THE 4 DIMENSIONAL POINCARÉ CONJECTURE

DANNY CALEGARI

ABSTRACT. This book aims to give a self-contained proof of the 4 dimensional Poincaré Conjecture and some related theorems, all due to Michael Freedman [7].

CONTENTS

PREFACE

My main goal writing this book is threefold. I want the book to be clear and simple enough for any sufficiently motivated mathematician to be able to follow it. I want the book to be self-contained, and have enough details that any sufficiently motivated mathematician will be able to extract complete and rigorous proofs. And I want the book to be short enough that you can get to the 'good stuff' without making too enormous a time investment.

Actually, fourfold — I want the book to be fun enough that people will want to read it for pleasure. This pleasure is implicit in the extraordinary beauty of Freedman's ideas, and in the Platonic reality of 4-manifold topology itself.

I learned this subject to a great extent from the excellent book of Freedman–Quinn [9] and Freedman's 2013 UCSB lectures (archived online at MPI Bonn; see [8]), and the outline of the arguments in this book owes a great deal to both of these. I taught a graduate course on this material at the University of Chicago in Fall 2018, and wrote up notes as I went along. As such, the main reason that there are not substantially more errors is due to the careful eyes of the attendees of that course, especially those of Emmy Murphy, who kept bringing me back to key points that I'd hitherto missed completely.

Date: March 27, 2019.

1. The Poincaré Conjecture

The *Generalized Poincaré Conjecture* says that a closed simply-connected n-manifold M with the homology of an n-sphere is homeomorphic to an n-sphere. The adjective *closed* for a manifold means compact without boundary, and for a space to have the homology of an n-sphere is for its reduced homology to be equal to $\mathbb{Z}$ in dimension n, and 0 elsewhere. Notice that stated in this way, the case $n = 1$ is vacuous, since there are no closed simply-connected 1-manifolds. The *Poincaré Conjecture* is the special case that $n = 3$.

1.1. Homotopy spheres.

Lemma 1.1. *For $n > 1$, a closed simply-connected n-manifold M has the homology of S^n if and only if it is homotopy equivalent to S^n.*

Proof. One direction is clear: $\pi_1(S^n) = 0$ for $n > 1$, and homology is a homotopy invariant. Conversely, suppose M is simply-connected, and the reduced homology is nontrivial and equal to $\mathbb{Z}$ only in dimension n. By Hurewicz, $\pi_n(M) = \mathbb{Z}$ and $\pi_k(M) = 0$ for $k < n$. Thus there is a map $f : S^n \to M$ inducing an isomorphism on π_n, and therefore also on homology in every dimension. Again, since M is simply-connected, by relative Hurewicz f induces an isomorphism in π_k for *all* k. Thus, by Whitehead, f is a homotopy equivalence. □

We can therefore reformulate the Generalized Poincaré Conjecture as the conjecture that an n-manifold homotopy equivalent to the n-sphere is homeomorphic to the n-sphere.

1.2. Statement of results. One usually refines the Poincaré Conjecture slightly, depending on the category one is interested in. So: one asks how many oriented PL n-manifolds, or how many oriented smooth n-manifolds, homotopic to S^n, are there are up to orientation-preserving isomorphism?

The answers to these questions, up through dimension 12, are summarized in the following table.

Table 1. closed manifolds homotopic to S^n by category

$n =$	1	2	3	4	5	6	7	8	9	10	11	12
TOP	1	1	1	1	1	1	1	1	1	1	1	1
PL	1	1	1	?	1	1	1	1	1	1	1	1
DIFF	1	1	1	?	1	1	28	2	8	6	992	1

In particular, the Generalized Poincaré Conjecture is known in every dimension. The answers in the **PL** and **DIFF** case are unknown in dimension 4, but are known to be equal.

A smooth n-manifold M is said to be an *exotic sphere* if it is homeomorphic to S^n but not diffeomorphic.

1.3. An exotic 7-sphere. The entries for 7 and 11 in the third row of Table 1.2 stand out. A smooth oriented manifold M of dimension 7 always bounds a smooth oriented manifold W of dimension 8, by Thom's theory of cobordism. For a *closed* smooth oriented W of dimension $4n$ the Hirzebruch signature formula says that the signature $\sigma(W)$ is related to the Pontriagin classes of the tangent bundle $p_j := p_j(W)$ by the formula

$$\sigma(W) = L(p_1, p_2, \cdots)[W]$$

where L is a universal power series whose homogeneous term of degree $4n$ is a rational polynomial L_n, where

$$L_0 = 1, \quad L_1 = \frac{p_1}{3}, \quad L_2 = \frac{7p_2 - p_1^2}{45}, \quad L_3 = \frac{62p_3 - 13p_2p_1 + 2p_1^3}{945}, \quad \cdots$$

Now let W^{4n} be smooth and oriented with boundary M. If M is a homology sphere, every Pontriagin class p_j of W comes from a unique class in the relative cohomology $H^*(W, M)$ with the exception of the top class p_n. Let $L_n[W]$ denote the polynomial in these relative classes, setting p_n formally to zero, and consider the difference $\sigma(W) - L_n[W]$. If W' is another smooth oriented manifold with boundary M, we may glue W to $-W'$ (i.e. W' with the opposite orientation) along M, and then apply the signature formula for closed manifolds.

Now, $\sigma(W \cup -W') = \sigma(W) - \sigma(W')$ because M is a homology sphere. Likewise, by Mayer-Vietoris and the fact that M is a homology sphere, $p_j(W \cup -W') = p_j(W) - p_j(W')$ for every $j < n$ (here we abuse notation, writing e.g. $p_j(W)$ for the class in $H^{4j}(W \cup -W')$ induced by inclusion $W \to W \cup -W'$).

We deduce that

$$\sigma(W) - L_n[W] = \sigma(W') - L_n[W'] + L_{n,n}[W \cup -W']$$

where $L_{n,n}$ is the part of L_n involving only p_n. For example,

$$L_{0,0} = 1, \quad L_{1,1} = \frac{p_1}{3}, \quad L_{2,2} = \frac{7p_2}{45}, \quad \cdots$$

For a smooth oriented 7-manifold M which is a homology sphere, this implies that

$$2p_1^2[W] - \sigma(W) \bmod 7$$

is independent of the choice of W.

Suppose therefore that we can find a smooth manifold M which is *homeomorphic* to S^7, and which smoothly bounds some W as above, but for which $2p_1^2[W] - \sigma(W)$ is not divisible by 7. It follows that M does not smoothly bound a ball, or else the difference would be zero. Thus, such an M would be an exotic 7-sphere.

Milnor constructed such exotic 7-spheres as follows. Let M be a smooth S^3 bundle over S^4. Then M smoothly bounds a D^4 bundle over S^4. The homology of M can be computed from the Serre spectral sequence: there is a map $d_4 : \mathbb{Z} \to \mathbb{Z}$ which is multiplication by e, the Euler class of the bundle. Thus, M is a homology S^7 (and in fact a homotopy S^7) provided $e = \pm 1$.

Oriented S^3 bundles over S^4 are classified up to homotopy by the class of the clutching map $S^3 \to \mathrm{SO}(4, \mathbb{R})$. Now, $\mathrm{SO}(4, \mathbb{R})$ is double-covered by $S^3 \times S^3$, which acts on S^3 by left and right multiplication of unit quaternions respectively. Thus, $\pi_3(\mathrm{SO}(4, \mathbb{R})) = \mathbb{Z} \oplus \mathbb{Z}$, and the homotopy type of M is determined by a pair of integers k, l where the clutching map is homotopic to the map taking $u \in S^3$ to the element $v \to u^k v u^l \in \mathrm{SO}(4, \mathbb{R})$. The Euler number e and the class p_1 in $H^4(W) = H^4(S^4) = \mathbb{Z}$ are linear in k and l. Thus by computing on two suitable examples, we get that $e = k + l$ and $p_1 = 2(k - l)$ (up to a choice of orientation).

Letting k, l be arbitrary with $k+l = \pm 1$ and $2(k-l) \neq 0$ mod 7, we obtain a homotopy 7-sphere which is not diffeomorphic to S^7.

1.4. **Handlebodies.** How can we show Milnor's manifolds are homeomorphic to S^7 via smooth techniques without showing (erroneously!) that they are diffeomorphic? The answer is to use Morse theory.

If M is a closed, smooth manifold, a *Morse function* $f : M \to \mathbb{R}$ is a smooth function with isolated critical points, and at each such point p the Hessian Hf is a *nondegenerate* quadratic form on T_pM. This implies that near p, there are smooth local coordinates $x := x_1, \cdots, x_n$ so that the function f has the form

$$f(x) = f(0) + x_1^2 + \cdots + x_k^2 - x_{k+1}^2 - \cdots - x_n^2$$

The *index* of the critical point is the number of negative eigenvalues of Hf.

We can obtain a nice description of the topology of M from this Morse function; to do so requires us to discuss the theory of handles, which we now do.

Suppose N is a smooth compact manifold possibly with boundary, and let K be a smooth S^{q-1} embedded in N with a trivial normal bundle. A *framing* of K is a smooth trivialization of a tubular neighborhood of K; i.e. a diffeomorphism to $S^{q-1} \times D^{n-q}$ taking K to $S^{q-1} \times 0$.

A *q-handle* is a smooth product of disks $D^q \times D^{n-q}$ (note that its interior is diffeomorphic to D^n). The *core* of the handle is $D^q \times 0$ and the *co-core* is $0 \times D^{n-q}$. The boundary of the q-handle is the union $S^{q-1} \times D^{n-q} \cup D^q \times S^{n-q-1}$; the first factor is the *attaching part* of the boundary.

Now both N and the q-handle have a parameterized $S^{q-1} \times D^{n-q}$ in their boundary, and we can glue these together. The result

$$N' := N \cup_{S^{q-1} \times D^{n-q}} D^q \times D^{n-q}$$

is said to be obtained from N by attaching a q-handle along the framed knot K. Note that the homotopy type of N' does not depend on the framing but only on the knot K. Notice that in this gluing the knot K is identified with the boundary of the core of the attaching handle.

Although the interior of N' is smooth, the boundary has a 'corner' along $S^{q-1} \times S^{n-q-1}$. However, this corner can be smoothed in a canonical way to give N' the structure of a smooth manifold with boundary; the details are neither difficult nor important.

Now we return to our smooth manifold M with its Morse function f. Enumerate the critical points as $p_1, \cdots, p_m$ where $f(p_j) = t_j$ and let's suppose $t_j > t_i$ if and only if $j > i$ (this can be obtained by reordering and a smooth perturbation). For each real t, define $M_t = f^{-1}(-\infty, t]$. For $t < t_1$ the manifold M_t is empty. The manifold M is obtained from the empty manifold by an explicit finite process tied to the Morse function, as follows.

(1) If $[s, t]$ is an interval disjoint from the t_j, then the gradient flow induces a smooth isotopy from ∂M_s to ∂M_t and a diffeomorphism from M_s to M_t.

(2) If $[s, t]$ is an interval with a single t_j in the interior, and p_j is a critical point of index q, then M_t is obtained from M_s by attaching a q-handle.

In the latter case, we describe the framed attaching knot K for the q-handle as follows. Recall that near a critical point of index q, the function f has the form

$$f(x) = f(0) + x_1^2 + \cdots + x_{n-q}^2 - x_{n-q+1}^2 - \cdots - x_n^2$$

In these local coordinates, the subspace where the first $n - q$ coordinates are zero is the *descending manifold*, and the subspace where the last q coordinates are zero is the *ascending manifold*. Thus the descending manifold has dimension q, and the link of 0 in this manifold is a sphere K of dimension $q - 1$. There is a natural framing of K, given by the translates of the ascending manifold.

Now let's return to Milnor's homotopy spheres M. Milnor writes down an explicit Morse function on M, and shows that it has exactly two critical points: a minimum (where the index is 0) and a maximum (where it is 7). The complement of the maximum is diffeomorphic to the interior of a single 0-handle, which is to say, it is diffeomorphic to $\mathbb{R}^7$. Since M is obtained from this complement by the 1-point compactification. we deduce that it is *homeomorphic* to $D^7 \cup \infty = S^7$.

1.5. **General Position.** Smale's theory of smooth manifolds depends on the use of two powerful tools: Morse theory, and general position. We have discussed the first; now we explain the second.

For integers a b both less than or equal to n, we let $\alpha : \mathbb{R}^a \to \mathbb{R}^n$ and $\beta : \mathbb{R}^b \to \mathbb{R}^n$ be affine linear injections. It is evident that there are an open dense set of choices of α or β or both for which $\alpha(\mathbb{R}^a) \cap \beta(\mathbb{R}^b)$ is an affine subspace of dimension $a + b - n$ where $a + b - n < 0$ means the intersection is empty. For such a generic choice we say these subspaces are in *general position*.

Now for smooth manifolds A^a, B^b, N^n of these dimensions, we say that immersions $\alpha : A^a \to N^n$ and $\beta : B^b \to N^n$ are in general position if there are local charts around the singular set diffeomorphic to the linear model. This means in particular that the intersection of A^a and B^b and the self-intersections of A^a are smooth immersed submanifolds of dimension $a + b - n$ and $2a - n$ respectively.

Proposition 1.2. *For smooth manifolds M, N (not necessarily connected, or with all components of the same dimension) the set of immersions in general position is open and dense in the space of immersions with the C^∞ topology.*

One application of general position is to Morse theory. Suppose p and q are critical points of f with index a and b where $f(p) < f(q)$ but $a > b$. Suppose further that these critical points are adjacent, so that there are no critical r with $f(p) < f(r) < f(q)$. The ascending manifold of p has dimension $n - a$ and the descending manifold of q has dimension b. It follows that we can modify f slightly so that these manifolds are in general position, and therefore *disjoint*. Then observe there is a subset of the descending manifold of q diffeomorphic to D^q, with $f(\partial D^q) < f(p)$. We can easily modify f in a neighborhood of this D^q, without creating new critical points or changing indices, so that at the end $f(p) > f(q)$.

The conclusion is that we can build a Morse function on M which is *self-indexing*; this means that every critical point p of index i has $f(p) = i$. In particular, the submanifold $M_{i+1/2}$ can be isotoped (by pushing along descending flowlines) to a regular neighborhood of a finite embedded CW complex of dimension $\le i$.

Notice that if f is Morse, so is $-f$, and if p is a critical point of f of index i, then p is a critical point of $-f$ of index $n-i$. It follows that we can split M with a self-indexing Morse function into two submanifolds $M_{i+1/2}, M - M_{i+1/2}$, each of which can be isotoped disjointly into disjoint regular neighborhoods of complexes of dimension $\le i$ and $\le n-i-1$ respectively.

1.6. **Engulfing.** We now prove the following theorem: if M^n is a smooth homotopy sphere of dimension at least 7, then it is homeomorphic to S^7. This shows that Milnor's examples are exotic but not fake.

This theorem was first proved by Smale; we give his proof in the sequel. But first we give an elegant argument due to Stallings (and independently Zeeman), discovered only days after Smale's announcement. The proof we give depends on an intermediate result — the Schoenflies Conjecture — that was known at the time; we defer the proof of this result until the next section.

Theorem 1.3 (Engulfing)**.** *Let M be a smooth homotopy sphere of dimension at least* 7*. Then M is homeomorphic to S^7.*

Proof. Let f be a self-indexing Morse function on M. Then we can split M along a level set between level $\lfloor n/2 \rfloor$ and $\lfloor n/2 \rfloor + 1$ into pieces X and Y, each of which is isotopic to a regular neighborhood of a subcomplex of dimension at most $k \le \lfloor n/2 \rfloor$. We shall show that X and Y are embedded ('engulfed') in the interior of disjoint balls B_X, B_Y. From this the theorem follows; for, ∂B_X is a smooth S^{n-1} contained in the interior of B_Y, and the Schoenflies Conjecture says that the complement of $B_X \cap B_Y$ in B_Y is homeomorphic to a ball. Thus M is the union of two balls glued along their boundary, and is therefore homeomorphic to a sphere.

It suffices to show that X and Y can be engulfed. Since M is a homotopy sphere, if we remove a point p disjoint from X, the complement $M - p$ deformation retracts to a point. Thus the inclusion of X into M extends to a map $CX \to M$. If this map were an embedding, we could push the points of X by an isotopy along the intervals of the cone, and produce an isotopy into a small neighborhood of a point; evidently this would imply that X can be engulfed.

