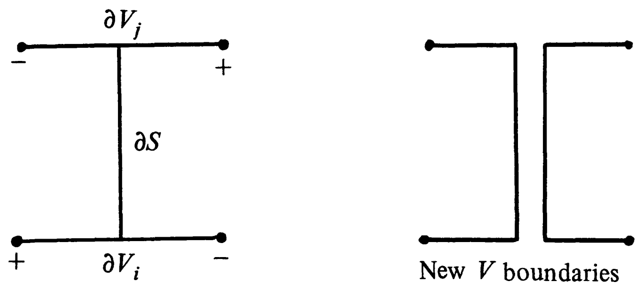


FIGURE 5

move, where B is pushed across $X \times I$, "X" denoting a cross-shaped graph. It is easy to arrange that the deformation of Whitney isotopies converges to this $X \times I$ move. Then note that the new V disks can be deformed to the same $X \times I$ object, so the isotopies they specify also can be deformed to the $X \times I$ move. Putting these together gives a deformation between the isotopies.

On the 2-sphere A (or B) this operation is depicted in Figure 5. Note in particular the boundary of the square divides the V arcs into two sides, and there is one $+1$ and one -1 intersection point on each side.

We begin assembling the data needed to construct sum squares. Suppose ∂f is a map from the boundary of a square which satisfies the evident necessary conditions; two edges are embedded in V, one in A, and one in B. Those in A and B should intersect V disks only in the endpoints and should have one $+1$ and one -1 intersection point on each side (in the sense above). Then ∂f extends to an immersion f of a neighborhood of the square in the model. To see this note that map ∂f extends to some immersion of a disk, which can be arranged to agree with the model except near some small interval on the boundary. Here the framings of the boundary from the data in the model, and from the contractibility of the disk may not agree. This can be fixed by spinning the disk [4, Chapter 1]. These immersions may intersect themselves and anything else 2-dimensional. The next step is to give ways to avoid these intersections.

4.3 Transverse spheres. We recall that "transverse spheres" for a family of surfaces $\{S_i\}$ consist of a family of framed immersed 2-spheres $\{S_i'\}$, so that each S_i intersects only S_i', and then only in one point.

The first remark is that the A_* have embedded transverse spheres, disjoint from each other and the B_*. Since the birth points are independent (before the modification at the end of 4.1), at birth the B_* themselves are embedded transverse spheres for the A_*. Denote by A^t parallel copies of the B_*. The B_*

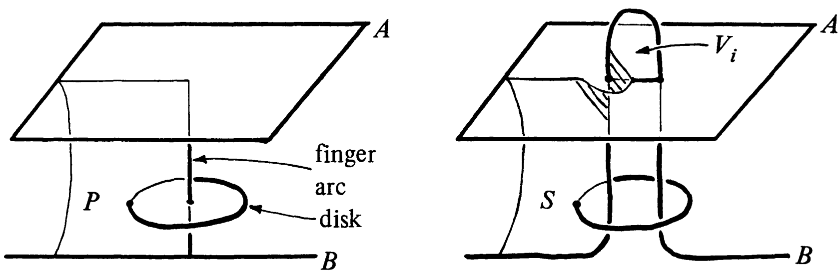


FIGURE 6

are then changed by finger moves through the A_*, which we may assume take place along arcs disjoint from the spheres A^t. This gives, at the intermediate time, embedded A^t disjoint from the B_* and the disks V_* from the finger moves.

Similarly the B_* have embedded transverse spheres disjoint from each other and the A_*. (Note however that the A^t and the B^t intersect each other.) This implies that the inclusion $N - A \cup B \cup V \to N$ is an isomorphism on fundamental group. It is onto by general position. To see it is injective, suppose a loop bounds a disk which intersects $A \cup B \cup V$. First push the disk off V through A or B. Then remove intersections with A and B by connected sums with A^t and B^t. This gives a disk which is disjoint from $A \cup B \cup V$. The same construction shows more precisely: if there is a framed immersion of a surface with boundary disjoint from $A \cup B \cup V$, then there is a framed immersion completely disjoint, and with the same framed boundary.

The next conclusion is that an immersed sum square has a framed embedded transverse sphere. More precisely suppose S is an immersion of the model which intersects V_i in exactly one boundary arc. Then there is a framed embedded sphere disjoint from $A \cup B \cup V$ which intersects S in exactly one point. The easiest way to see this is to temporarily undo the finger move, by pushing B across V_i. S is converted into a surface P with an edge on the finger move arc. A small linking 2-sphere to the arc intersects P in one point, and is disjoint from everything else. Now perform the finger move in a smaller neighborhood of the arc, to recover the original B, V, S.

Figure 6 shows A and P in the 3-dimensional "present," while B extends into the past and future. The linking 2-sphere is represented as a disk which intersects the finger arc in one point. The rim of the disk is in the present, and a sphere is obtained by shifting copies of the disk into both the past and future. S is also in the present except for a twist into the future near the AB intersection point, to get to V_i.

As a corollary of this we can conclude that if N is 1-connected, then an appropriate embedding of the boundary of a sum square extends to an embedding of the entire model, with interior disjoint from $A \cup B \cup V$. In 4.3 it was shown that an embedding of the boundary extends to an immersion of the model. The transverse spheres for A and B can be used to construct from this an immersion with interior disjoint from $A \cup B \cup V$. Then connected sums with the framed embedded transverse sphere disjoint from $A \cup B \cup V$ can be used to remove the self-intersections.