We can't expect this map to be an embedding, but if we put it in general position, we can arrange for the singular set to be (a regular neighborhood of) a subset Σ of CX of dimension at most $2k - n + 2$. The cone $C\Sigma$ embeds in CX, and we can isotop X along CX to a regular neighborhood of the image of $C\Sigma$. Thus we can repeat this procedure with $C\Sigma$ in place of X. Since the dimension of $C\Sigma$ is at most $2k - n + 3$, we can make progress providing $2k - n + 3 < k$. This is true provided $n \ge 7$, and we are done. □

In fact, asking for an embedding of CX is superfluous to engulf X: it is sufficient to ask for a map which is an embedding *on each $X \times t$ slice*. For, this gives us a 1-parameter family of embeddings of X starting at the identity and ending inside a small ball; undoing an ambient isotopy produces the engulfing ball.

General position implies that we can make progress providing $2k-n+2<k$, improving on our naive count by one. Thus the argument extends to $n \geq 5$.

1.7. **Modifying Morse functions.** Let M be a smooth manifold of dimension n. A Morse function $f : M \to \mathbb{R}$ gives a way to obtain M from the empty manifold by attaching finitely many handles corresponding to the critical points as in § 1.4. It's convenient to move back and forth between the perspective of Morse functions and handlebodies. A Morse function (equivalently a smooth handle decomposition) admits the following three basic modifications:

(1) Birth–Death;
(2) Handle slide; and
(3) Whitney move.

We explain each of these modifications in turn.

1.7.1. *Birth–Death.* Morse functions form an open and dense subset of the space of smooth functions on a compact manifold, but the space of Morse functions is not connected. If f_t is a generic 1-parameter family of smooth functions f_t for which f_0, f_1 are Morse, there are finitely many intermediate values $0<s<1$ for which the Hessian of f_s is degenerate at some critical point p. Near p, the Hessian Hf_s has 0 as an eigenvalue of multiplicity 1, and there are smooth local coordinates x_i and an ϵ so that for $t \in (-\epsilon, \epsilon)$ we have

$$f_{s+t}(x) = x_1^2 + \cdots x_{n-q}^2 - x_{n-q+1}^2 - \cdots - x_{n-1}^2 + x_n^3 - tx_n$$

Let's look at what happens as t switches from negative to positive.

We compute $\partial_n f = 3x_n^2 - t$. When $t<0$ this is strictly positive, so f has no critical points near $x=0$. When $t=0$ the function gets a critical point at $x=0$. The Hessian Hf is degenerate, with ∂_n as a zero eigenvector. When $t>0$ this degenerate critical point splits into a pair of nondegenerate critical points $p^{\pm}$ at $x_n = \pm\sqrt{t/3}$. The index at p^- is q and the index at p^+ is $(q-1)$. For $t>0$ there is an ascending flowline from p^+ to p^- tangent to ∂_n; in an intermediate level set the descending and ascending manifolds of the critical points intersect transversely in exactly one point. When t passes through 0 in the positive direction a pair of critical points are *born*, and when it passes in the negative direction a pair of critical points *die*.

Near any point we may modify a Morse function so as to give birth to a pair $p^{\pm}$ of critical points of any adjacent indices $(q-1)$, q. Conversely, when $p^{\pm}$ are a pair of critical points with adjacent indices, and their descending and ascending manifolds intersect transversely in an intermediate level in a single point, the Morse function can be modified in a neighborhood of the flowline joining them so as to kill the pair. At the level of handlebodies, when the attaching sphere of the core of a q-handle intersects the attaching sphere of the co-core of a $(q-1)$-handle transversely in one point, we may remove the pair by a diffeomorphism. One also says that a pair of handles *geometrically cancel*.

Birth–death is the only modification we consider that changes the number of critical points.

1.7.2. *Handle slide.* When we build up a smooth manifold M as a handlebody by attaching handles one by one, the diffeomorphism type of M only depends on the isotopy classes of the attaching maps of the framed boundary spheres. If we attach h' to N and then h, we may replace h by a new h'' whose attaching map is isotopic in $\partial(N \cup h')$ but not necessarily in ∂N. A handle slide is a modification of this form.

Suppose h and h' are a pair of $(i+1)$ handles attached to an intermediate level $\partial M_{i+1/2}$ along framed neighborhoods of i-spheres S, S'. Note that these framed neighborhoods are disjoint in $\partial M_{i+1/2}$. Now, let α be an embedded arc in $\partial M_{i+1/2}$ from S to S'. We can extrude a finger from S and slide it along the arc α, all the way to S'. Then we can slide S over the core of the attaching handle h'; this replaces S with a *new* framed i-sphere S'' corresponding to a new $(i+1)$-handle h'' such that S'' is contained in the boundary of a tubular neighborhood of $S \cup \alpha \cup S'$. The handlebody obtained by attaching handles along S' and S'' is diffeomorphic to the original handlebody; to see this, just 'slide' h over h' to produce h''; this sliding can evidently be accomplished by isotopy.

This modification doesn't change the number of critical points, but it changes the collection of attaching spheres.

1.7.3. *Whitney move.* Let A^a and B^b be smooth submanifolds of M^n of complementary dimension; i.e. so that $a + b = n$. Let's assume A and B are in general position, so $A \cap B$ is a finite collection of points. If A, B, M are oriented, each intersection point has a sign. Let's suppose there's a pair of intersection points p, q of opposite sign. Let α and β be embedded arcs in A and B respectively running between p and q. The union $\alpha \cup \beta$ is a loop γ.

A *Whitney disk* is a smooth embedded disk D with $\partial D = \gamma$, interior disjoint from both spheres, together with a trivialization of its normal bundle ν and a splitting of this trivialization $\nu = E_1 \oplus E_2$ so that $E_1|_\alpha$ is some trivialization of the normal bundle to α in A, and $E_2|_\beta$ is some trivialization of the normal bundle to β in B. This trivializations is called a *framing* and is part of the data of a Whitney disk.

If there's a Whitney disk, it turns out we can push A over D by an isotopy to eliminate the two points of intersection. To do this, cut out the unit normal disk bundle of α from A and glue in the unit sphere bundle of E_1 to produce a new A'.

Actually when $n \ge 5$, one can always find a framing whenever there's an embedded D. That's because D is contractible, so its normal bundle is already trivial. Restricting this trivialization to α and β gives the framing over ∂D, which can be thought of (relative to a given trivialization of ν) as a loop in the Stiefel manifold $V_{a-1}(\mathbb{R}^{n-2})$ of oriented framed $\mathbb{R}^{a-1}$s in $\mathbb{R}^{n-2}$. This is an iterated sphere bundle, where the fibers range from spheres of dimension $n-3$ to $n-a-1$. Since $n \ge 5$ we can choose $a \le b$ so that $n - a \ge 3$ and $n - a - 1 \ge 2$. Thus the Stiefel manifold is simply-connected, and the framing can always be found.

When $n = 4$ things are more complicated. For instance, if $a = b = 2$ the Stiefel manifold $V_1(\mathbb{R}^2) = S^1$ is not simply-connected. We'll return to this in § 3.3.

1.8. **The h-Cobordism theorem.** The aim of this section is to prove the *h-Cobordism Theorem*, due to Smale. This theorem says that if M and N are smooth simply-connected closed n-manifolds that arise as the oriented boundaries of a smooth compact $n+1$ manifold

W, and if the inclusions of M and N into W are homotopy equivalences, then W is diffeomorphic to a product. We shall give Smale's proof, using Morse theory. We begin with some Morse function $f : W \to \mathbb{R}$ for which M and N are regular values, and then modify f so as to eliminate the critical points. Eventually we obtain a new Morse function g with no critical points at all. The gradient flow of g exhibits W as a product.

Theorem 1.4 (h-Cobordism Theorem)**.** *Let $n \geq 5$. Suppose W is a compact smooth simply-connected $(n+1)$-manifold with oriented boundary $\partial W = M \cup -N$, and suppose that the inclusions of M and N into W are homotopy equivalences. Then W is diffeomorphic to a product. Consequently M is diffeomorphic to N.*

Proof. Evidently it suffices to consider the case that M, N and W are connected, since otherwise we can work component by component.

Let f be a self-indexing Morse function on W with $f^{-1}(-1/2) = M$ and $f^{-1}(n+3/2) = N$. Each critical point of index i corresponds geometrically to an i-handle in a handle decomposition of W. Each $W_{i+1/2}$ is obtained from $W_{i-1/2}$ by attaching i-handles; the cores of these i-handles are the descending manifolds, and the co-cores are the ascending manifolds. Their boundaries are (attaching) spheres of dimension $i-1$ and $n-i$ in $\partial W_{i-1/2}$ and $\partial W_{i+1/2}$.

First we deal with $i = 0$. Attaching a 0-handle produces a new component. Only a 1-handle can join up different components, since S^0 is the only disconnected sphere. Since W is connected, for every 0 handle, some 1-handle must join it to another component. But this 0–1 pair geometrically cancel, so we can eliminate them. At the end of this procedure there are no 0-handles.

Now we deal with $i = 1$. Let h be a 1-handle. We'll show that we can create a 2–3 pair so that the 2 handle cancels h; in this way we can trade 1-handles for 3-handles.

How can we arrange this? The ascending sphere of h has codimension 1 in $\partial W_{1+1/2}$. We claim that there is a circle S in $\partial W_{1+1/2}$ transverse to this ascending sphere in exactly one point, and disjoint from the ascending sphere of every other 1-handle.

Actually, once we have cancelled all but one 0-handle, there is only one way to attach the remaining 1-handles to its boundary, so $\partial W_{1+1/2}$ is always a connect sum of $S^{n-2} \times S^1$s, and the ascending spheres of the 1-handles are $S^{n-2} \times$ point factors. In this picture, the existence of the transverse circle S is clear.

Now, if $\pi_1(W)$ is trivial, so is $\partial W_{2+1/2}$, since handles of index 3 and greater don't change the fundamental group. The descending sphere of a 2-handle is a circle, and in dimension 3 and greater, two circles can be isotoped to be disjoint. So we can push S up along gradient flowlines while avoiding all descending spheres of 2 handles by general position, so that it comes to a new circle S' in $\partial W_{2+1/2}$. Now, introduce a cancelling pair of handles of index 2 and 3, where the attaching circle of the 2-handle is in $\partial W_{2+1/2}$. Since this manifold is simply-connected, we can slide this attaching circle around until it is equal to S'. Then by construction, the descending sphere of this new 2-handle geometrically intersects the ascending sphere of the 1-handle transversely in a single point in $\partial W_{1+1/2}$ and we may cancel the 1–2 handle pair, thereby 'trading' a 1-handle for a 3-handle.

Define relative chain groups C_i freely generated by the handles of index i, and boundary maps $\partial C_{i+1} \to C_i$ whose matrix entries are given by the algebraic intersection number of the boundary of the core of an $(i+1)$-handle with the boundary of the co-core of an i-handle. The homology of this chain complex computes the relative cellular homology of W, which is therefore trivial in every dimension.

Let i be the least dimension in which there are handles. It follows that the boundary map $\partial C_{i+1} \to C_i$ is surjective. In the given basis the boundary map is expressed as a matrix. Handle slides of i or $i+1$ handles change the bases for these chain groups in such a way as to perform row and column operations on this matrix. Since the map is surjective, we can perform such a sequence of operations until there is a pair h_{i+1}, h_i with algebraic intersection ± 1. The co-core of h_i bounds an S^{n-i} in $\partial W_{i+1/2}$, and the core of h_{i+1} bounds an S^i. Call these spheres A and B. If the geometric intersection number of A and B is 1, we can cancel h_i and h_{i+1}. Otherwise there are a pair of intersection points of opposite sign.

Join these points by arcs α and β in the two spheres, and let $\gamma = \alpha \cup \beta$. Now, $\pi_1(\partial W_{i+1/2}) = 1$ if $i > 1$. We assume for the moment that $i > 2$ and $n - i > 2$. In this case, removing $A \cup B$ doesn't change the fundamental group so we can find an immersed disk in the complement bounding γ. If $n \geq 5$ then a disk in general position is embedded, and the framing problem can be solved. Thus we can find a Whitney disk and eliminate a pair of intersection points. After all but one intersection point is eliminated, we can cancel h_i and h_{i+1}.

If $i = 2$ this argument is a little more delicate. Let S_R be the boundary of the co-core of h_2 in $\partial W_{2+1/2}$ and let S_L be the boundary of the core of h_2 in $\partial W_{1+1/2}$. Let N be a neighborhood of S_L in $\partial W_{1+1/2}$. For the moment by pushing other handles up, we assume h_2 is the single 2-handle beneath level $2 + 1/2$. We also assume that we have traded all the 1-handles for 3-handles, so that $\partial W_{1+1/2}$ is simply-connected. Then $\partial W_{1+1/2} - S_L$ is diffeomorphic by the gradient flow to $\partial W_{2+1/2} - S_R$. But $\pi_1(\partial W_{1+1/2} - S_L) = \pi_1(\partial W_{1+1/2}) = 1$ because S_L is a circle, and $n > 3$. Now we can repeat the argument above for $i > 2$.

Thus by induction we can eliminate all i handles of W of dimension $\leq n - 3$. Likewise, by replacing f by $-f$ we can also eliminate all i handles of dimension ≥ 4 (remember that W has dimension $n+1$). If $n \geq 6$ we can therefore eliminate all handles and exhibit W as a product. If $n = 5$ this method eliminates all handles of W except possibly in dimension 3; but if there are no handles of any other dimension, the handles of dimension 3 freely generate the 3-dimensional relative homology of W, which by hypothesis must vanish. Thus, a posteriori, we conclude there are no 3 handles, and the theorem is proved. □

Here is the application to the Poincaré Conjecture. Suppose M^n is a smooth homotopy sphere. Remove two round balls from M^n and observe that the remainder forms an h-Cobordism between the two boundary spheres. Thus, providing $n \geq 6$, we can conclude that this remainder is a (smooth) product, and M is homeomorphic to S^n.

The case $n = 5$ requires a modification of this strategy. One shows that an S^5 and a smooth homotopy S^5 are smoothly cobordant. Then one surgers the cobordism until it is an h-Cobordism.

2. Decompositions

This section is something of a detour. Smooth topology will play a big role in the eventual proof of the 4-dimensional Poincaré Conjecture. We will work with handlebodies, and make crucial use of smooth tools like transversality. The main problem is the Whitney trick. If we try to prove an h-Cobordism Theorem for 4-manifolds, we will need to deal with pairs of smooth embedded 2-spheres which intersect algebraically once, but geometrically any number of times. If we try to use a Whitney disk to eliminate some pairs of intersections, we'll run into trouble, because a disk in general position in a 4-manifold is typically not embedded, but only immersed. Pushing a sphere over an immersed Whitney disk removes some intersections, but possibly produces many more!

OK, maybe you can try to look for an embedded Whitney disk, but it turns out you might be looking for a long time. Smooth Whitney disks can't always be found. What can be found are a sequence of more and more complicated objects — kinks, gropes, towers and so on. The situation gets wilder and wilder with no apparent end in sight.

Decomposition theory is a tool for taming this wildness. It's a remarkably powerful tool. In § 2.3 we'll see how it reduces the proof of the Schoenflies Conjecture to a few paragraphs. But this power comes at a cost, and the cost is that we no longer get to see explicitly what's going on. The magician will take care of your rabbit problem — but she won't reveal her tricks.

2.1. **Cellular subsets.** Let $f : D \to E$ be a map between two disks. The *fibers* of f are the point preimages. A fiber is *nontrivial* if its cardinality is greater than 1.

A subset $F \subset D$ is said to be *cellular* if there are a nested sequence of closed disks $D_i \subset D$, each D_{i+1} contained in the interior of D_i, such that $\cap_i D_i = F$.

Lemma 2.1. *Let D and E be disks, and let $f : D \to E$ be a continuous map with a single nontrivial fiber $F = f^{-1}(y)$ for some $y \in E$. Then F is cellular.*

Proof. Let E_i be a nested sequence of small round neighborhoods of y contained in $f(D)$, with intersection y. Let $r_i : E \to E_i$ be radial homeomorphisms shrinking points inwards, but equal to the identity in a neighborhood of y. Define maps $g_i : D \to D$ by setting $g_i(x) = x$ for $x \in F$, and otherwise define $g_i(x)$ by the formula $fg_i(x) = r_i f(x)$; i.e. $g_i = f^{-1} r_i f$. Because each r_i is the identity near y, the same is true of g_i on some open neighborhood of F; thus g_i is continuous. The map g_i is further injective, and therefore it is a homeomorphism onto its image. So $g_i(D) = D_i$ is a nested sequence of closed disks and $\cap_i D_i = F$. □

Note that invariance of domain implies that D and E have the same dimension, since f is open away from F.

2.2. **Shrinkability.**

Definition 2.2. A *decomposition* of a compact metric space X is a partition into nonempty closed sets called the *elements* of the decomposition. A decomposition is *upper semicontinuous* (hereafter USC) if for each element A and each positive ϵ there is δ, so that if an element B intersects the ϵ-neighborhood of A, it is contained in the δ-neighborhood of A. An element with more than one point is *nondegenerate*, and the set of nondegenerate elements is the *nondegeneracy set*.

For a compact metric space X, upper semi-continuity is equivalent to the statement that any Hausdorff limit of a sequence of decomposition elements is *contained in* (though not necessarily *equal to*) some decomposition element. If $\mathcal{D}$ is a decomposition of X, we write $X/\mathcal{D}$ for the quotient space in which each element of $\mathcal{D}$ is crushed to a point. Upper semi-continuity is equivalent to the condition that $X/\mathcal{D}$ is Hausdorff, in which case it is metrizable. For the sake of brevity, we often assume that we have chosen a metric on $X/\mathcal{D}$; none of our conclusions will depend on the choice of metric.

The most important property of a decomposition we consider is called *shrinkability*. In the sequel we let $\mathcal{D}$ be an USC decomposition of a compact metric space X, and let $q : X \to X/\mathcal{D}$ be the quotient map. Assume we have chosen metrics on X and $X/\mathcal{D}$.

Definition 2.3. A USC decomposition $\mathcal{D}$ of X is *shrinkable* if for every $\epsilon > 0$, there is a surjective homeomorphism $h : X \to X$ satisfying

(1) qh is close to q: for any $x \in X$, the distance from $q(x)$ to $qh(x)$ is less than ϵ; and
(2) decomposition elements are small: for any decomposition element $Y \subset X$, the diameter of $h(Y)$ is less than ϵ.

A quotient map $q : X \to X/\mathcal{D}$ is *approximable by homeomorphisms* (hereafter ABH) if there are homeomorphisms $q_i : X \to X/\mathcal{D}$ converging to q in the uniform topology.

Lemma 2.4. *A USC decomposition $\mathcal{D}$ of X is shrinkable if and only if the quotient map $q : X \to X/\mathcal{D}$ is approximable by homeomorphisms.*

Proof. Suppose $q : X \to X/\mathcal{D}$ is ABH $q_i : X \to X/\mathcal{D}$. For any ϵ there is i so that the distance from $q(x)$ to $q_i(x)$ is at most $\epsilon/2$. Now, since q_i^{-1} is a homeomorphism between compact metric spaces, there is a $\delta > 0$ such that if $A \subset X/\mathcal{D}$ has diameter at most δ, the diameter of $q_i^{-1}(A)$ is at most ϵ. Now, without loss of generality, let's suppose $\delta < \epsilon$, and let j be such that the distance from $q(x)$ to $q_j(x)$ is at most $\delta/2$.

Then for any decomposition element Y, the image $q_j(Y)$ has diameter at most δ, so $q_i^{-1}q_j(Y)$ has diameter at most $\epsilon/2$. Furthermore, for any $x \in X$ the triangle inequality gives

$$d(q(x), qq_i^{-1}q_j(x)) \le d(q(x), q_j(x)) + d(q_i(q_i^{-1}q_j(x)), q(q_i^{-1}q_j(x))) \le \epsilon$$

This proves one direction.

Conversely, suppose $\mathcal{D}$ is shrinkable. We give an argument due to Edwards showing that q is ABH. Consider the space of continuous maps from X to $X/\mathcal{D}$ with the uniform topology (this is a Baire space), and let H denote the closure of the set of maps of the form qh where $h : X \to X$ is a homeomorphism. Shrinkability implies (by conjugation) that the subset H_ϵ of H whose fibers all have diameter $< \epsilon$ is open and dense. Thus the intersection H_0 of all H_ϵ is dense. But any element of H_0 is a homeomorphism, so we are done. ∎

Now let $\mathcal{D}$ be a decomposition of a disk D with a single nontrivial element X, and suppose X is cellular. Let D_i be nested disks with $D_i \to X$. For any ϵ, there is some i so that $q(D_i)$ has diameter less than $\epsilon/2$. It follows that for any homeomorphism $h_i : D \to D$ supported in D_i the map qh_i is ϵ-close to q. But since D_i is a disk, there is certainly some $h_i : D \to D$ for which the diameter of $h_i(X)$ is less than ϵ. It follows that $\mathcal{D}$ is shrinkable, and therefore $D \to D/\mathcal{D}$ is a homeomorphism.

Putting this together we deduce:

Corollary 2.5. *If $f : D \to E$ is a map between disks with a single nontrivial fiber, then $f : D \to f(D)$ is ABH.*

It is straightforward to extend this Corollary to the case when f has finitely many nontrivial fibers. Let's suppose $f : D \to E$ has two nontrivial fibers F_1, F_2 mapping to y_1, y_2.

There is an open neighborhood U of y_1 avoiding y_2, and a radial map $r : E \to U$ fixing y_1. Define a map $g : D \to D$ as before by $f^{-1}rf$ on $D - F_1$, and the identity on F_1. The map g is not a homeomorphism, since it crushes F_2 to the single point $f^{-1}rf(F_2)$. But it is a homeomorphism everywhere else. It follows by the Lemma that F_2 is cellular, and evidently so is F_1. Thus the decomposition of D by fibers of f is shrinkable, and $f : D \to f(D)$ is ABH. The general case of finitely many nontrivial fibers follows by induction.

The next Lemma can be used to show that an USC decomposition is shrinkable without actually exhibiting a shrinking sequence.

Lemma 2.6 (Big Shrinking)**.** *Suppose we have a compact metric space X with an USC decomposition $\mathcal{D}$. Let $q : X \to X/\mathcal{D}$ be the quotient map, and suppose for any $\epsilon > 0$ there is an open neighborhood U of the union of nontrivial elements, so that for each component V of U the image $q(V)$ has diameter less than ϵ in $X/\mathcal{D}$.*

Now let U_i be a sequence of such neighborhoods for $\epsilon = 1/i$. Suppose we can find a sequence of homeomorphisms $h_i : X \to X$ supported in U_i and fixing it componentwise, so that for any decomposition element Y, the product $g_n := h_1 h_2 \cdots h_n(Y)$ has diameter at most $1/n$ in X. Then $\mathcal{D}$ is shrinkable.

Proof. Since X is metric, q is uniformly continuous, and therefore the sequence g_n eventually satisfies the second bullet of shrinkability, though they will typically not satisfy the first bullet: the g_n will typically be uniformly far from q.

To remedy this, for any $n \le m$, let $f_{n,m} := g_n^{-1} g_m = h_{n+1} h_{n+2} \cdots h_m$. Now for any fixed n the sequence $f_{n,m}$ eventually satisfies the second bullet of shrinkability, since this is true of the sequene g_m, and this property of a sequence is preserved by precomposition with any fixed homeomorphism. On the other hand, for any n, any $f_{n,m}$ is supported in U_n, and therefore any sequence of $f_{n,m}$ with $n \to \infty$ eventually satisfies the first bullet. A diagonal subsequence therefore satisfies both properties, and shows that $\mathcal{D}$ is shrinkable. □

Certain kinds of decomposition always satisfy the first hypothesis of the lemma.

Definition 2.7. An USC decomposition $\mathcal{D}$ of a compact metric space X is *null* if for any $\epsilon > 0$ there are only finitely many elements of $\mathcal{D}$ with diameter at least ϵ.

Note that a null decomposition has at most countably many nontrivial elements. If $\mathcal{D}$ is null, we can always find a family of open neighborhoods U_i as in the hypothesis of the lemma. We shall not use this fact in the sequel.

2.3. **The Schoenflies Conjecture.** In this subsection we prove the Schoenflies Conjecture, following Brown. To first approximation, the Schoenflies Conjecture says that a codimension one sphere in S^n (or D^n) bounds a ball. We've already used it in the proof of the Engulfing Theorem 1.3. In that context the codimension one sphere and the ball were smooth. As is well-known, and can be proved using the exponential map and an auxiliary

Riemannian metric, smooth submanifolds have *tubular neighborhoods*; i.e. the embedding extends (smoothly) to an embedding of some smooth vector bundle. A topological submanifold might not have a tubular neighborhood; if it does we say it is *collared*. In this context, a collared S^{n-1} in S^n has a neighborhood homeomorphic to a product $S^{n-1} \times [-1, 1]$.

Without the hypothesis that the sphere is collared, the Schoenflies Conjecture is false, as we shall see in § 2.4.

Theorem 2.8 (Schoenflies Conjecture)**.** *Let S be a collared S^{n-1} in S^n. Then S bounds a closed ball on each side; equivalently there is a homeomorphism from S^n to itself taking S to the round equator.*

Proof. We would like to apply Corollary 2.5. Let N denote the collar of S, and fix a homeomorphism $\phi : S^{n-1} \times [-1, 1] \to N$ taking $S^{n-1} \times 0$ to S. Let D be a round open disk in N and let E be a round open disk in $S^{n-1} \times [-1, 1]$ for which the closure of $\phi(E)$ is contained in the interior of D. The complement of D in S^n is a closed disk, and is contained in the complement of $\phi(E)$. Let $\psi : S^{n-1} \times [-1, 1] \to S^n$ be obtained by crushing each boundary sphere to a point. The map ϕ^{-1} is defined on $N - D$, but the composition $\psi\phi^{-1}$ extends to the entire disk $S^n - D$ by crushing each component of $S^n - N$ to a point. Since E is a round disk in $S^{n-1} \times [-1, 1]$ its image $\psi(E)$ is a round disk in S^n so $S^n - \psi(E)$ is a closed disk. Thus $\psi\phi^{-1} : S^n - D \to S^n - \psi(E)$ is a map between closed disks, and by construction it has exactly two nontrivial fibers. In particular, both components of S^n minus the interior of N are cellular.

Now, let A be one component of S^n minus the interior of N, and let M be the compact manifold $A \cup S^{n-1} \times [-1, 0]$ (say). Notice that M is one of the closed complementary regions to S, and we are trying to prove that M is a ball. But now we know that A is cellular, and therefore shrinkable in M. It follows that $M \to M/A$ is a homeomorphism. Since M/A is the cone on S^{n-1}, we deduce that M is homeomorphic to D^n (and similarly for the other complementary component).

The last claim of the theorem follows by coning. □

2.4. **Wild spheres and crumpled cubes.** It's important in the proof of the Schoenflies Conjecture that we only consider *collared* (also: tame) spheres. Alexander gave an example of a wild S^2 in S^3 — the *Alexander horned sphere* — which does not bound a D^3 on one side. We describe a particular horned sphere Σ in $\mathbb{R}^3$ that bounds a compact subset B whose interior is not simply-connected. The subset B is known as a *crumpled cube*.

It's convenient for later applications to let our construction depend on a parameter $0 < \lambda < 1/2$. It's fine to think of λ as fixed for the moment; Figure 2 indicates the picture when $\lambda \sim 0.49$.

Embed two solid cylinders H_0, H_1 inside a bigger solid cylinder C as shown in Figure 1, and let F denote the closed complement of C in S^3.

If we parameterize each H_i as $D^2 \times [0, 1]$ we let C_i denote the 'middle' piece $D^2 \times [\lambda, 1-\lambda]$ of each H_i, and let F_i denote $H_i - C_i$. There's an affine isomorphism $\phi_i : C \to C_i$ (in radial coordinates) for each of $i = 0, 1$, and we can use this to build new solid handles $H_{ij} \subset C_i$ for $j = 0, 1$ by the formula $H_{ij} = \phi_i(H_j)$. There is likewise a middle piece $C_{ij} \subset H_{ij}$, a complement $F_{ij} = H_{ij} - C_{ij}$, and an affine isomorphism $\phi_{ij} : C \to C_{ij}$. Inductively we can construct a C_σ, H_σ, F_σ and $\phi_\sigma : C \to C_\sigma$ for each finite binary string σ.

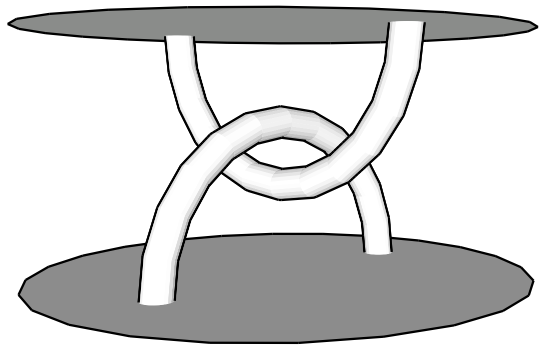


FIGURE 1. The handles H_0 and H_1 in C

Define E to be the *closure* of $\cup_\sigma F_\sigma$ where the union is taken over all binary strings σ. The interior of E is an open ball; it is the increasing union of the interiors of $\cup_{|\sigma|\le n} F_\sigma$, each of which is itself an open ball. The boundary of E is the wild sphere Σ. The complement of the interior of E is the crumpled cube B; the interior of B is a 3-manifold with infinitely generated π_1, though B itself is simply-connected.

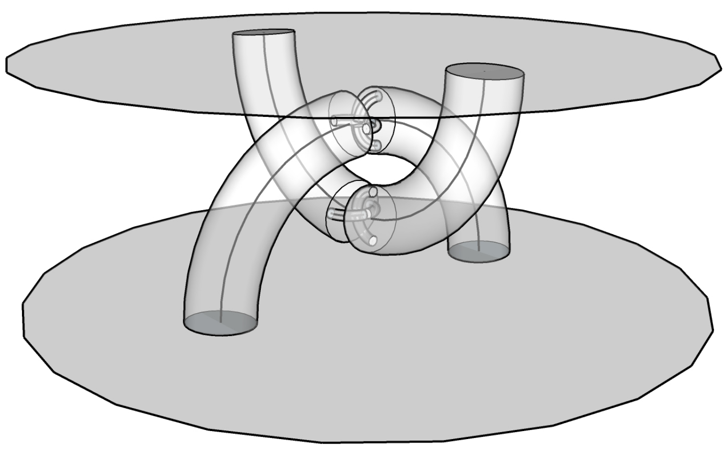


FIGURE 2. 2 stages in the construction of the wild sphere

The infinite intersection $\mathcal{C} = \cap_\sigma C_\sigma$ is a Cantor set, consisting precisely of the points at which the horned sphere Σ is not locally tame.

We can think of each F_σ as a 'finger' extruded by the exterior into the interior. It grows for a while, and then extrudes its own pair of smaller fingers $F_{\sigma 0}, F_{\sigma 1}$ which each extrude their own pair of smaller fingers and so on. The fingers at every stage are geometrically entangled with each other at some finite scale; in the limit the entanglement is topological in nature.

Now consider how this construction depends on the parameter λ. The smaller λ is, the stubbier the fingers F_σ are, and the slower their diameters shrink. Nevertheless, for any two $0 < \lambda' < \lambda$ the resulting spaces are obviously homeomorphic.

On the other hand, if we take λ to 0, we can still construct the sequence C_σ as above, with $C_\sigma = H_\sigma$ at each stage. Now the diameters of the components of C_σ stay bounded below, and for any infinite increasing sequence of binary strings $\sigma_1 \subset \sigma_2 \subset \dots$ the intersection $\cap_i C_{\sigma_i}$ is a tame arc from ∂C to one of the points in the Cantor set $\mathcal{C}$ from before.

The union of these tame arcs is an USC decomposition $\mathcal{D}$. To see what the quotient space is, shrink the elements of $\mathcal{D}$ to points by successively pushing in fingers F_σ with bigger and bigger $|\sigma|$. Evidently the quotient space $C/\mathcal{D}$ is homeomorphic to the crumpled cube B.

Since B and C are not homeomorphic, $\mathcal{D}$ is not shrinkable *in* C, although it *is* shrinkable in the ambient S^3. This is true even though the decomposition of C is null, and the nontrivial elements are cellular.

2.5. **Bing doubles and an exotic involution on S^3.** Now, although the crumpled cube B is not a manifold, it does have a 'boundary' S^2, namely the wild Alexander horned sphere. It makes sense therefore to construct the *double* of B, denoted DB, by gluing two copies of B together along their boundary spheres.