4.4 Simplification of boundary curves. We begin the actual proof of the Theorem 1.4. The situation is a pair of spheres A, B in a 1-connected 4-manifold N, and two collections V_*, W_* of Whitney disks which each eliminate all but one of the AB intersections. The V_* intersect A in disjointly embedded arcs, with AB intersection points as endpoints. The W_* intersect similarly, so the union is a collection of immersed circles and either one arc, or an isolated point. The objective of this section is to deform the situation to one in which $V \cup W$ intersects A and B in single embedded arcs.

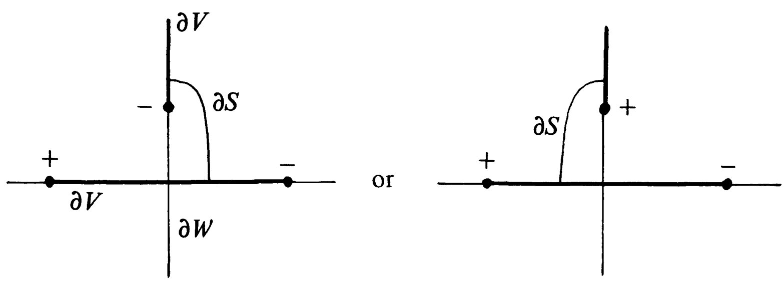


FIGURE 7

First we show how to reduce the number of intersections between ∂V and ∂W on A. Choose an intersection point which is adjacent on ∂W to an endpoint in common with a ∂V arc. Join the two ∂V components by an arc $\partial_0 S$ with interior disjoint from $V \cup W$ and with one $+1$ and one -1 V endpoint on each side, as in Figure 7. (If the two pieces of V in Figure 7 are actually on the same component of ∂V, then split it. This does not introduce intersection points.) This can be extended to an embedding of the boundary of a sum square, hence by 4.3 to an embedded sum square. Performing the sum operation does not reduce intersections, but it leaves a disk in A which can be used for the "rotating" move of 4.2 to remove an intersection (see Figure 8).

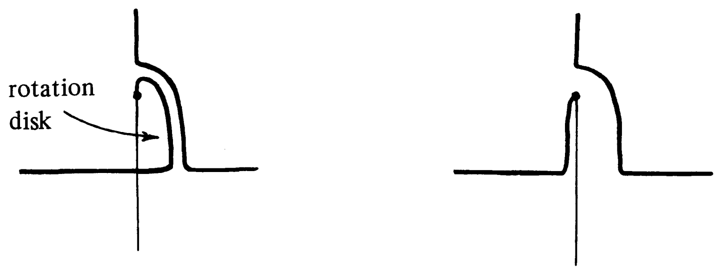


FIGURE 8

Proceeding by induction we can arrange that $\partial V \cup \partial W$ have no interior intersections on A. This means it is a union of embedded circles, and an arc or isolated point. We assume there is an arc, since an isolated point can be converted to an arc by introducing a trivial finger move, then cancelling the isolated point with one of the new intersections. By joining a V segment on an "outermost" circle to one on the arc and extending to a sum disk as above, the circles can be added to the arc. (By outermost we mean that it can be connected to the arc without crossing another circle.) Repeating this operation reduces $(\partial V \cup \partial W) \cap A$ to an embedded arc.

The next step is to improve the boundary arcs on B, but with as little damage as possible to the ones on A. During the procedure we suppose: The boundaries on A are the one embedded arc and disjointly embedded circles, all "outermost" and positively oriented with respect to an orientation of A. Note that going from the $+1$ toward the -1 intersection points on a V segment determines an orientation of the circle, so choice of an "outside" orients a neighborhood of the circle. The improvement above yields a single arc on A, so we begin with this hypothesis satisfied.

As above locate a useful $\partial_0 S$ arc in B, and add segments going across the V disks. We extend this to an embedding of the boundary of a sum square by joining the endpoints on A. If the endpoints do not lie on a single circle, then the "outermost" hypothesis implies that they can be joined by an embedded arc with interior disjoint from $\partial V \cup \partial W$. If they both lie on circles, then the orientation hypothesis implies that the $\pm$ intersection condition is satisfied. If one is on the arc, then coming in on the correct side of the arc gives the condition. If the endpoints lie on a single circle, then join them through the interior of the circle. In all cases when the resulting ∂S is extended to a sum square and the sum operation performed, we again get an arc and outermost circles. The orientations will be correct except possibly in one case. If both endpoints are on the arc, then they can be joined by a ∂S segment coming in on either side of the arc. One side splits off a circle with the incorrect orientation, but there is always a choice which gives the correct orientation.

Proceed as above to eliminate VW intersections on B, and then simplify the pattern to be an embedded arc. When this is done the pattern on A will also be an embedded arc, since for each V or W disk, its intersection arc with A joins the same pair of points of $A \cap B$ as its intersection with B.

4.5 A replacement criterion. *Suppose a* 1*-parameter family has a single* AB *pair, and* V *and* W *disks intersecting* A *and* B *in embedded arcs* (*i.e. the output of* 4.4). *Suppose there is a second set of disjoint Whitney disks* $W^{\sim}$ *with boundaries parallel to those of* W, *and with interiors disjoint from* W. *Then there is a deformation from the given family to the family with disks* V, $W^{\sim}$.