The double DB is obviously not a manifold — or is it?!? Wilder conjectured (!) and Bing proved (!!) that it is! In fact, DB is homeomorphic to S^3. There is an obvious involution of DB that interchanges the two sides. Thus there is an *exotic* involution on S^3, one whose fixed point set is a wild S^2.

We now give (a modification of) Bing's argument, proving this. The double of the solid cylinder C is evidently S^3. We push the annulus A of the cylinder C slightly into its interior; after doubling, A doubles to a torus T bounding a solid torus N. Each C_σ has boundary disks contained in the boundary of C, so when we double it becomes a solid torus N_σ with boundary torus T_σ. Notice that the core K_σ of every N_σ is an unknot in S^3, though it is knotted in $N_{\sigma'}$ where σ' is the prefix of σ obtained by removing the last letter.

The intersection of a sequence $\cap_i N_{\sigma_i}$ is a tame interval, obtained by doubling one of the components of $\mathcal{D}$. We define a new decomposition $\mathcal{D}'$ of S^3 whose components are the doubles of the components of $\mathcal{D}$. We shall show that $\mathcal{D}'$ is shrinkable; this will imply that S^3 is homeomorphic to $S^3/\mathcal{D}' = D(C/\mathcal{D}) = DB$.

Theorem 2.9 (Bing)**.** *The decomposition $\mathcal{D}'$ of S^3 is shrinkable.*

Proof. Let U_i denote the union of the open N_σ with $|\sigma| = i$. Then the components of U_i nest down to the components of $\mathcal{D}'$, and evidently satify the first property of the Big Shrinking Lemma 2.6. It remains to find a sequence of homeomorphisms h_i of S^3, each supported in U_i, whose compositions $g_i := h_1 h_2 \cdots h_i$ shrink all components of N_σ with $|\sigma| = i$ to diameter $\le \epsilon_i$, for some $\epsilon_i \to 0$.

Our sequence will have the property that the image of each (shrunken) N_σ will be a (very!) thin tubular neighborhood of a knot K_σ, so that the diameter of any decomposition element in the image of N_σ can be estimated from the diameter of K_σ.

Now, the knot K is essential in N, but every other K_σ is inessential in N, and therefore lifts to the universal cover of N, in which the preimage $\tilde{K}$ of K is a line. Since N can be as thin as we like, we imagine K_σ being laid out somehow along this line. Measure length along $\tilde{K}$ so that each fundamental domain has length 1.

Consider the (rooted) binary tree of all finite binary sequences, and partially order it by prefix of sequences. Each sequence σ has two *children* $\sigma 0$, $\sigma 1$ and is associated to a component N_σ of $U_{|\sigma|}$.

To each σ we will associate a number t and a cyclic word W in the letters L and R, and write $\sigma \to (t, W)$. The meaning of this is that K_σ consists of $|W|$ segments of length t aligned with $\tilde{K}$, where a segment labeled L goes 'left' and a segment labeled R goes 'right'. If we lay out K_0 and K_1 in the obvious way, we write

$$0 \to (2^{-1}, LR) \text{ and } 1 \to (2^{-1}, LR)$$

If $\sigma \to (t, W)$ then the length of K_σ is $t|W|$. The word W has as many Ls as Rs because K_σ 'closes up'. For each subword I of W let $\#L(I)$ and $\#R(I)$ denote the number of Ls and Rs in I respectively. Let $s(W)$ be the maximum of $|\#L(I) - \#R(I)| > 0$ over all I. Then the diameter of K_σ (i.e. the length of its projection to $\tilde{K}$) is $ts(W)$.

By induction we'll suppose that W never has 3 or more Ls or Rs in a row. We call an LL or an RR a *chunk*. The only chunkless words are $(LR)^k$ for some k. If W is chunkless, the diameter of K_σ is t, no matter how big $|W|$ is.

Now, we can always think of each segment of K_σ of length t as the concatenation of two segments of length $t/2$. In our notation, $\sigma \to (t, W)$ is equivalent to $\sigma \to (t/2, W')$ where W' is obtained from W by replacing each L with LL, and each R with RR. If W is chunkless, then W' consists entirely of chunks. We call this operation *subdivision*; we shall apply it only to chunkless words.

Now, K_σ has two children, $K_{\sigma 0}$ and K_{σ_1}. Here is how we lay the $K_{\sigma i}$ along K_σ. We split K_σ into two segments A, B. Then K_{σ_0} goes once along A and then back along A^{-1}, while K_{σ_1} does the same with B. This can be arranged by a homeomorphism supported in N_σ.

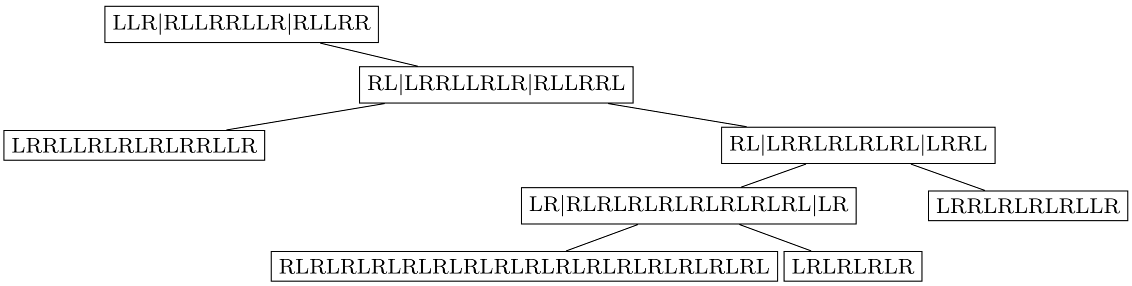
At the level of words, we split W into two subwords A, B. We denote by A' the word obtained from A by reversing the order of letters, and exchanging Ls for Rs and vice versa. With this notation, we have

$$\sigma 0 \to (t, AA'), \quad \sigma 1 \to (t, BB')$$

Notice that AA' has twice as many chunks as A, and similarly for BB'; this is because every chunk in AA' is entirely contained in A or in A'. Let's observe by induction that all our words have an even number of chunks.

We claim that every W arising as above is *either* chunkless (and should be subdivided) *or* has a pair of subwords A, B so that both AA' and BB' have fewer chunks than W. To see this, enumerate the chunks around W in order, and split along the middle of a pair of opposite chunks.

But now we are done: for any initial W, by repeatedly splitting, we obtain a finite binary subtree with root word W, all of whose leaves have chunkless words. Subdivide all the leaves, and continue. Eventually the diameters of all the leaves are as small as we like. But any decomposition element is contained in some leaf, and therefore successive application of homeomorphisms gives a big shrinking sequence. □



Part of an inefficient splitting sequence

The graph above shows part of an inefficient splitting sequence. Actually, a judicious choice of splitting ensures that for every σ with $2^n < |\sigma| \leq 2^{n+1}$ we have $|W_\sigma| = 2^n$ and $t_\sigma = 2^{-n}$. Each K_σ has length 1 and diameter 2^{1-n} unless $|\sigma|$ is a power of 2, in which case the diameter is 2^{-n}.

2.6. **Embedded gropes.** We have seen how the crumpled cube arises as an infinite intersection, by successively drilling out smaller and smaller fingers as they poke in. Now let's see how it arises as an infinite union.

Let's start again with C, and let γ be a meridianal loop in ∂C. γ bounds a meridional disk D in C dividing the top cap from the bottom. This disk intersects each of the handles H_0 and H_1 transversely in two meridional circles. Tubing either H_0 or H_1 changes D into a once-punctured torus τ. Note that the standard meridian and longitude of τ are meridians of H_0 and H_1 respectively, and a thickened neighborhood of τ is equal to $C - (H_0 \cup H_1)$. We refer to the meridian and longitude of τ as γ_0 and γ_1.

Now, cut C_0 and C_1 out of H_0 and H_1. This has the effect of gluing 2-handles (i.e disks D_0, D_1 to the meridian and longitude of τ, one on either side. But if we now drill H_{i0}, H_{i1} out of each C_i, the disk D_i intersects each of these handles transversely in two meridional circles, and tubing produces once-punctured tori τ_0, τ_1 capping off γ_0, γ_1. A thickened neighborhood of the various τ_i fills out the complement of the second stage construction of the crumpled cube.

Continue inductively: for each once-punctured torus τ_σ with meridian or longitude $\gamma_{\sigma i}$, there is a once-punctured torus $\tau_{\sigma i}$ that caps it off inside C_σ by tubing a disk around the meridian of $H_{\sigma i}$. For each i the union

$$\Gamma_i := \cup_{|\sigma| \leq i} \tau_\sigma$$

is called an *i-stage grope*, and the infinite union is an infinite stage grope. We can think of the infinite stage grope as a *spine* for the crumpled cube, since a tapering union of neighborhoods of the τ_σ is precisely equal to the interior of B.

Algebraically, the fundamental group of each i-stage grope is free, and its abelianization (i.e. H_1) is freely generated by the γ_σ. Including each grope into the next stage precisely kills all the H_1, at the cost of introducing twice as much new H_1. Thus the union (i.e. the interior of the crumpled cube) has trivial H_1, although its fundamental group is infinitely generated.

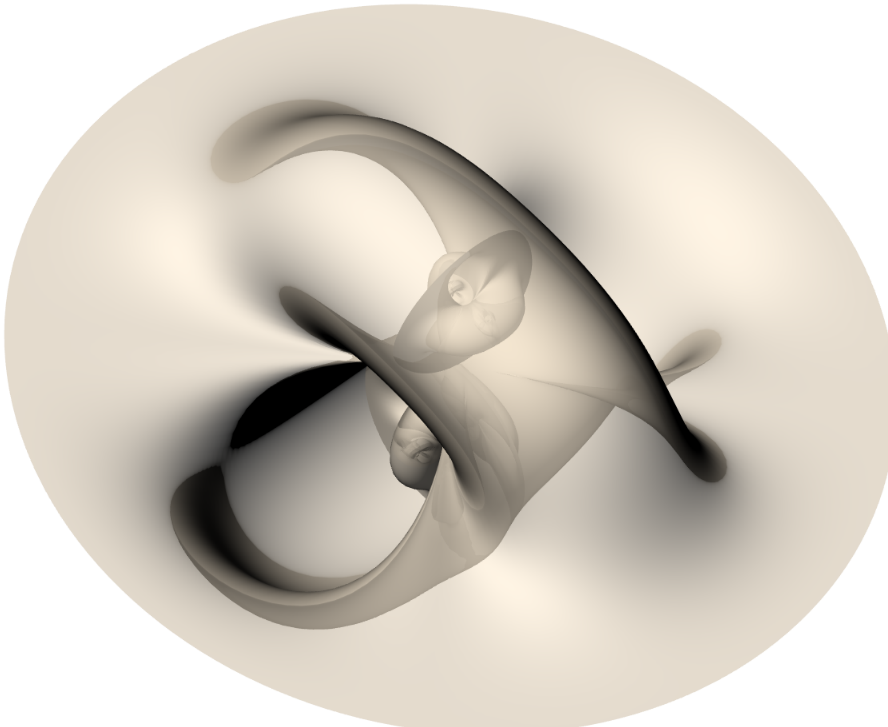


FIGURE 3. The grope is the spine of the crumpled cube.

Actually, we already know H_1 of the crumpled cube interior is zero; this follows from Alexander duality. For a reasonable subset X of S^n the homology of the complement of X depends only on the homeomorphism type of X, and not on the way X sits in S^n.

2.7. S^3/Wh **is a manifold factor.** Suppose K is a knot in S^3 and N is a regular neighborhood. The *Bing double* of K is the link consisting of a pair of unknots clasped nontrivially in N, i.e. as K_0 and K_1 clasp in N in the proof of Theorem 2.9. If K is an unknot in S^3, then the union of $K_0 \cup K_1$ with a meridian of N is the Borromean rings.

Finite Bing doubling is a topological notion; the result of n-fold Bing doubling is a 2^n-component link in N whose isotopy class is well-defined. However the result of *infinite* Bing doubling is a *geometric* notion — in fact that is the whole point of Theorem 2.9: the result depends on the particular choice of embeddings of each successive $N_{\sigma 0}, N_{\sigma 1}$ in N_σ. But however it is done, the *complement* is an open crumpled cube.

Now, if K is a knot in S^3 and N is a regular neighborhood, the *Whitehead double* is a *single* knot K_1 in N clasping itself. If K is an unknot in S^3, the union of K_1 with a meridian of N is the Whitehead link. Under iterated Bing doubling, the fundamental group of the open complement gets more and more complicated; but under Whitehead doubling (beginning with the unknot!) the complement at every stage is a solid torus.

Let $Wh = \cap_i N(K_i)$ denote the infinite intersection of all the $N(K_i)$ obtained by successive Whitehead doubling of the unknot. *Wh* is called the *Whitehead continuum*. Since each (closed) $N(K_i)$ is compact and connected, so is *Wh* (however it is not path connected).

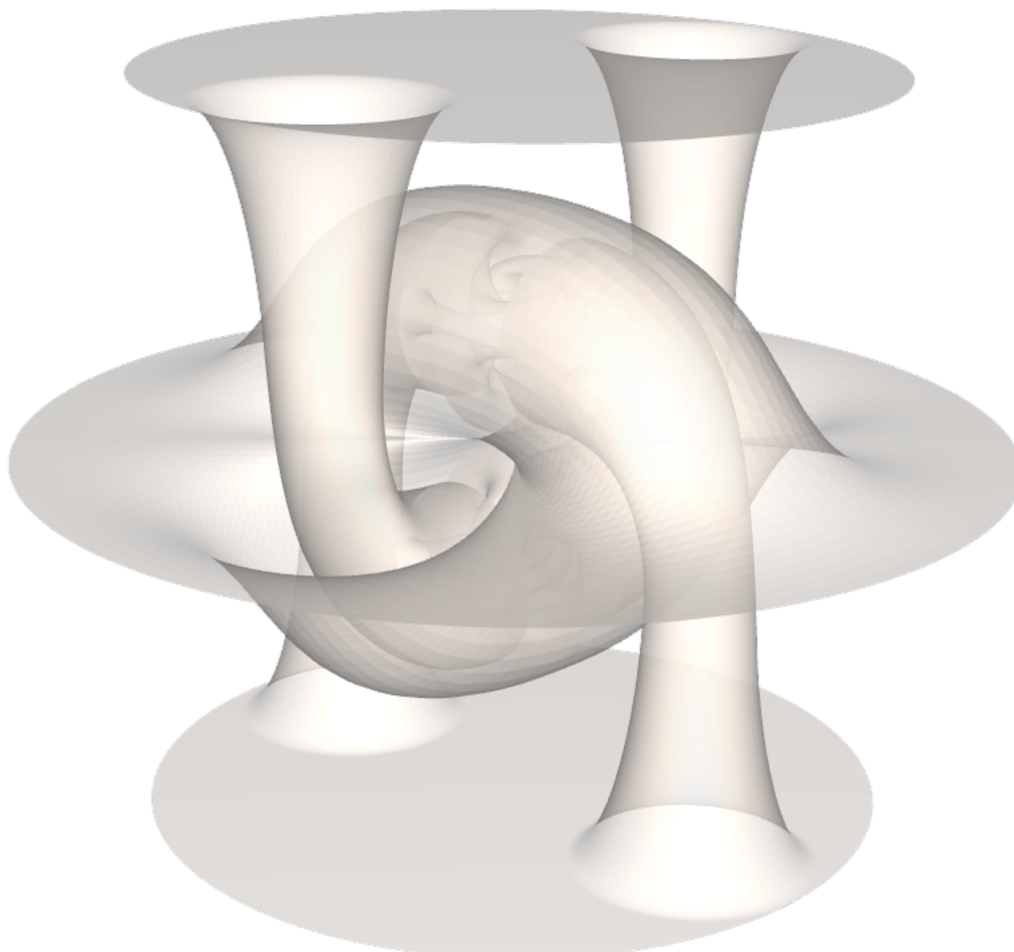


FIGURE 4. The grope locked in complement with the horned sphere

The complement $S^3 - Wh$ is called the *Whitehead manifold*, and unlike the interior of the crumpled cube, it *is* simply-connected, and even contractible. However it is 'wild at infinity': its end is not homeomorphic to a product of a surface with an interval, and therefore it is not homeomorphic to an open D^3.

The one-point compactification of the Whitehead manifold is evidently equal to the quotient S^3/Wh, and is not homeomorphic to S^3 (in fact, it is not even a manifold) since the complement of the compactifying point has a wild end. Therefore the following is quite surprising:

Theorem 2.10 (Manifold factor). *The product $S^3/Wh \times \mathbb{R}$ is homeomorphic to $S^3 \times \mathbb{R}$.*

Proof. We think of S^3/Wh as the quotient by a decomposition whose only nontrivial element is the Whitehead continuum. Since the quotient is not S^3, the nontrivial element Wh is not cellular.

Let $\mathcal{D}$ denote the decomposition of $S^3 \times \mathbb{R}$ with nontrivial elements consisting precisely of the elements $Wh \times t$ for each $t \in \mathbb{R}$. We shall show that $\mathcal{D}$ is shrinkable in $S^3 \times \mathbb{R}$, proving the result. Since our results on shrinking only hold for compact metric spaces, technically we should first compactify $S^3 \times \mathbb{R}$ by adding two points at infinity to form S^4, and extend the decomposition $\mathcal{D}$ trivially over these two points. It is evidently sufficient to show that $\mathcal{D}$ is shrinkable in S^4. This property will hold for the shrinking sequence we now construct.

Wh is a nested sequence $\cap_i N_i$ of closed solid tori in S^3. Let's think about why we can't shrink N_i in N_{i-1} keeping $S^3 - N_{i-1}$ fixed. The reason is topology (winding number): because the core of N_i is a nontrivial knot in N_{i-1}, any isotopy of this core in N_{i-1} must intersect every meridian disk of N_{i-1}, so its diameter is uniformly bounded below. Informally, we can't undo the 'clasps' of N_i in 3-dimensions.

But we can undo these clasps in 4 dimensions. First, each N_1 slice can be unknotted by a tiny perturbation in $N_0 \times \mathbb{R}$. To distinguish the $\mathbb{R}$ factor, and for the sake of brevity, we refer to it as the 'time' coordinate (this is purely a notational convenience). In this language, we unclasp N_1 from itself by nudging one clasp very slightly forward into the future, and the other very slightly back into the past. After the nudge, N_1 will not clasp itself, but it will clasp a 'future' N_1 on one side, and a 'past' N_1 on the other. Instead of N_1 clasping itself in a circle, we get a chain of successive N_1s, each clasping the next, in a slowly ascending *spiral*. Let's let $\epsilon/4$ be the size of the perturbations of each clasp in the time direction, so that the projection of each N_1 to the time coordinate after it's been nudged has total length $\epsilon/2$.

Nudging adjusts points in $N_1 \times \mathbb{R}$ by sliding each point $\times \mathbb{R}$ slightly backward or forward in time. Nudging extends to a self-homeomorphism ν of $N_0 \times \mathbb{R}$, fixed on the boundary.

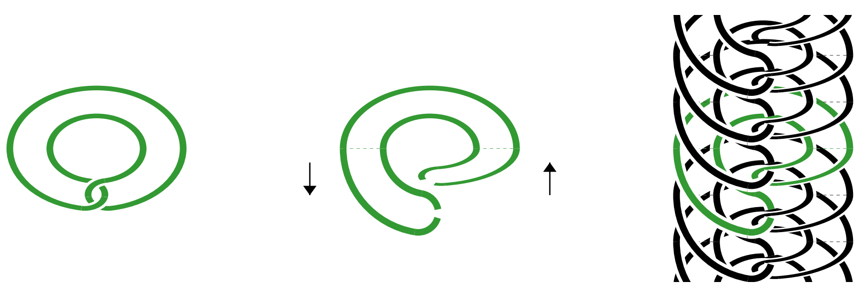


FIGURE 5. Folding the clasps of each N_1 back and forth in time nudges the union of all N_1s into a collection of spirals

By the way, there's not just one spiral, there's a circle's worth of them, filling the whole of $N_1 \times \mathbb{R}$. Two slices $\nu(N_1 \times t)$, $\nu(N_1 \times s)$ are in the same spiral if and only if $t - s$ is an integer multiple of $\epsilon/2$.

After nudging, the next move will straighten out this and every other spiral so that its projection to the S^3 factor is small (let's say for concreteness has diameter $< \epsilon/2$) *without* affecting the projection to the $\mathbb{R}$ factor.

The cylinder $K_0 \times \mathbb{R} \in S^3 \times \mathbb{R}$ has polar co-ordinates (θ, t) where $\theta \in \mathbb{R}/\mathbb{Z}$. Extend these polar co-ordinates to a small tubular neighborhood of $K_0 \times \mathbb{R}$ containing $N_1 \times \mathbb{R}$, with closure contained in the interior of $N_0 \times \mathbb{R}$.

We can 'untwist' every spiral simultaneously by the map

$$(\theta, t) \to (\theta - 2t/\epsilon, t)$$

on our small tubular neighborhood. Twisting extends to a self-homeomorphism τ of $N_0 \times \mathbb{R}$, once again fixed on the boundary.

In summary, first we nudge, then we twist. After doing this, every $\tau\nu(N_1)$ slice projects to subsets of diameter at most $\epsilon/2$ in *both* the $\mathbb{R}$ and the S^3 directions. So $\tau\nu(N_1)$ has diameter at most ϵ.

In other words, $h_1 := \tau\nu$ simultaneously *shrinks* all the N_1 slices in $S^3 \times \mathbb{R}$ as small as we like, while keeping $(S^3 - N_0) \times \mathbb{R}$ fixed pointwise.

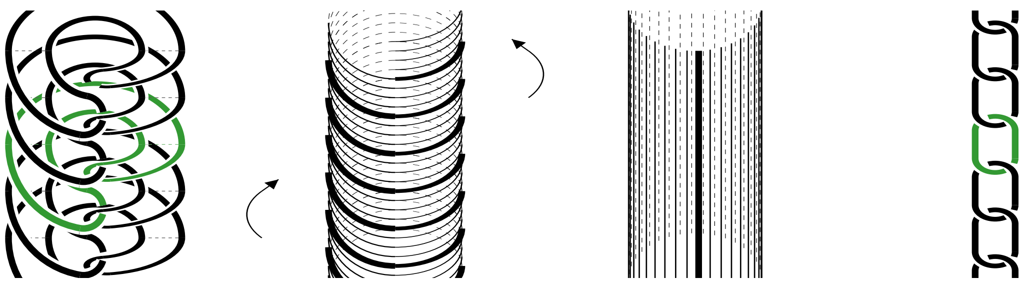


FIGURE 6. Screw top and bottom in opposite directions like you're taking the lid off a pickle jar

Take a sequence $\epsilon_i \to 0$, and repeat this operation for each $i > 1$ in place of 1 with ϵ_i in place of ϵ. We get a sequence of self-homeomorphisms $h_i : S^3 \times \mathbb{R} \to S^3 \times \mathbb{R}$, each supported in $N_{i-1} \times \mathbb{R}$, as a composition of a nudge–and–twist $h_i := \tau_i\nu_i$. Each N_i slice gets smaller and smaller in diameter as we apply consecutive h_is.

Applying the Big Shrinking Lemma, we are done. □

As we shall see in § 3, products $S^3/Wh \times \mathbb{R}$ arise as the simplest kind of Casson handle. Thus the fact that these products are standard is a big hint (if you're still skeptical) that the 4d Poincaré Conjecture might actually be true.

Remark 2.11. We didn't seem to use many properties of the Whitehead link in the proof of Theorem 2.10. This seems suspicious. But in fact, the proof really uses almost no topological properties of the Whitehead link beyond the fact that the two components have linking number zero.

The argument works just as well — even with links L_i instead of knots K_i at each stage — providing only that each component of L_i is homologically (equivalently homotopically) trivial in N_{i-1} (i.e. has linking number 0 with the meridian of N_{i-1}). The proof is exactly the same: successively nudge and twist each component of $N_{i-1} \times \mathbb{R}$ by a homeomorphism fixed on the boundary, which shrinks every N_i slice, and then apply Big Shrinking.

2.8. **Bing versus the world.** We can produce many interesting decompositions $\mathcal{D}$ of S^3 by starting with the unknot, and mixing Whitehead and Bing doubling at each stage. One can even throw in various other kinds of doubling, e.g the Whitney doubling in § 3.9.3.

Let's call a *generalized double* any operation which inserts a finite framed link in a neighborhood of the core of a framed knot. Then the following generalization of Theorem 2.9 is straightforward:

Theorem 2.12 (Bing beats anything)**.** *Let $\Lambda := \cap_i N_i$ be obtained as the intersection of a nested family of closed subsets of S^3, where each N_i is a finite disjoint union of solid tori, and where N_i is obtained from N_{i-1} by applying a generalized doubling move to each component of N_{i-1}. Suppose that generalized double moves are only allowed at stages $n(i)$, where every intermediate stage is Bing doubling of every component. Then there is an integer $m(i)$ depending on the particular sequence of doubles resulting in $N_{n(i)}$ so that if $n(i+1) - n(i) > m(i)$, the components of Λ form a shrinkable decomposition, and the quotient of S^3 by this decomposition is homeomorphic to S^3.*

In words, if we have enough Bing doubles between any other kind of doubles, the result is shrinkable.

Proof. It suffices to isotop each finite stage so that successive diameters of every nested sequence of components goes to 0.

Let N be a solid torus whose core has length T. Then for any $\epsilon > 0$ the result of Bing doubling N approximately $2^{T/\epsilon}$ times will produce components in N whose diameters are at most ϵ as measured in N. Of course, this implies that the diameters are at most ϵ as measured in S^3.

Now, no form of doubling increases diameters, but some doubles will increase the length of the core curve by some amount. But whatever the length of the core curves of the $N_{n(i)}$, they are all bounded by some finite T, and therefore for any ϵ, if we take $m(i)$ to be of order $2^{T/\epsilon}$, the diameters of every component of the $N_{n(i+1)}$ are bounded above by ϵ as measured in S^3. □

Such shrinkable decompositions literally arise as the frontiers of the Flexible handles we shall construct in the sequel.

If N is a solid torus and N' is a union of solid tori in N obtained by a generalized doubling of the core, the complement $N - N'$ is called a *drilled solid torus*. Since the components in a generalized doubling are framed, the drilled solid torus inherits framings of its boundary tori. If $\Lambda \subset S^3$ is obtained by recursive generalized doubling, the complement $S^3 - \Lambda$ is exhibited in a natural way as a union of drilled solid tori.

2.9. **Starlike and birdlike.** A compact subset X of $\mathbb{R}^n$ is *starlike* with respect to a point p, if it is a (closed) union of rays ending at p. A set is *starlike equivalent* if there is a (compactly supported) self-homeomorphism of $\mathbb{R}^n$ taking it to a starlike set. A starlike equivalent set is cellular.

Remember that a decomposition is *null* if there are only finitely many elements with diameter bigger than any fixed ϵ. In particular, a null decomposition is countable. Not every null decomposition of D^n is shrinkable, even if all components are cellular.

Example 2.13 (2-Bing). We produce a nested sequence of 2-component links in S^3, just as in the construction of the Bing decomposition, with the difference that the clasped links at each stage wind *twice* around the core of their parent torus. The decomposition elements consist of the infinite nested intersections.

These elements are all cellular, since each solid torus is unknotted in the previous one. Moreover, the decomposition can be chosen to be null: call the two initial components B and L. Each component corresponds uniquely to an infinite word in the alphabet B, L.

At each stage the union of the B and L components can't be too small, but we can arbitrarily insist that one of them (the L component) is as small as we like. No component has bigger diameter that its parent. Thus we can arrange that any sequence with a B at the nth stage has diameter (at this and subsequent stages) no bigger than 2^{-n}. With this choice, nontrivial decomposition elements correspond to sequences with finitely many B's, and there are only finitely many of these with diameter (ultimately) bigger than any fixed ϵ.

Despite this example, Bean proved the following:

Theorem 2.14 (Starlike null is shrinkable)**.** *Let $\mathcal{D}$ be a null decomposition of D^n in which every decomposition element is starlike equivalent. Then $\mathcal{D}$ is shrinkable.*

Proof. Let X be starlike with respect to the origin (without loss of generality). Then X is cellular, and therefore shrinkable in $\mathcal{D}$. Since the decomposition is null, what's wrong with just shrinking the components of the decomposition one by one in order from biggest to smallest?

The problem is that when we try to shrink some big decomposition element X, there might be some other sequence of smaller and smaller decomposition elements Y_i closer and closer to X, so that any self-homeomorphism ϕ shrinking X must also stretch some Y_i big.

With this in mind, we will show for any ϵ that there is a self-homeomorphism ϕ of D^n, supported in an arbitrarily small neighborhood of X, and satisfying

(1) the diameter of $\phi(X)$ is at most ϵ;
(2) for every other decomposition element Y either the diameter of $\phi(Y)$ is at most ϵ, or no point of Y moves more than ϵ.

This will be sufficient for our needs, since for any Y, either $\phi(Y)$ will not be much bigger than Y, or it will be of size less than ϵ. This is evidently good enough to let us shrink the decomposition elements one by one without accidently stretching one too much and undoing our progress. Incidentally, to construct ϕ as above we will not need to use the fact that every other Y is starlike-equivalent.

The homeomorphisms we construct will be radial; i.e. they take each ray centered at the origin to itself. Furthermore, the restriction to each ray will be compactly supported, and will be piecewise affine. Here's the construction.

Let X_i be a sequence of closed disks, each contained in the interior of the previous one, and with $\cap X_i = X$. Since X is starlike, we can choose such a X_i which is likewise starlike from the origin. We choose the X_i satisfying the following conditions:

(1) X_1 does not intersect any decomposition element Y of size $> \epsilon/2$; and
(2) No Y intersects both $X_i - X_{i+1}$ and X_{i+2} for any i.

Let B_i be a sequence of round closed balls of radius r_i centered at the origin, such that

(1) $r_i \to \epsilon$; and
(2) $0 < r_i - r_{i+1} < \epsilon/8$.

Now, define $Z_i := X_i \cup B_i$ for each i. Each Z_i is a starlike disk, and $\cap Z_i = X$.

We shall define ϕ to be the unique radial homeomorphism so that for each ray r and each i, the radial segment $r \cap (Z_i - Z_{i+1})$ is taken affinely to the radial segment $r \cap (B_i - B_{i+1})$ for all $i < n$, while Z_n is taken radially to B_n. Notice that any point $x \in B_i - X_i$ moves no more than $\epsilon/8$.

We now check that ϕ as above satisfies the desired properties, i.e. that for every Y, either ϕ moves no point of Y more than ϵ, or the diameter of $\phi(Y)$ is less than ϵ.

The proof falls into a small number of cases, all straightforward to analyze.

(1) If Y has diameter $> \epsilon/2$ then it doesn't meet any X_i. Thus every point is in some $B_i - X_i$, and therefore moves no more than $\epsilon/8$.
(2) If Y has diameter $\le \epsilon/2$ but still has every point in some $B_i - X_i$, then it is still true that no point moves more than $\epsilon/8$.
(3) If Y has diameter $\le \epsilon/2$ and *no* point is in any $B_i - X_i$, then there is some index j so that every point in Y is in $X_j - X_{j+2}$. It follows that $\phi(Y)$ is in $B_j - B_{j+2}$, an annulus of thickness at most $\epsilon/4$. If $x, y \in Y$ are arbitrary, let x', y' in ∂B_{j+2} be on the rays containing x, y. Note that the distance from x' to y' is less than the distance from x to y, which is bounded by $\epsilon/2$. On the other hand, the distance from x' to $\phi(x)$ is at most $\epsilon/4$, the thickness of the annulus, and similarly for y' and $\phi(y)$. By the triangle inequality the distance from $\phi(x)$ to $\phi(y)$ is at most ϵ.
(4) If Y has diameter $\le \epsilon/2$ with points x, y such that x is in $X_j - X_{j+2}$ and y is in $B_j - B_{j+2}$, define x' and y' in ∂B_{j+2} as before. It is still true that the distance between x' and y' is no more than the distance from x and y. And it is still true that the distances from x' to $\phi(x)$ and from y' to $\phi(y)$ are bounded by $\epsilon/4$. So the diameter of $\phi(Y)$ is at most ϵ.

□

It's important to stress that the condition of being starlike-equivalent does not just refer to the topology of a decomposition element, but how it sits in the ambient space.