Unlike the manipulations of the previous sections, new handles are added and cancelled during this deformation. The argument is similar to an unpublished one of K. Igusa for pseudoisotopies of 5-manifolds.

Corollary. *Suppose* A, B, V, W *are as above, and there are disks* $W^{\sim}$ *with interiors disjoint from both* V *and* W. *Then there is a deformation of the pseudoisotopy to an isotopy.*

Proof of the corollary. According to 4.5 we can deform the family to the one with disks V, $W^{\sim}$. Since these disks are disjoint they can be "cancelled." By this we mean that by gradually pushing across more and more of $W^{\sim}$ before completing the finger move gives a deformation to an isotopy such that A and B always have a single point of intersection. The handles can then be cancelled, leaving an isotopy.

Proof of 4.5. Begin with the A, B, V, W family. Choose a product $A \times I$ near A, with A on one end and a sphere $A^{\sim}$ on the other. Displace copies of $W^{\sim}$ to give Whitney disks, disjoint from the W, for the $A^{\sim}B$ intersections (see Figure 9). We interpret this data as specifying a 1-parameter family: introduce the A, B as usual, then do the finger moves, then push B across *both* the W and $W^{\sim}$ to reduce both AB and $A^{\sim}B$ intersections back to one point.

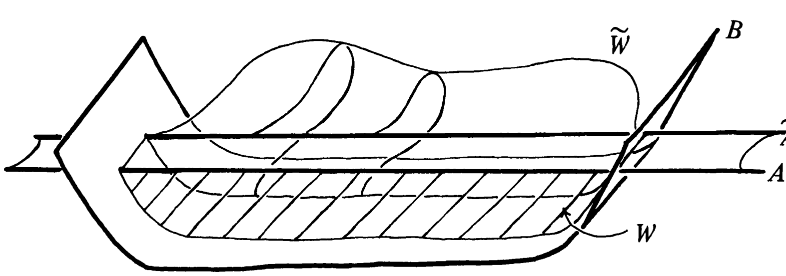


FIGURE 9

Then cancel A and B as usual. Since $A^{\sim}$ is not part of a handle, the push across $W^{\sim}$ does not effect the intersections. Gradually pushing across less and less of the $W^{\sim}$ therefore gives a deformation from the new family to the original one. If, by some maneuver, we could deform the new family so that $A^{\sim}$ rather than A was the dual sphere of the 2-handle, then the pushes across W would become inessential. Phasing them out would give a deformation to the family with V, $W^{\sim}$ data. Putting the deformations together would give the deformation required for 2.5.

The proposition is therefore reduced to finding handle moves which switch the dual sphere of the 2-handle from A to $A^{\sim}$. Consider the level below the 2-handle. This is obtained from N by cutting out A and replacing it by a 1-sphere. In particular the product $A \times I$ gets filled in to a 3-disk with $A^{\sim}$ as boundary. Use this disk to introduce a cancelling pair of 1, 2 handles. Proceeding downward this looks like adding a 3-handle on $A^{\sim}$, and then adding a 4-handle on the 3-sphere gotten from the core of the 3-handle and the 3-disk bounding $A^{\sim}$. Dually (from the bottom up) we see a 1- and a 2-handle, with $A^{\sim} \subset N$ as the dual sphere of this new 2-handle. Cancelling these near the ends of the family gives a new family with graphic depicted by Figure 10. Here the finger move of B through A takes place at t_f, and the Whitney move at t_w. The boundary sphere of the upper 2-handle intersects the dual of the 1-handle in a single point. Cancel these from before t_f to after t_w, to get a family with graphic depicted by Figure 11.

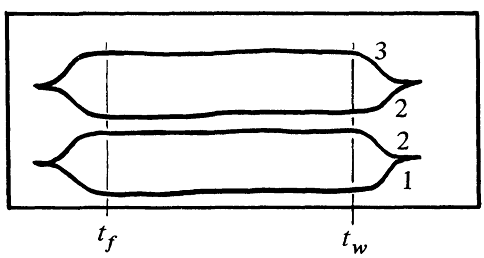


FIGURE 10

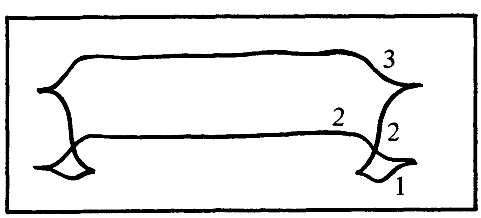


FIGURE 11

The 1-handle pieces on the ends intersect both 2-handles in exactly one point, so they can be cancelled by pushing through a dovetail singularity [6, Chapter V, 1.4]. The resulting family is as before, except that the sphere $A^{\sim}$ is now the dual sphere of the 2-handle.

4.6 Completion of the proof. Suppose the situation is as in 4.4; V and W disks with union intersecting A and B in embedded arcs. We show that W can be changed to have interior disjoint from V. Since 4.5 shows that such a situation can be cancelled, this completes the proof in the case under consideration, when the pseudoisotopy can be joined to the identity by a 1-parameter family of finite handlebody structures.