A set is *birdlike* if it is 'recursively starlike': i.e. there are finitely many starlike retractions that shrink to a point. Denman and Starbird [4] observed that birdlike equivalent null decompositions are shrinkable too:

Theorem 2.15 (Birdlike null is shrinkable)**.** *Let $\mathcal{D}$ be a null decomposition of D^n in which every decomposition element is birdlike equivalent. Then $\mathcal{D}$ is shrinkable.*

Proof. We restrict attention to the case of a polyhedral birdlike set, since this is the only case we will need for applications.

As in the proof of Theorem 2.14, we only need to show for any $\epsilon > 0$ that we shrink a single big birdlike (bigbird-like?) component to diameter ϵ without increasing the diameter of any other component by size ϵ. Furthermore, we don't actually need to exhibit a homeomorphism performing this shrinking; it's enough to exhibit a map which is ABH, since an approximating homeomorphism will do the job.

The point of this is that we can replace the radial affine map ϕ on each $Z_i - Z_{i+1}$ by a new map that merges some of the radial segments. In its intrinsic path metric as a subset of Euclidean space, a (polyhedral) birdlike set X is CAT(0) so if we fix a point p in the interior, there is a unique geodesic segment in X (which is PL in the ambient space) from any $q \in X$ to p.

Using this geodesic flow, we can build a new 'radial' function in a neighborhood of X whose gradient trajectories are obtained by first flowing at unit speed along the intervals of a product collar into X, and then flowing in X to p. If we define balls and radial affine maps with respect to this function, the proof works exactly as above. □

2.10. **Ball to ball theorem.** The last substantial shrinking result we need for the Poincaré Conjecture is the *ball to ball theorem*, due to Freedman.

Theorem 2.16 (Ball to ball)**.** *Let $f : D^n \to D^n$ be onto, and let the associated decomposition of D^n be null. Further, let's suppose that the singular image (the image of the nontrivial elements) is nowhere dense. Let's also suppose that there's a closed subset E of D^n containing ∂D^n, so that the restriction of f to E is a homeomorphism. Then f is ABH, where the approximating homeomorphisms can be chosen to agree with f on E.*

Notice that we do *not* assume the decomposition is cellular.

Proof. Let's denote the nontrivial elements by X_i mapping to x_i. Since the decomposition is null, there's a largest nontrivial element X mapping to x.

Let U be a tiny open ball around x, and let B be tiny closed round ball containing x whose boundary avoids the singular image. Let U' and B' be the preimages of U and B under f.

If we knew B' was a ball, we could replace $f|B'$ by any homeomorphism $f' : B' \to B$ agreeing with f on $\partial B'$ and eliminate the largest nontrivial element of the decomposition. But we don't even know that X is cellular.

Let $\sigma : D^n \to B$ be a radial homeomorphism squeezing $D^n - U$ down to $B - U$ while staying fixed on U. If it made sense to conjugate σ^{-1} by f we would get a homeomorphism $f^{-1}\sigma^{-1}f : B' \to D^n$, and then we could choose any homeomorphism $h : D^n \to B$ agreeing with σf on E, and replace f by $hf^{-1}\sigma^{-1}f$ on B'. Notice that this map restricts to a homeomorphism h on U' and agrees with f on $D^n - B'$.

But this doesn't make sense as written, because f is not a homeomorphism, so f^{-1} is not a map; it's a relation. Nevertheless we have well-defined maps $\sigma^{-1}f : B' \to D^n$ and $fh^{-1} : B \to D^n$, and we can form the fiber product $\Delta \subset B' \times B$.

Now, Δ as so defined might contain product regions $X_i \times h(X_j)$ whenever $x_i = f(X_i) = \sigma^{-1}f(X_j) = \sigma^{-1}x_j$. Since the singular image is nowhere dense, we can perturb σ very slightly so no such coincidences occur for $X_i, X_j \subset B' - U'$; this ensures that the subset of Δ in $B' - U' \times B - U$, though it has both horizontal and vertical segments, has no product regions.

Since σ is the identity on U, we get product regions $X_i \times h(X_i)$ for all $x_i \subset U$; thus we should replace Δ by the graph of h^{-1} in $U \times U'$. Finally, we extend Δ by the graph of f in $D^n - B' \times D^n - B$. This new Δ has three properties:

(1) Δ is as close as we like to Γ_f (they only differ in $B' \times B$, and B is as small as we like);

(2) The horizontal and vertical segments of $\Delta \cap (B' \times B)$ have projections to both factors as small as we like.
(3) Away from horizontal and vertical segments, Δ is the graph of an injective function. If Q is the quotient of Δ obtained by crushing both horizontal and vertical segments to points, then Q is a disk, and the singular image is nowhere dense in Q.

But now we can continue this process, inductively crushing the biggest horizontal or vertical segment at each stage, converging in the limit to the graph of a homeomorphism approximating f. □

3. Kinks and Gropes and Flexible Handles

The idea of the proof of the Poincaré Conjecture in dimension 4 is the same as the proof in higher dimensions: we need to find embedded Whitney disks to promote algebraic intersection information to geometric intersection information. We can find immersed Whitney disks rather easily; by general position they intersect themselves in isolated points. Somehow we need to replace these immersed disks with embedded ones.

3.1. **Slice knots and signatures.** In 3-dimensions the famous *Haken lemma* says that if $f : D^2 \to M^3$ is a singular map which is an embedding near the boundary, then we can find a new $g : D^2 \to M^3$ which is an embedding, and agrees with f near ∂D^2. The new g is obtained by cut-and-paste argument from f; thus $g(D^2)$ can be found in any neighborhood of $f(D^2)$.

The same sort of argument can't be true in 4 dimensions: embedded Whitney disks can't be found 'locally'.

Example 3.1 (Slice Knots). Let K be a knot in S^3, which we think of as the boundary of D^4. Then K bounds a singular disk in D^4, obtained by coning the knot to a point. Actually, this disk is embedded. But it is not necessarily locally flat, and in fact K might not bound a locally flat embedded disk at all. A knot which does bound a locally flat embedded disk in D^4 is said to be (topologically) *slice*.

It is a fact that some knots are not slice. In fact, a slice knot K has an Alexander polynomial of the form $f(t)f(t^{-1})$, a fact due to Fox–Milnor. The figure 8 knot (for instance) has Alexander polynomial $A(t) = -t + 3 - t^{-1}$. Since $A(-1)$ is not a square, the knot is not slice.

We now explain the observation of Fox–Milnor.

Let F be a Seifert surface for K. Then $H_1(F) = \mathbb{Z}^{2g}$ where g is the genus, and there is a Seifert pairing

$$S : H_1(F) \times H_1(F) \to \mathbb{Z}$$

given by $S(x, y) =$ linking number of x with y^+, where y^+ is obtained by pushing y off R in the 'positive' direction. Note that this linking number is also given by the algebraic intersection of surfaces A, B in D^4 bounding x and y^+ respectively.

Lemma 3.2. *If K is slice, there is subspace L of $H_1(F)$ Lagrangian for S.*

Proof. Build F' by gluing a slice disk D onto F. Then F' is a closed, embedded, locally flat surface in D^4. We can push it slightly into the interior. We claim that there is an embedded 3-manifold M in D^4 that bounds F'. If F' is smooth, this is easy. By Alexander duality, $H^1(D^4 - F') = \mathbb{Z}$, so we can choose a map $D^4 - F' \to S^1$ in general position realizing the generator, and take the preimage of a regular point.

Now the kernel of $H_1(F') \to H_1(M)$ is Lagrangian L for the intersection form. If x and y are in L, they bound surfaces A and B in M. Push off y to y^+, at the same time pushing B off M to B^+. Then B^+ and A are disjoint, so $S(x, y) = 0$. □

The Alexander polynomial of a knot may be obtained from the Seifert pairing as $A(K) := \det(S - tS^T) \in \mathbb{Z}[t]$, up to sign and multiplication by a power of t. If K is slice, there are A, B, C so that $S = \left(\begin{smallmatrix} 0 & A \\ B & C \end{smallmatrix}\right)$ and

$$A(K) = \det \begin{pmatrix} 0 & A - tB^T \\ B - tA^T & C - tC^T \end{pmatrix} = \det(A - tB^T)\det(B - tA^T) = f(t)f(t^{-1})$$

3.2. **Immersions, framings and Euler classes.** It's important to find immersed disks and surfaces in 4-manifolds.

Lemma 3.3 (Immersion exists)**.** *Let R be an oriented surface and let W be an oriented 4-manifold. If $f : R \to W$ is any map, then f can be homotoped to an immersion. Moreover, if the restriction to some subsurface S is already an immersion, then f can be homotoped to an immersion rel. S.*

Proof. Since every surface can be immersed locally, we suppose f is already an immersion on a subsurface S. Since we can replace W in the hypothesis by any neighborhood of the image $f(R)$, it suffices to show that *some* immersion extending f on ∂S exists. This follows from the immersion theorem, which says (in the context where the domain has smaller dimension than the range) that any bundle map $T(R - S) \to TW$ is homotopic to an integrable bundle map — i.e. an immersion. There is also a relative version of this theorem: the map can be homotoped rel. its restriction to a subset where it is already integrable.

Now, f^*TW is a trivial $\mathbb{R}^4$ bundle over ∂S. The immersion near the boundary gives a loop in the Stiefel manifold $V_2(\mathbb{R}^4)$ of oriented 2-frames in $\mathbb{R}^4$. But $V_2(\mathbb{R}^4)$ is an S^2 bundle over S^3 and is therefore simply-connected. So the bundle map over the boundary always extends, and f can be approximated rel. boundary by an immersion. □

Actually, the same proof works whenever the target has dimension at least 4. The lemma is false in dimensions 3 and 2. The most important special case will be to a closed surface R for which $f : R \to W$ represents a homology class, and a map of a disk $f : D^2 \to W$ which is already an immersion on a collar neighborhood of ∂D.

Suppose R is a closed connected oriented surface, and suppose $f : R \to W$ is a smooth immersion where W is oriented. The normal bundle ν is an oriented $\mathbb{R}^2$ bundle over R. A framing of R is a trivialization. The obstruction to finding a framing is the self intersection number of the class $[R]$ in the 4-manifold ν; equivalently, it is the signed number of zeros of a generic section of ν. We write this $e(\nu)$.

At each self-intersection point of $f(R)$ the sign of the intersection is ± 1; let's let $I(R) = \sum \pm 1$, the sum taken over all the self-intersection points. If we perturb $f(R)$ normally to $f'(R)$ in W, then $f(R)$ and $f'(R)$ intersect $e(\nu)$ times at zeros of $f' - f$, but they also intersect ± 2 times at a self-intersection point of $f(R)$. In other words,

$$[R] \cdot [R] = e(R) + 2I(R)$$

where $[R] \cdot [R]$ is the intersection form in W. Notice that $e(R)$ is well-defined mod 2, and is equal to the evaluation of the second Stiefel-Whitney class w_2 of W on $[R]$.

Any connected oriented surface R with boundary can be framed, but if we have a framing on the boundary, the obstruction to extending this over the interior is the relative Euler number. We write this $e(R)$. Any closed connected surface R can be framed in the complement of a small disk D. On D the relative Euler class $e(D)$ is equal to $e(R)$.

An immersion can be modified locally by an *interior twist*. This inserts a new self-intersection point with sign ± 1, and changes the Euler class by ∓ 2.

If R has boundary, we can also modify R locally by a *boundary twist*. This changes the relative Euler class by ∓ 1. The boundary twist doesn't add a new self-intersection point for R, but it changes the homology class of R rel. its boundary. If R is attached transversely to S along ∂R, then the boundary twist adds a new intersection point of R with S. See Figure 7.

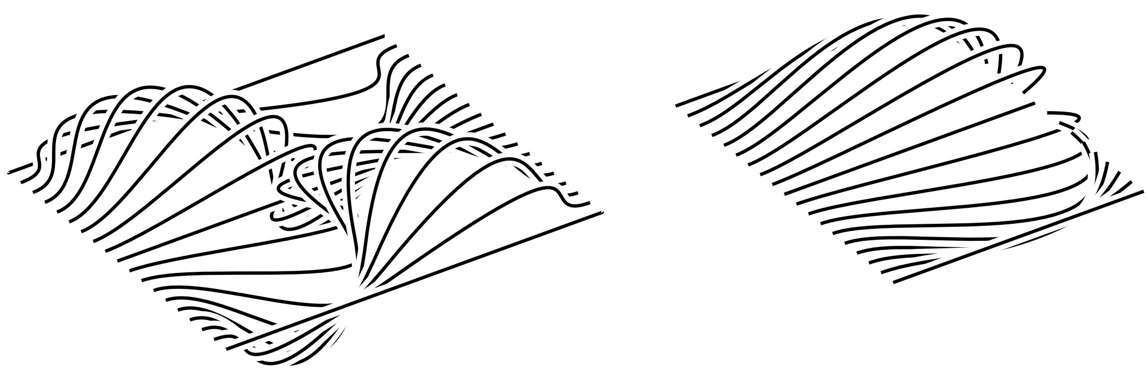


FIGURE 7. Interior twist and boundary twist

We can think of both twists as a 1-parameter family of immersions of the interval (the 'horizontal' slices in the figure), moving from top to bottom. The boundary twist is easier to understand: it is obtained by dragging the right hand tangent of D through a 2π revolution perpendicular to the right edge, and dragging the rest of the interval along. If we do this in $\mathbb{R}^3$ we get a 'kink' at angle π. But we can push this kink out into the 4th dimension to straighten it. The framing along the (right hand) boundary evidently changes by ± 1. It introduces one new point of intersection with S. The interior twist is two copies of the boundary twist, one obtained from the other by a rotation through 2π around the boundary arc.

Here's another picture of the interior twist. Start with two transverse planes in $\mathbb{R}^4$ intersecting D^4. That's the local picture near the self-intersection point. The link in $S^3 = \partial D^4$ is a Hopf link L. Attach a collar $S^3 \times [0, 1]$ to ∂D^4. The intersection of our surface with $S^3 \times [0, 1/2)$ is a pair of linked cylinders $L \times [0, 1/2)$. In the level $S^3 \times 1/2$ we tube the components of the Hopf link together by an unlinking tunnel for L. Summing these circles along this tunnel gives an unknot in $S^3 \times 1/2$ and we extend this by a product to give an unknot in $S^3 \times 1$. This gives a framed disk properly immersed in D^4 whose boundary is an unknot; thus we can insert this picture locally in any surface.

3.3. **Whitney moves and finger moves.** Let's recall the definition of a Whitney disk. Suppose we have immersed submanifolds A^p and B^q in W^{p+q} and let's suppose they intersect transversely at two points x, y. Let α and β be embedded arcs in A and in B between x and y, and let $\gamma = \alpha \cup \beta$.

An *immersed Whitney disk* D is a disk with the following properties:

(1) $\partial D = \gamma$;
(2) D is immersed and in general position with respect to itself and $A \cup B$;
(3) There is a splitting of the normal bundle νD as $E_A \oplus E_B$ where $E_A|A$ is the normal bundle of α in A, and $E_B|B$ is the normal bundle of β in B.

An *embedded Whitney disk* is embedded, and must also be interior disjoint from $A \cup B$.

Let's suppose we have an immersed disk D with $\partial D = \gamma$ in general position with respect to itself and $A \cup B$. When does it admit the desired splitting of the normal bundle?

Since α is an embedded arc in A, the normal bundle $\nu_A(\alpha)$ is trivial. It sits as a $(p-1)$-plane bundle in ν_D along α, and splits off an orthogonal $(q-1)$-plane bundle we call ξ. The splitting $\nu_D = \nu_A \oplus \xi$ along α extends to a splitting over all of D. Thus we need to compare the $(q-1)$-plane bundles ξ and ν_B over β.

Since A is transverse to B, the $(q-1)$-planes $\xi|x$ and $\xi|y$ have the same orientation as $\nu_B|x$ and $\nu_B|y$ if and only if the points x and y have opposite sign, with respect to a consistent choice of orientations for A, B and W in a neighborhood of D.

The trivialization can therefore be found if the bundles $\xi|\beta$ and $\nu_B|\beta$ are isomorphic rel. endpoints. The difference of any two trivializations agreeing at the endpoints is a based loop in the Stiefel manifold $V_{(q-1)}(\mathbb{R}^{p+q-2})$ of $(q-1)$-frames in R^{p+q-2}. Now, $V_1(\mathbb{R}^{p+q-2})$ is a $(p+q-3)$-sphere, and each $V_i(\mathbb{R}^{p+q-2})$ is a $(p+q-2-i)$-sphere bundle over $V_{i-1}(\mathbb{R}^{p+q-2})$. So the Stiefel manifold is simply-connected whenever $p > 2$. When $p = 2$ and $q > 2$ the fundamental group is $\mathbb{Z}/2\mathbb{Z}$, but when $p = q = 2$ we have $V_1(\mathbb{R}^2) = S^1$ with fundamental group $\mathbb{Z}$.