The first step is to arrange the disks in V and W to have trivial algebraic intersection numbers. Begin with a small linking sphere to the boundary arc on A of each component of V. These intersect V in one point, but intersect A in two points of opposite orientation. According to 4.3, A has an embedded transverse sphere disjoint from $B \cup W$. Add copies of this to the linking spheres to remove the intersections with A. This yields a collection of embedded spheres disjoint from $A \cup B \cup W$, and since copies were added in pairs with opposite orientation, the algebraic intersections with V are the same as transverse spheres. Adding copies of these to W at VW intersection points gives new disks $W^{\sim}$, which have trivial algebraic intersections with V. Since the W and the spheres are disjointly embedded, $W^{\sim}$ can be arranged to be embedded with interiors disjoint from W. Applying 4.5 exchanges $W^{\sim}$ for W, and accomplishes the desired improvement.

In the topological case, W satisfies the isotopy criterion of [3], [4] to be isotopic rel $A \cup B$ to have interior disjoint from V. In the smooth case, we now take connected sum with the identity pseudoisotopy on copies of $S^2 \times S^2$. This changes the intermediate level N in the handlebody by connected sum with $S^2 \times S^2$, so according to [14] after some number of such modifications there is a smooth isotopy of W rel $A \cup B$ to have interior disjoint from V. As above this implies that the handle pair A, B can then be cancelled, so completes the proof.

For the convenience of nonexperts we sketch how the isotopy theorems follow from the more familiar embedding theorems. First, it is possible to do finger moves of W through V so that V has framed immersed transverse spheres disjoint from $A \cup B \cup W$. For this note it was shown above that there are algebraically transverse spheres disjoint from $A \cup B \cup W$. Choose Whitney disks (using the 1-connectedness of N) for the extra intersections between these spheres and V. Make these disks disjoint from V by pushing through A, and then from A and B using transverse spheres disjoint from V.

Then make W disjoint from these disks by pushing off through V. This changes W by finger moves through V. Finally the spheres can be pushed across these Whitney disks to eliminate extra intersections with V, without introducing intersections with anything else.

Next, again use the 1-connectedness of M to choose immersed Whitney disks for all the VW intersections. Again use transverse spheres for A and B disjoint from W to get Whitney disks disjoint from $A \cup B \cup W$. Then use the transverse spheres for V just found to make them disjoint from V as well. Deleting an open tubular neighborhood of $A \cup B \cup W \cup V$ leaves a 1-connected manifold with framed immersed disks (the Whitney disks). These disks have algebraic duals (the linking torii of the VW intersections), so the topological embedding theorem applies to give disjointly embedded framed disks with the same framed boundaries. These are embedded Whitney disks for the VW intersections, so can be used to construct an isotopy of W disjoint from V.

In the smooth case, connected sums with $S^2 \times S^2$ introduce a lot of pairs of embedded spheres intersecting in points. If we form the connected sum of a framed Whitney disk with one of these spheres, the other gives a framed embedded transverse sphere for the result. These spheres can be used to remove intersections and self-intersections, yielding smoothly embedded Whitney disks. Then there is a smooth isotopy separating W from V.

This completes the sketch of the isotopy theorem. As a final note on the method we observe that the simplification procedure does not work for intersections between A and B spheres with different indices. Here rather than arcs we have circles of boundaries of W and V disks. When constructing V' we used connected sums with A' along arcs in A, to remove intersection of a linking sphere with A. With a circle of boundaries one of these arcs must cross the circle, introducing a nontrivial algebraic intersection between V' and W. It is therefore impossible to find transverse spheres V' disjoint from W.

5. Codimension 2

Suppose M is a compact 1-connected 4-manifold, and $X \subset M$ is obtained by adding handles of index $\leqslant 2$ to ∂M. Suppose that h is a pseudoisotopy which is the identity on ∂M.

Proposition. *h is isotopic rel ∂M to a pseudoisotopy which is the identity on X.*

The point is to prove this without making assumptions about handlebody structures, since otherwise the result of §4 implies h is isotopic to the identity on all of M.

This result and 1.5 complete the proof of the pseudoisotopy Theorem 1.4: apply it to the N of 1.5, then h is isotopic to h_1 which is the identity except on $M - N$, which is contained in a disk. But the Alexander trick (pushing toward 0) shows pseudoisotopies of the disk are isotopic to the identity. Applying this to h_1 gives an isotopy of h to the identity.

We may suppose that h is the identity on a ball B disjoint from X. Let 0 be the origin of B, and denote $M - \{0\}$ by M_0. h restricts to give a pseudoisotopy of M_0. Since M_0 is noncompact and $H^3(M_0, \partial M_0) = 0$, smoothing theory ([7] and [13, part III]) apply to show M_0 has a smooth structure, and that h is isotopic to a smooth pseudoisotopy. This means it is joined to the identity by a 1-parameter family of handlebodies, but with infinitely many handles. Only finitely many will intersect X. The plan of the proof is to cancel the handles in the family which intersect X, using the result of the previous section. Cancelling a handle rearranges the remaining handles in the family, and care is necessary to avoid introducing new intersections with X.

Let $M_0 \to [0, \infty)$ be the map which is (1/radius) $- 1$ on the ball, and takes the complement of the ball to 0. We will use the controlled topology of [13], with control in $[0, \infty)$. With this particular control space we can avoid a lot of technical detail. Generally in controlled arguments when $\varepsilon > 0$ is given there is some $\delta > 0$ so that data with δ control implies a conclusion with ε control. However if α control is given, any $\alpha > 0$, then we can obtain δ control by composing with a compressing homeomorphism $[0, \infty) \to [0, \infty)$. Allowing such compressions means we can speak generically of "control" without having to continually specify ε, δ, etc. Note also that control allowing compressions is equivalent to requiring that everything be proper.

5.1 Lemma (*Reduction to "eyes"*). *The* 1-*parameter family of handlebody structures on* $M_0 \times I$ *can be deformed to one with handles of controlled diameter, only* 2- *and* 3-*handles, independent birth and death points, and no handle additions.*

This is the controlled analog of 4.1, which comes from Hatcher and Wagoner [6, §IV3]. The clearest "reason" this should be so is that the arguments of Hatcher and Wagoner can be modified to work with control, the modifications being specified in [13], particularly part IV. This is satisfactory up to the elimination of handle additions, which requires the vanishing of a Wh_2 group. For the controlled analog, note M_0 is controlled relatively 1-connected over $[0, \infty)$ (because inverses of points in $(0, \infty)$ are copies of D^4, the inverse of 0 is M-int(B), and these are all 1-connected). The controlled analog of Wh_2 then vanishes for two reasons. First it is a locally finite homology group [13, II and IV], and the locally finite homology of $[0, \infty)$ is trivial with any (constant) coefficient spectrum. The second reason is that the controlled pseudoisotopy obstruction group for trivial local fundamental group vanishes over any space [13, IV, Theorem 1.2]. Once the controlled Wh_2 is known to vanish, the arguments of Hatcher and Wagoner can be used to eliminate handle additions in a controlled manner.

The difficulty with this "reason" is that the details of the controlled analog of Wh_2, and its connection to handle additions, have not yet been worked out. To get a technically complete proof of 1.5 we use the fact that in this situation higher-dimensional controlled pseudoisotopies are trivial, and use the proof of the stability theorem as far as possible. Since this is somewhat technical we postpone it until after using it to complete the proof of the proposition.

5.2 Proof of the proposition. As in 4.1 a family consisting only of "eyes" is determined by an isotopy of the intermediate level, which first introduces and then removes intersections of boundary and dual spheres. As in §4 we denote the level between the handles by N, the dual spheres of the 2-handles by A_*, and the attaching spheres of the 3-handles by B_*. Denote by V^{ij} the disks associated to the finger moves of B_i through A_j, and by W^{ij} the Whitney disks which remove these intersections.

When they first appear in the family the handles can be described as follows: Begin with a disjoint collection of embedded 2-disks $\{D_i\}$ in M_0. Then add 2-handles attached on the boundary circles of the disks to a collar on M_0, then 3-handles on the 2-spheres obtained by glueing together the disks and the cores of the 2-handles. The disjointness of the 2-disks corresponds to the "independence" of the births in the 1-parameter family.

The isotopy of N takes place in a neighborhood of this data, so the pseudoisotopy is isotopic to the identity on any part of M_0 which can be moved disjoint from the data. As a first application we note that the data is all 2-dimensional, so can be made disjoint from the spines of the 0- and 1-handles of X. The pseudoisotopy is therefore isotopic to the identity on these handles. The problem arises with the 2-handles of X, whose spines may intersect the data in isolated points.

Since X is compact and the D, V, W are controlled, there is some n so that D_j, V^{ij}, and W^{ij} are disjoint from X for $j \geqslant n$. We claim that the pair A_1, B_1 can be cancelled from the family by the method of §4 in such a way that the data for the remaining handles is still disjoint from X for $j \geqslant n$. If this is the case, then we can cancel the first $n-1$ handles in the same way, and end up with all the data disjoint from X. As above this means the pseudoisotopy is the identity on X, which is the desired conclusion of the proposition.

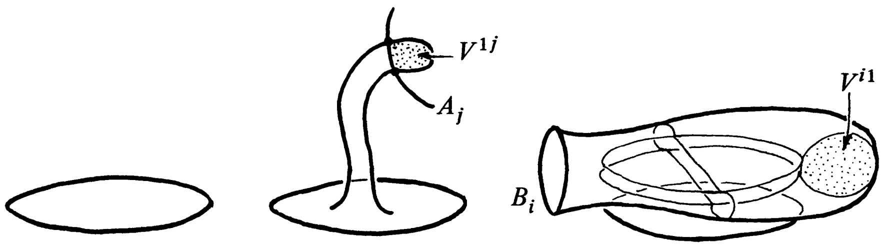
The modifications in 4.4 and 4.5 involve pushing things across disks. Note that if, in modifying the disks V^{11}, the modifications take place in the complement of the other V disks (which is easily arranged) then these other disks remain part of the data for the modified family. These modifications can usually not be arranged to be disjoint from the W disks. However the W intersect the modification disks in points, so are changed by finger moves. The 2-handles of X also may intersect the modification disks in points, but by general position the finger moves of W can be arranged to miss these points. The V modifications therefore do not introduce new XW intersections. Similarly the W modifications do not introduce new XV intersections.

In 4.5 it is shown that a W disk can be replaced by a disk $W^{\sim}$ provided it is disjoint from all W disks. The particular $W^{\sim}$ considered in 4.5, and constructed in 4.6, allowed modification of A_1, B_1 so that they always intersect in a single point during the isotopy, hence could be cancelled. The replacement does not change any of the other disks. The modification involves pushing across a disk, so as above does not introduce new intersections of the other disks and X. This leaves consideration of the cancellation.

Change the order of attachment in the handlebodies so that the 2-handles of index > 1 are attached before that of index 1. Let P denote the level between the first and other 2-handles; then P is the intermediate level in the handlebody obtained by cancelling the first pair of handles. As above the cancelling pair of handles arises from a disk D_1 in P. Changes of the pair A_1, B_1 in N which do not introduce new intersections between them can be realized (up to ambient isotopy) by changes in D_1. In particular we describe intersections with other handles from this point of view.