For $n = 4$ and $p = q = 2$ there is a $\mathbb{Z}$ obstruction to finding a suitable framing for D, which is just the relative Euler number. If this number is zero and the framing problem can be solved, we modify A by cutting out a tubular neighborhood of α, gluing in a copy of the sphere bundle of D in ν_A, and capping this off with a pushoff of $D^{p-1} \times \beta$ (in the direction of the outward normal of β in TD). This is the *Whitney move*.

If D is embedded, the result of the Whitney move is to remove the points x, y of intersection of A with B. Thus by a sequence of Whitney moves we can find new representatives A and B with $|A \cap B| = |[A] \cap [B]|$.

If D is immersed, then the result of pushing A over D will create two new intersections of A for each intersection of D. If D is disjoint from B then this move at least reduces the number of intersections of A with B.

Note that by applying enough boundary twists we can always solve the framing problem, at the cost of introducing new intersections of D with A or B (or both). In the cases we'll consider, it will usually be possible for us to push these intersections off A or B.

The inverse of a Whitney move is a *finger move*. Let's suppose A and B are surfaces, and C is a surface with boundary on B. Suppose A intersects C transversely in a point x, and y is a point on $\partial C \cap B$. If C is connected, we can find an arc α in C from x to y, and push A along this arc and through B off of C. This removes the point of intersection of A with C, but it creates two new points of intersection of A with B (of opposite sign). When we push A off α, it drags a segment in ∂C along with it, sweeping out a tiny disk W. This W is a Whitney disk, that can be used to cancel the two new intersection points.

Inserting a twist changes the framing, and a finger move changes self-intersection, but neither move changes the homology class. There is a third move which does, called (framed) sum. It's the immersed analog of a handle slide. Given immersed framed surfaces A, B we choose a path α from A to B, slide a finger of A along α, and then connect sum it with B at the end and push off using the framing of B. This produces a new framed surface A' representing the homology class of $[A] + [B]$.

Here is one of the the main applications of a framed sum. If there is an immersed framed sphere F that intersects B in exactly one point (such an F is called a *geometric dual* for B), then for each $p \in A \cap B$ we can do a finger move of A along an arc in B from p to F, and then do a framed sum with F to produce A' intersecting B in fewer points. For instance, if D is a disk with some boundary on B but otherwise disjoint from B, we can frame D by doing some boundary twists, creating new intersections of D with B, then we can push these intersections off B into copies of F.

It's very useful for B to have a geometric dual. In § 3.5 we'll explain how to systematically produce and use them.

3.4. **Accessories and Whitneys.** Suppose there's a framed immersed surface R with a single self-intersection point p. There's nothing to pair p with, so what do we do? There's another kind of disk we use in this context, called an *accessory disk*. Let's explain.

Let α be an arc in R from p to itself. It forms a loop in W. If this loop is null-homotopic in the complement of R, there's an immersion $D \to W - R$ taking $\partial D \to \alpha$, and maybe we can even find an embedded D. The normal to α in R is a line bundle. At p these two normals don't agree, but they are perpendicular, so there's a canonical one-parameter family of normals sweeping out a quadrant in the 2-plane they span. This 2-plane is perpendicular to D at p, so the net result of inserting this family of normals is to give a section of the normal bundle to D over ∂D, and since W is orientable, this gives a framing of $\nu(D)$ over ∂D. An extension of this to a framing over the interior of D is called an *accessory disk*. If D is immersed but otherwise framed in this manner, it's called an *immersed accessory disk*.

A neighborhood of $\alpha \cup D$ is D^4 and R intersects the boundary of this neighborhood in an unknotted 0-framed circle in the boundary S^3. So we can cut this neighborhood out and insert an embedded framed disk, and thereby eliminate the single self-intersection point. We call this an *accessory move*. It's nothing more than the inverse of an interior twist move.

Here's another way to describe it: an accessory move is an interior twist followed by a Whitney move. In more detail: near a self-intersection point p with a null-homotopic loop γ we introduce a new self-intersection point q by interior twist, so that p and q have opposite signs. Join p to q by α. Then $\gamma^\alpha := \alpha\gamma\alpha^{-1}$ can be perturbed to an embedded bigon, and an accessory disk A for γ can be perturbed to a Whitney disk W for γ^α. Now perform the Whitney move to eliminate p and q. Notice: if A is embedded then so is W.

One important difference between the Whitney move and the accessory move is that the former can be carried out by isotopy (resp. homotopy through immersions) if the Whitney disk is embedded (resp. immersed) but the accessory move changes the Euler number of R, and therefore can't be carried out by a homotopy through immersions. By the way, a homotopy through immersions is usually called a regular homotopy, but we won't usually use that term because it's kind of opaque.

3.5. **Transverse spheres.** Let V be a smooth simply-connected 4-manifold, and suppose $H_2(V)$ contains a hyperbolic factor; i.e. a pair of (primitive) homology classes α, β with $\alpha^2 = \beta^2 = 0$ and $\alpha \cdot \beta = 1$. Since W is simply-connected, α and β are represented by immersed two-spheres A, B.

One possibility is that A and B are embedded with trivial normal bundles, and intersect transversely in a single point. This is precisely the case where V can be written as $V'\sharp(S^2 \times S^2)$, and understanding V reduces to understanding the (simpler) manifold W'.

Now, $w_2(A) = w_2(\alpha) = \alpha^2 \bmod 2$, so the Euler class of the normal bundles of A and B are even. Therefore we can trivialize these bundles by doing interior twists; this will typically produce new self-intersections. Since $\alpha \cdot \beta = 1$, all but one of the intersections of A with B come in pairs with opposite signs.

How can we use global information (e.g. the fact that $\pi_1(V) = 0$) to 'improve' this geometric picture? Evidently we should start to look for Whitney disks.

Since V is simply-connected, we can find many immersed Whitney disks to pair oppositely oriented intersection points. But these disks might have many intersections with A and B, and the result of Whitney moves might make things more complicated, not less.

It's too much at this stage to ask for embedded Whitney disks, but at least we can ask for Whitney disks whose interiors are disjoint from $A \cup B$. This amounts to asking whether $\pi_1(V - (A \cup B))$ is trivial (one says in this case that $A \cup B$ is π_1-*negligible*).

Proposition 3.4. *Suppose V is simply-connected, and suppose we have framed immersed spheres A, B representing the generators α, β of a hyperbolic subspace of H_2. Then we can find new framed immersed spheres A, B representing α, β whose union is π_1-negligible, and all of whose intersections but one can be spanned by immersed Whitney disks interior disjoint from $A \cup B$.*

Proof. By Seifert-van Kampen, $\pi_1(V - (A \cup B))$ is (normally) generated by meridian loops around A and B. Evidently, π_1-negligibility is equivalent to the existence of *geometric duals*: immersed spheres S_A and S_B where S_A is disjoint from B and intersects A in one point; and conversely. We shall modify A and B so that they admit geometric duals.

B itself is a sphere with $A \cdot B = 1$ and $B \cdot B = 0$, so it is already an *algebraic dual*. Let's let our first approximation to S_A be B itself. Likewise we can let the initial S_B be A. The pairs A, S_A and B, S_B are algebraic duals, but not geometric ones. We improve the geometric situation in a sequence of steps.

1. Pick a pair of points of $A \cap S_A$ with opposite sign, and a Whitney disk W that might intersect A and S_A (not to mention B and S_B) in isolated points. By disjoint finger moves, we can push the intersections of W with A into A, and likewise with S_A. This creates new self-intersections of A and of S_A, but makes W interior disjoint from both. Then we can push A over W to remove two points of $A \cap S_A$. After finitely many such moves, we can get a new A and a new S_A, each homologous to the old, and $|A \cap S_A| = 1$.

2. Since S_A and B are homologous, and $B \cdot B = 0$, they intersect in an even number of points with opposite signs. A Whitney disk W for a pair of these intersections might cross the new A, but we can push these intersections into B by finger moves (along an arc in W). Also, push intersections of S_A with W through S_A, and intersections of B with W through B. At the end of this, W is interior disjoint from A, S_A, B and we can push S_A across W to remove two intersections of S_A with B. Repeat until S_A and B are disjoint, while preserving $|A \cap S_A| = 1$.

3. At this point we have constructed A, B, S_A where $|A \cap S_A| = 1$ and $|B \cap S_A| = 0$. Thus the meridian of A is trivial in $\pi_1(V - (A \cup B))$. We can therefore construct a Whitney disk W for each pair of intersections of B with S_B which is *disjoint* from A. As in step **1.** we can push interior intersections of W off B and S_B without creating new intersections with A, then push B over W to remove two points of $B \cap S_B$. After finitely many steps we get S_B with $|B \cap S_B| = 1$.

4. Finally, since S_B is homologous to A, and $A \cdot A = 0$, we can pair up points of intersection of S_B with A. Each of these spans a Whitney disk that is disjoint from B. Push interior intersections of W with A or S_B into A or S_B. Then push S_B across W to reduce $|A \cap S_B|$ by two. Eventually, A and S_B are disjoint.

After these four steps, $|S_X \cap Y| = \delta_{XY}$ for $X, Y = A, B$, and therefore $A \cup B$ is π_1-negligible in V.

The four steps taken so far are indicated in Table 2. Notice that every step takes framed surfaces to framed surfaces; thus we can assume A, B, S_A, S_B as constructed are framed.

TABLE 2. Promoting algebra to geometry

$*$	A	S_A	S_B	B
A	alg(0)	alg(1) $\to_1$ 1	alg(0) $\to_4$ 0	alg(1)
S_A		alg(0)	alg(1)	alg(0) $\to_2$ 0
S_B			alg(0)	alg(1) $\to_3$ 1
B				alg(0)

It remains to construct (framed!) Whitney disks in $V - (A \cup B)$ for all but one intersection point of $A \cap B$.

5. Once $A \cup B$ is π_1-negligible, we can find disks W interior disjoint from both associated to pairs of intersection points with opposite orientations. We must be careful: these disks are not yet necessarily framed, and a boundary twist might produce a new point of intersection of W with A (say). But we can tube this intersection along a path in A to the geometric dual S_A, and thereby push the intersection off A. Since S_A is framed, we obtain (framed) immersed Whitney disks interior disjoint from both A and B, pairing up all extraneous intersection points. □

There is a relative version of Proposition 3.4, with essentially the same proof.

Proposition 3.5. *Let V be a simply-connected 4-manifold with boundary, and suppose some $\alpha \in H_2(V, \partial V)$ represented by an immersed framed proper disk A has an algebraic dual β in $H_2(V)$ with $w_2(\beta) = 0$. Then there is some framed A' properly homotopic to A which is π_1-negligible and has a framed geometric dual.*

Proof. Represent β by B, a sphere which is an algebraic dual to A. Note that we cannot assume a priori that B is framed. Since $w_2(\beta) = 0$, the Euler number of the normal bundle is even, and we can frame B by performing finitely many interior twists.

Pair up intersections of A and B and span them by framed immersed Whitney disks that might intersect A and B. Push intersections of b with each Whitney disk W off of W, and then push A over W, eliminating two points of $A \cap B$. After finitely many steps, $|A \cap B| = 1$ so A is π_1-negligible, and the new B is a framed geometric dual to A. Furthermore, the new A is obtained from the old by homotopies. □

Insisting that $w_2(\beta) = 0$ is overkill. If $w_2(\beta) = 1$ we just need to know that α does not represent w_2. For, this implies there's another sphere F with $w_2(F) = 1$ and $\alpha \cap [F] = 0$, and we can sum B to F to change its Euler number by an odd integer while staying algebraically dual to A.

3.6. **Casson Handles.** We're just getting started. In 1973–74 Andrew Casson delivered a series of lectures in Cambridge and Paris introducing a powerful new infinite method in 4-manifold topology. This method produces some sort of weak substitute for an embedded Whitney disk when only immersed ones can be easily found. Casson's lectures were written up by Cameron Gordon and published in [3].

Theorem 3.6 (Casson)**.** *Let V be a simply-connected 4-manifold with boundary, and suppose some $\alpha \in H_2(V, \partial V)$ represented by an immersed framed proper disk A has an algebraic dual β in $H_2(V)$ with $w_2(\beta) = 0$.*

Then we can find a new framed properly immersed disk W_0 homotopic to A with the same boundary, and a sequence of open manifolds $N_i \subset V$ which are the interiors of compact manifolds with boundary $\overline{N}_i \subset V$ and such that the following are true:

1. *N_0 is a tubular neighborhood of W_0;*
2. *Each N_j is obtained from N_{j-1} by attaching tubular neighborhoods of a finite collection W_j of framed immersed Whitney disks properly embedded in $V - \overline{N}_{j-1}$, with boundaries attached to disjoint embedded circles in ∂N_{j-1};*
3. *The self-intersections of the collection W_{j-1} at each stage can be paired and joined up by disjoint arcs to make framed loops which are the attaching circles of the W_j.*

Proof. The idea is to induct on Proposition 3.5. The hypothesis implies that we can find a framed proper disk W_0 homotopic to the original A with a geometric dual S_A; i.e. an immersed framed sphere intersecting W_0 transversely in one point. Let N_0 be a tubular neighborhood of W_0.

The sphere S_A implies that N_0 is π_1-negligible, so we can find contractible Whitney circles for the self-intersections of W_0, and span these by immersed Whitney disks W_1 in $V - N_0$. Note that if an initial choice of bounding disk is not framed, and the obstruction is odd, we can do a boundary twist, and then sum with the geometric dual S_A to slide the new intersection point off W_0.

Let's write $V_1 = V - N_0$. This is a compact manifold with boundary. Each component of W_1 is a framed proper disk in V_1. We'd like to apply Proposition 3.5 to each component of W_1. To do so we need to find an algebraic dual in V_1 with even self-intersection number.

First of all, since W_0 is an immersed disk in a 4-manifold, and V is simply-connected, the fundamental group of N_0 is freely generated by its self-intersection points. In particular, each component of ∂W_1 is homologically essential and primitive in W_0, and therefore also in ∂N_0. For simplicity, suppose W_1 consists of a single disk; multiple disks can be handled consecutively. The fundamental class $[W_1]$ in $H_2(W_1, \partial W_1) \in H_2(V_1, \partial V_1)$ has boundary a primitive element in $H_1(V_1)$ and therefore it already had primitive image in $H_1(\partial V_1)$. Thus $[W_1]$ is essential and primitive in $H_2(V_1, \partial V_1)$, so by Lefschetz duality, there is a class $\beta \in H_2(V_1)$ with $[W_1] \cap \beta = 1$.

Let's let B be a sphere in V_1 representing β. If β^2 is even, we can apply Proposition 3.5. If β^2 is odd, we can frame B in the complement of a small embedded 2-disk $E \subset B$. Now take one point of intersection of $B - E$ with W_1 and push it off W_1 until it creates two points of intersection with W_0 of opposite sign. Now push these points of intersection off W_0 by doing framed sums with S_A. Let's call the resulting surface $B' - E$, and let B' be the result of gluing back E. Note that $B' - E$ is framed, but B' isn't; in particular $[B'] \cap [B']$ is odd. But

$$[B'] \cap [W_1] = [B] \cap [W_1] - 1 + 2[S_A] \cap [W_1] = [B] \cap [W_1] - 1 \bmod 2$$

is even. The existence of $[B']$ certifies that $[W_1]$ does not represent w_2 in V_1 after all, so we can apply Proposition 3.5 with $[B'] - [B]$ in place of $[B]$. This completes the induction step. □

Remark 3.7. If in the proof of Theorem 3.6 it turns out that the collection W_j is embedded at some finite stage, then we can take all W_k empty for $k > j$. In this case, $\overline{N}_j$ is already a 2-handle (we give a combinatorial proof of this in the next section).

The infinite union $N := \cup_{i=0}^{\infty} N_i$ is an open submanifold of V. It has an open $S^1 \times D^2$ denoted ∂^- in its frontier which is an open neighborhood of the original ∂A in ∂V. Every $\pi_1(N_i)$ is free, and dies in $\pi_1(N_{i+1})$; thus N is simply-connected. Likewise, the homology of each N_i is killed by the attaching maps of the W_{i+1}. Thus N has the homology of a point, and is therefore contractible. It turns out, though we shall not prove this, that N has the proper homotopy type of an open 2-handle (rel. the attaching surface ∂^-).