A finger move of B_1 through A_j corresponds to a finger move of the disk D_1 through A_j, with cancelling Whitney disk V^{1j}. An intersection of B_i with A_1 corresponds to a disk in B_i parallel to D_1. Therefore a finger move of B_i through A_1 corresponds to pushing B_i across a parallel of D_1, and the disk V^{i1} joins the two resulting pieces of B_i parallel to D_1 (see Figure 12). Putting these together we see that finger moves of B_1 through A_j and finger moves of B_i through A_1 yield, after cancellation, many finger moves of B_i through A_j. There is, however a choice of disks for these new moves. If we imagine the finger move of D_1 through A_j being done first, and then B_i pushed parallel to D_1, the new intersections have V disks parallel to V^{i1}. If we push B_i parallel to D_1 first, then do the finger moves of D_1 (and all its parallels), the new intersections get disks parallel to V^{j1}. By reversing this picture we see there is also a choice of Whitney disks W which undo these intersections.

Now consider an intersection formed in this way between A_j and B_i when one of the indices is at least n. According to the above we may arrange that the new V and W disks are parallel either to ones with index 1, j or i, 1. Choose the larger of i, j, which by supposition is at least n. Then by choice of n the



D_1 finger move of D_1 through A_j finger move of B_i parallel to D_1 (= through B_1)

FIGURE 12

original V, W disks are disjoint from X, so parallels will be also. This means that the new data after cancellation is still disjoint from X when one index is at least n. This completes the proof of the proposition.

5.3 Proof of Lemma 5.1. We will discuss technical details of proofs in [13, part IV], and in the remainder of this section references of the form Proposition IV.3.2 are to that paper.

First, it is possible to find a 1-parameter family with only 2- and 3-handles on $M_0 \times I$ joining the pseudoisotopy to the identity. (This is the controlled version of Hatcher-Wagoner [6, V.3.1].) Further, using the "exchange lemma" as in Hatcher-Wagoner [6, p. 214], it can be arranged that there are no handle additions among the 2-handles. The objective is to eliminate handle additions among the 3-handles.

The next step is to consider deformations of a stabilization of this family through 2-parameter families. These 2-parameter families will be used as guides for simplification of the 1-parameter family. An appropriate product family on $(M_0 \times D^k) \times I$ has a deformation through a 2-parameter family to one with no handles, and that only handles of two indices occur during the deformation. That there is a deformation to the trivial family is Theorem IV.1.2. That the deformation only involves two indices follows from the general simplification procedure of Hatcher and Wagoner if the total dimension $k + 4$ is at least 7. If $k + 4 \geqslant 6$ the proof of the stability Theorem IV.3.2 also gives such a deformation. The statement of the theorem includes $k + 4 = 5$, but the proof in that case uses a low-dimensional argument of K. Igusa (similar to 4.5 in this paper) which involves three indices.

Now consider the proof of Proposition IV.3.2, which simplifies the 2-parameter family. At the end of the proof of the lemma on 1-parameter families we cannot eliminate death points. Therefore when $m = 4$ that statement should assert there is a deformation to a family with graphic constant except for death points. In the main body of the proof this is used to clear lines in the trace off an arc. When $m = 4$ we can clear all lines off except intersections with death curves. But using the uniqueness of death these curves can usually be pushed off the $\{1\}$ end of the arc. This can be done in steps 1, 2, 4, and 5 of the proof, to eliminate dovetails, $i/i+1$ points, crossings of addition curves, and closed loops.

The 2-parameter family is parametrized by $I \times I$, with the family on $M_0 \times I$ over $\{0\} \times I$ and no handles over the rest of the boundary of $I \times I$. After the modifications above the birth-death and addition curves are generically arcs and circles in $I \times I$. We claim we can modify the family so that on each component projection to the second coordinate is single-valued. To see this consider horizontal arcs to points where a curve locally fails to be single-valued. Use the lemma to clear everything but death curves off this, and note no multivalued phenomena is introduced. Push the death curves out past the end of the arc, again in a convex way, and pull the bad point on the curve along the arc and off the edge.

We can now distinguish curves of births and deaths, depending on which happens when we cross the curve going horizontally from $\{0\} \times I$. Birth curves can be pulled over the edge using the independence of birth hypothesis. (In the graphic on the edge this corresponds to joining eyes which lie over disjoint intervals.) Death curves can be pushed out past the addition curves, near the boundary of $I \times I$. Inside the death curves this leaves disjoint addition curves, with their boundaries on $\{0\} \times I$.

Next we may arrange that for any given handle the addition curves occur in unnested parallel collections. Again this is done by pushing pieces of the curves off the edge along horizontal lines (see Figure 13).