The space N is called a *Casson handle*. It's more usual to denote it CH, and that's what we'll do from now on.

3.7. **Kirby diagrams.** Smooth compact 4-manifolds can be described by handle decompositions. There is a visual convention for the display of such handle decompositions, called Kirby diagrams. A Kirby diagram, formally, is a link in S^3, each of whose components is decorated either with an integer, or with a dot. The union of the dotted components must be an unlink. The diagram represents a compact 4-manifold with boundary, obtained from D^4 by attaching 1-handles and 2-handles. A Dehn surgery diagram for the boundary 3-manifold is obtained from the Kirby diagram by replacing dots on circles with 0s.

Figure 8 shows a typical example. Here's how to interpret it. We start with a D^4, represented by nothing at all. Since S^0s can't link in S^3 there is only one way to attach

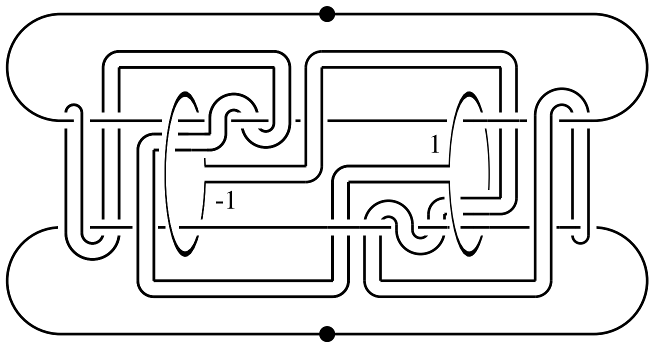


FIGURE 8. The Akbulut-Kirby sphere. Not a counterexample to the smooth Poincaré Conjecture.

1-handles to D^4. The result of attaching 1-handles is represented by an unlink on some number n of components, each with a dot, and the 4-manifold is $\natural_n S^1 \times D^3$ with boundary $\sharp_n S^1 \times S^2$. Passing through one of the dotted circles is shorthand for passing 'over' one of the $S^1 \times S^2$ handles. In the figure there are two 1-handles.

Now attach finitely many 2-handles. These are represented by their attaching circles — knots in the diagram, together with an integer label — the framing. In the figure there are two 2-handles, with framings -1 and 1 respectively. These numbers should be interpreted as follows. Suppose we have a circle K labeled by the integer n. The framing lets us push off a parallel circle K'. Choose the parallel such that $\text{link}(K, K') = n$. In other words, the framing is measured relative to the longitude — the boundary of a Seifert surface for K — where we're thinking of K here as a knot in S^3. By the way, the longitude framing differs from the so-called 'blackboard framing' (where the pushoff K' is drawn disjointly from K on the blackboard) by the writhe of the projection of K. So if K is an unlink, these framings agree, and an unlink labeled by 0 represents a framed embedded S^2 sitting as the core in $S^2 \times D^2$.

Figure 8 is a smooth contractible 4-manifold with boundary S^3. If we attach a 4-handle we get a homotopy 4-sphere. It turns out this example is diffeomorphic to S^4, but that's not so easy to see from the figure. This was demonstrated by Akbulut–Kirby [1], and the example is known as the Akbulut–Kirby sphere (in fact, the cited article is known to contain a gap, which by now has been filled).

3.8. **Homology and handle slides.** Suppose we have a Kirby diagram with no 1-handles representing a 4-manifold W. If K and K' are two attaching circles, we can cone them to the center of D^4 to make disks, and then attach the core disks of the 2-handles to build immersed spheres S_K, $S_{K'}$ in W. These spheres represent generators of $H_2(W)$. With respect to the intersection form on $H_2(W)$, the self-intersection number $[S_K] \cdot [S_K]$ is equal to the framing number (and similarly for $S_{K'}$), and the algebraic intersection $[S_K] \cdot [S_{K'}] = \text{link}(K, K')$. This follows from the definition of the framing. Handle slides are represented in these diagrams by doing a framed sum of the linking knots; i.e. pushing

K along an arc to K' and tubing it to a framed pushoff. The framing number of K changes by the framing number of K' under this move.

The 1-handles are a basis for $C_1(W)$, the cellular chain group, and the 2-handles are a basis for $C_2(W)$. If L_i are dotted circles representing the 1-handles h_i and K_j are attaching circles representing the 2-handles h'_j then the coordinate of $\partial h'_j$ in h_i is $\text{link}(K_j, L_i)$. Unless these linking numbers are all zero, the class of h'_j is not a 2-cycle and we can't interpret the framing number of K_j as a self-intersection number. If d is the greatest common divisor of the linking numbers of K_j with the L_i, then we may change the framing number of K_j by any multiple of d by cutting out disks D_i bounded by the L_i and regluing by an integer twist. This cut-and-paste move represents a diffeomorphism of W, so the resulting Kirby diagram represents the same 4-manifold. Note that the various strands of the K_j running through the L_i might change the way they link in the diagram under this move.

With the same notation, K_j is the attaching circle of the core of h'_j and D_i (or the sphere it represents) is the attaching sphere of the co-core of h_i. It follows that if there are D_i and K_j which intersect geometrically once, the handle pair may be canceled. When L_i and K_j are represented by a Hopf link, and the framing on K_j is 0, the effect of this cancellation on the rest of the diagram is to take the stuff linking K_j, and drag it over the stuff linking L_i with no change to the framing number of any component.

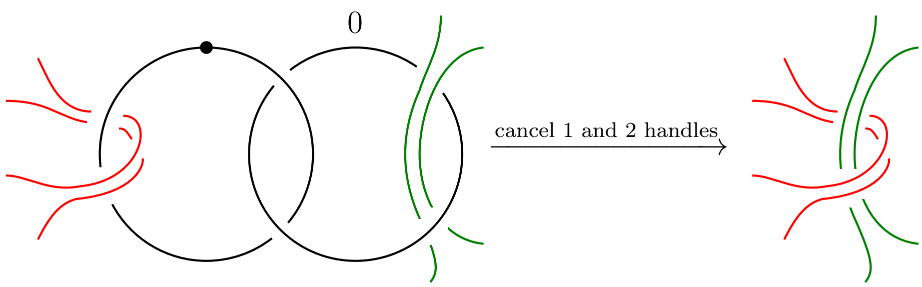


FIGURE 9. Cancelling a 1–2 handle pair

A tame embedded surface R in a 4-manifold is framed if the homology class it represents has self-intersection number 0. As a special case, if it represents the zero homology class, it's always framed. For a smooth surface in a smooth 4-manifold one can see this by mapping the complement of the surface to S^1 and observe that a regular preimage is a tame embedded 3-manifold M with $\partial M = R$. Then R is canonically framed by its normal bundle in M summed with the normal bundle of M in W. Since we are interested in framed spheres or surfaces, we almost exclusively consider diagrams in which every circle is dotted, or has framing 0. And as it happens, the framed circles we come across will typically be unknots.

3.9. **Some simple diagrams.** The 2-complexes that arise in the proof of the Poincaré Conjecture are made of simple pieces. In this section we draw Kirby diagrams of these simple pieces, so we can translate complicated 4-dimensional objects into pictures.

Each piece is attached to the previous pieces along a circle. In each diagram this circle is indicated in red. The places in this piece where subsequent pieces will be attached are indicated by green and blue circles. The effect on the boundary is to replace the complement of the red circle by a drilled solid torus associated to a generalized double in the sense of § 2.8.

3.9.1. *Kink with accessory disk.* A 2-handle in isolation is just D^4. The boundaries of the core and the co-core form a Hopf link in S^3, both components with the zero framing. A 2-handle with a single self-intersection is called a *kink*. The result of doing an interior twist in a surface is to insert a kink.

Let's draw a neighborhood of a 2-handle with a single kink, and see where the attaching circle of an accessory disk should go. See Figure 10. A kink is homotopic to a circle so a neighborhood is homeomorphic to $S^1 \times D^3$. The first drawing is a low-dimensional cartoon. The red arc represents the 2-dimensional core disk of the kinky handle, and the blue circle represents the 1-dimensional attaching circle of the accessory disk. The second drawing is more literal. A neighborhood of the intersection point is a D^4, and the red disk meets it in a Hopf link. A 1-handle connects up the two components of the Hopf link, and the blue circle runs over this 1-handle. Lastly there are two Kirby diagrams related by a (non-obvious) isotopy, coming from the symmetry of the Whitehead link that interchanges the dotted circle and the red circle.

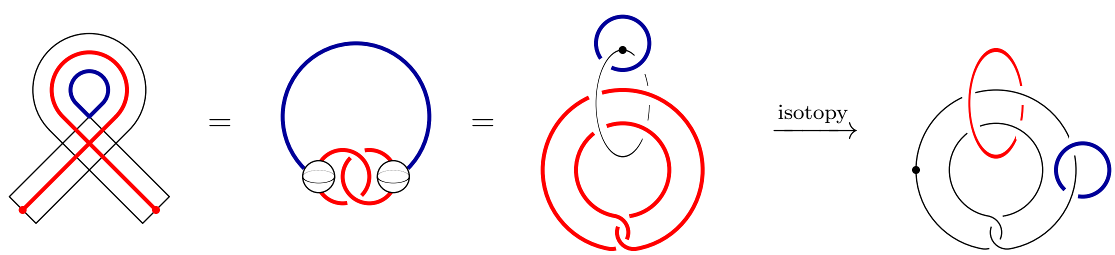


FIGURE 10. The red and blue circles with 0 framing are the attaching circles of the kink and the accessory disk respectively.

This corresponds to a positive kink, because the Whitehead component has a right-handed clasp. In a negative kink, the Whitehead component would have a left-handed clasp.

3.9.2. *Double kink with Whitney disk.* Two kinks can be paired by a Whitney disk if they have opposite signs. A neighborhood is homotopic to a wedge of circles, and is homeomorphic to the boundary connect sum $\natural_2 S^1 \times D^3$.

Notice that the two self-intersections have opposite signs, as they must to admit a Whitney disk. This translates into the two clasps in the red component having opposite handedness.

3.9.3. *Double kink with Whitney/accessory pair.* Here's a picture of a Whitney/accessory pair in a double kink. In the 2-complex the disks go through same singularity point. When we come to consider towers, we will find disks at the top with pairs of singularities; the tower caps will consist of Whitney/accessory pairs.

We call the operation that replaces the unknotted dual of the red curve with the green and blue curves a *Whitney double.*

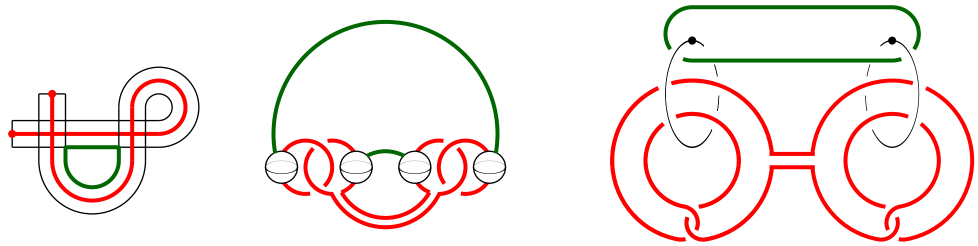


FIGURE 11. The red and green circles with 0 framing are the attaching circles of the double kink and the Whitney disk respectively.

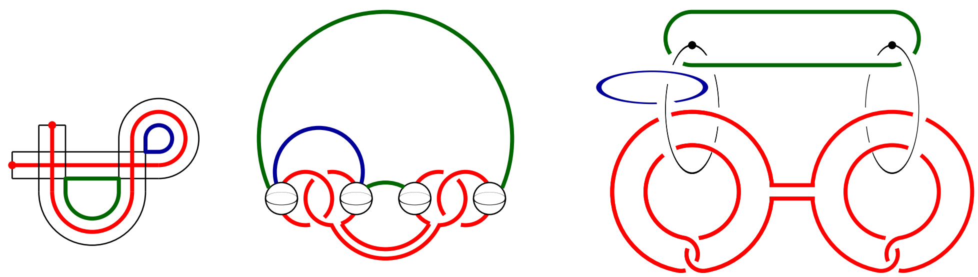


FIGURE 12. The green circle is Whitney, the blue circle is accessory.

3.9.4. *Surface with caps.* The effect on the boundary of attaching a kink in the construction of a Casson handle is to first glue in a solid torus and then drill out a neighborhood of the Whitehead double of the core. On the other hand, attaching a compact surface of genus 1 corresponds to gluing in a solid torus and drilling out a neighborhood of the *Bing double* of the core; see Figure 13. Higher genus surfaces add parallel copies of the Bing link.

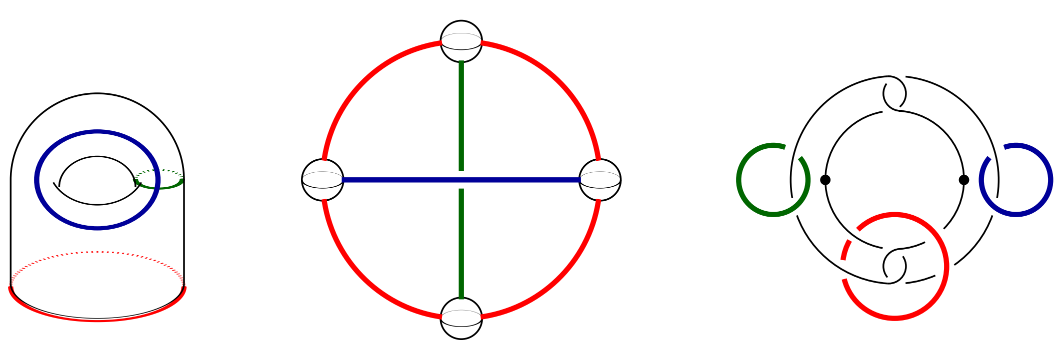


FIGURE 13. A surface of genus 1. This could be a stage in a grope.

3.9.5. *Casson handles.* The simplest (infinite) Casson handle has an accessory disk with exactly one self-intersection at each stage. This is represented by an infinite Kirby diagram of successive Whitehead links.

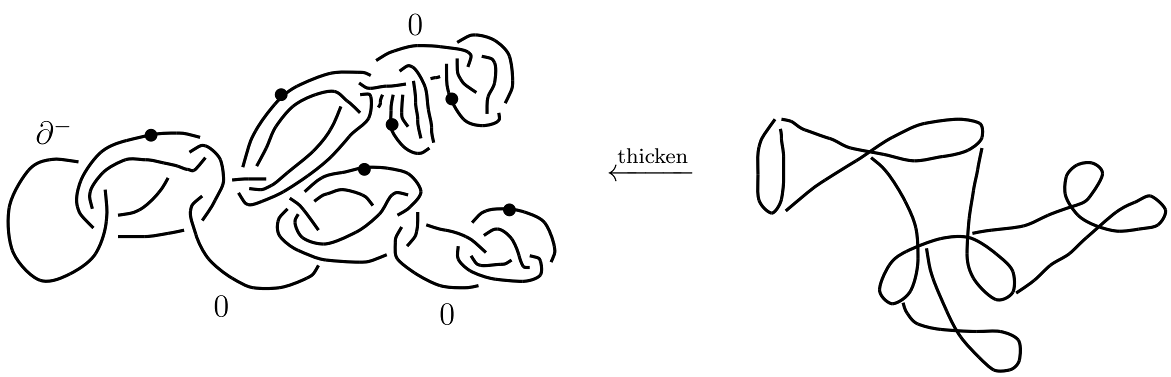


FIGURE 14. The first few stages of a Casson handle. The attaching handle is ∂^-.

The Casson handle CH is attached to the boundary of V along ∂^-, which recall is a $D^2 \times S^1$. The remainder of ∂CH is denoted ∂^+; as defined, it is an open infinite 3-manifold, which is the result of interpreting the infinite Kirby diagram from before as an infinite Dehn surgery diagram for a 3-manifold.

In the case of the simplest Casson handle, where the accessory disks have a single self-intersection at each stage, this 3-manifold is just $D^2 \times S^1 - Wh$. If we had enough geometric control, we might be able to arrange for the Whitney disks to converge to a single point, thereby compactifying the boundary to $D^2 \times S^1/Wh$. Of course this is not a manifold. But