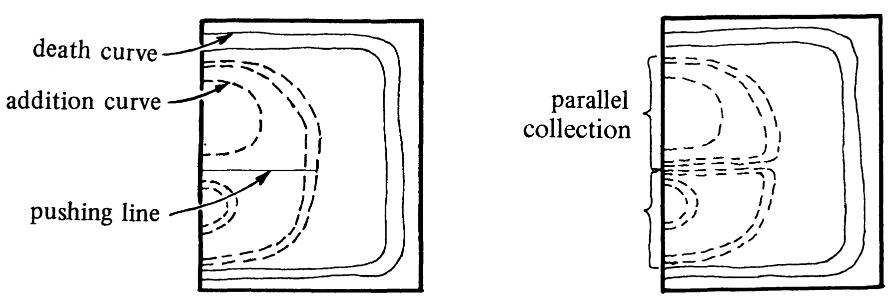


FIGURE 13

In the intermediate level N, passing through an addition corresponds to specifying an arc from some B_i sphere to a B_j, and changing B_i by connected sum with a parallel copy of B_j. Passing through additions on the ends of one of these families corresponds to sums along a collection of disjoint arcs from B_i to various B_j. Starting on both sides of the collection (above and below on the left boundary, in Figure 13) and moving into the interior specifies two such collections of arcs. The fact that the additions are endpoints of addition curves in the 2-parameter family means that these arcs are homotopic. But this means the picture obtained by going all the way through both ends of the collection is: B_i is changed by connected sum along embedded arcs, then is changed by "undoing" sums along homotopic arcs. Such inverse pairs of additions can be omitted from the family (Hatcher-Wagoner [6, p. 83]).

Applying this to all of the collections of addition curves eliminates all of the additions from the 1-parameter family. The point of arranging the curves in unnested collections is that this can be done all at once, therefore with control. The result therefore satisfies the conditions in the conclusion of the lemma.

6. The decomposition

In this section a 1-connected 4-manifold is shown to have a decomposition as described in Proposition 1.5. This is first done in the case M has a handlebody structure, then the general case is done using an infinite handlebody structure in the complement of a point.

Suppose M is a finite handlebody, and denote by C_* the cellular chain complex with $\mathbf{Z}$ coefficients. Since $H_1(M) = 0$ there is a splitting homomorphism $s\colon C_1 \to C_2$ so that $C_2 = (\operatorname{im}\partial) \oplus (\operatorname{im} s\partial)$. Stabily these summands are free, and we can choose bases for them so that the resulting basis for C_2 is related to the cellular basis by elementary matrices. Stabilization can be realized by adding cancelling pairs of handles, and elementary matrices can be realized by sliding handles over other handles. Therefore we can arrange that the summands $(\operatorname{im}\partial)$ and $(\operatorname{im} s\partial)$ are generated by handles.

Define Δ to be the union of the 0- and 1-handles, and the 2-handles in $(\operatorname{im} s\partial)$, and let N be the remainder of the handlebody. Dualizing we see that N is obtained from ∂M by adding handles of index $\leqslant 2$, as desired for the proposition. By construction Δ has the homology of a point, and the proposition is completed by showing that Δ is contained in a disk.

The boundary Σ of Δ is a homology 3-sphere. By Freedman [3] there is a 1-connected homology H-cobordism of Σ to S^3, say W. (More directly, note Δ minus an open ball gives an H-cobordism, and W can be constructed by killing the fundamental group of this.) We get an h-cobordism rel boundary from N

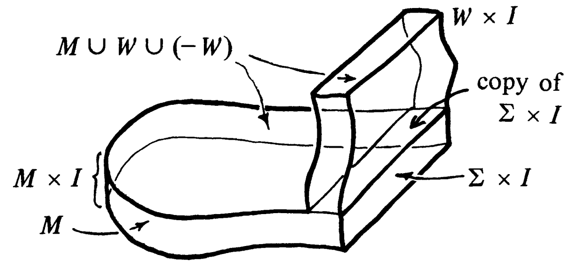


FIGURE 14

to $N \cup_{\Sigma} W \cup_{S} (-W)$ by attaching $W \times I$ to $N \times I$ on a copy of $\Sigma \times I$ in $N \times \{1\}$ (see Figure 14). The h-cobordism theorem implies that these two manifolds are homeomorphic, so Δ is contained in a copy of $(-W) \cup \Delta$ inside M. However $(-W) \cup \Delta$ is a homotopy 4-disk (1-connected because W is) with boundary S^3, so is homeomorphic to D^4.

This completes the proof when M has a handlebody structure. In general begin with a ball $B \subset M$, and denote $M - \{0\}$ by M_0. Let $M_0 \to [0, \infty)$ be the map which is (1/radius) $- 1$ on the ball, and takes the complement of the ball to 0. As in the previous section (see the note there) we use controlled topology with control in $[0, \infty)$. Let J be a regular neighborhood of a radial arc in the ball. Note the pair $(M_0 - \operatorname{int} J, \partial J)$ is controlled relatively 1-connected over $[0, \infty)$. This is because inverses of points in $(0, \infty)$ are copies of D^4 and the inverse of 0 is $M - \operatorname{int} B$, and these are all 1-connected.

According to [13, part III] M_0 has a smooth structure in which we may assume J is a smooth submanifold. Choose a handlebody structure on M_0 obtained by adding handles of controlled diameter to J. Let C_* $(= C_*(M_0, J))$ be the relative cellular chains, considered as geometric modules on $[0, \infty)$. The controlled 1-connectivity over $[0, \infty)$ implies there are controlled homomorphisms $s\colon C_0 \to C_1$ and $s\colon C_1 \to C_2$ such that $s\partial + \partial s = 1$ on C_1, and $\partial s = 1$ on C_0. Using s and simple chain constructions one can change C_* by stabilization and changing bases by controlled elementary matrices until it is a direct sum of $C_4 \to C_3 \to D$ and $A \to A \oplus B \to B$ (the dimensions in the latter are 2, 1, 0). This corresponds to the splitting of C_2 in the finite handle case. The algebraic manipulations can be realized by controlled handle moves, so there is a controlled handlebody structure whose cellular chain complex has this form. Define V to be the union of handles corresponding to $A \to A \oplus B \to B$, and N the rest of the manifold.

As above N is obtained from ∂M by adding handles of index $\leqslant 2$, but there are infinitely many of them. Since they are controlled with respect to the inverse of the radius in B, they converge to the point $0 \in B$. If we can find a locally flat 4-disk which contains $V \cup J \times \{0\}$, then it can be enlarged a little to contain all but finitely many handles of N. The enlarged disk and these finitely many handles then satisfy the statement of the proposition.

Define $V \cap N = \Sigma$. By construction V is a controlled $H(=$ homology$)$-cobordism between $\partial J = \mathbf{R}^3$ and Σ. Do controlled 1- and 2-surgeries on V to obtain a controlled 1-connected H-cobordism, say W. For this particular V this process can be described fairly explicitly. Cancel the 0-handles of V, introducing only 1- and 2-handles in the process [13, I, §6]. Then we get a structure on V with only 1- and 2-handles, and boundary homomorphism $C_2 \to C_1$ a controlled isomorphism. Join the ends of the 1-handles by embedded arcs to get framed embedded circles in V. Do surgery on these, then the result has only 2-handles, with chain group $2C_2$. Half a basis for this is represented by framed embedded 2-spheres (the ones coming from the surgery). According to the controlled embedding theorem of [13, part III] there are framed embedded 2-spheres controlled homotopic to these, which additionally have controlled 1-connected complement. Do surgery on these spheres to obtain W.

As above there is a controlled h-cobordism from N to $N \cup W \cup (-W)$. The controlled h-cobordism theorem [13, part III] asserts this has a product structure, so the ends are homeomorphic. In particular $V \cup J$ is contained, in M_0, in a copy of $(-W) \cup V \cup J$. This is contractible, 1-connected at ∞, and has boundary $\mathbf{R}^3$, so Freedman's characterization identifies it as a half-space. Adding the point $\{0\}$ gives a 4-disk. Denote this disk $(-W) \cup V \cup J \cup \{0\}$ by D.

According to the above, the proof will be completed by seeing that D is locally flat in M. It is except possibly at the point $\{0\}$, but a theorem of Kirby (see [3]) asserts that nonlocally flat points cannot be isolated in this dimension. Therefore it is locally flat at $\{0\}$ as well. This also follows, of course, from the much more general flattening theorems of [13, part III].

References

[1] D. Burghelea, R. Lashof & M. Rothenberg, *Groups of automorphisms of compact manifolds* (with an appendex by E. Pedersen), Lecture Notes in Math., Vol. 473, Springer, Berlin, 1975.

[2] J. Cerf, *La stratification naturelle des espaces de fonctions differentiables reeles et le theoreme de la pseudo-isotopie*, Inst. Hautes Études Sci. Publ. Math. **39** (1970).

[3] M. Freedman, *The topology of four-dimensional manifolds*, J. Differential Geometry **17** (1982) 357–453.

[4] M. Freedman & F. Quinn, *Topology of 4-manifolds*, in preparation.

[5] A. Hatcher, *A proof of the Smale conjecture*, Diff$(S^3) = O(4)$, Ann. of Math. **117** (1983) 553–607.

[6] A. Hatcher & J. Wagoner, *Pseudo-isotopies of compact manifolds*, Asterisque 6, Soc. Math. France, 1973.

[7] R. Kirby & L. Siebenmann, *Foundational essays on topological manifolds, smoothings and triangulations*, Ann. Math. Studies, No. 88, Princeton University Press, Princeton, NJ, 1977.

[8] M. Kreck, *Isotopy classes of diffeomorphisms of* $(k - l)$*-connected almost-parallelizable* $2k$*-manifolds*, Algebraic Topology, Aarhus 1978, Lecture Notes in Math., Vol. 763, Springer, Berlin, 1979, 643–661.

[9] R. Lashof & J. Shaneson, *Smoothing* 4*-manifolds*, Invent. Math. **14** (1971) 197–210.

[10] R. Lashof & L. Taylor, *Smoothing theory and Freedman's work on four manifolds* (Aarhus 1982 proceedings) Lecture Notes in Math., Vol. 1051, Springer, Berlin, 1984, 271–292.

[11] J. Milnor, *On simply-connected* 4*-manifolds*, Sympos. Internac. Topologia Algebraica (Mexico 1966), Mexico, 1968, 122–128.

[12] B. Perron, *Pseudo-isotopies et isotopies en dimension quatre dans la categorie topologique*, preprint.

[13] F. Quinn, *Ends of maps*. I, Ann. of Math. **110** (1979) 275–331; II, Invent. Math. **68** (1982) 353–424; III, J. Differential Geometry **17** (1982) 503–521; IV, Amer. J. Math., to appear.

[14] ______, *The stable topology of* 4*-manifolds*, Topology Appl. **15** (1983) 71–77.

[15] ______, *Smooth structures on* 4*-manifolds*, Contemp. Math. **35** (1984) 473–479.

[16] C. T. C. Wall, *Surgery on compact manifolds*, London Math. Soc. Monographs, No. 1, Academic Press, 1970.

[17] G. W. Whitehead, *Elements of homotopy theory*, Graduate Texts in Math., No. 61, Springer, Berlin, 1978.

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY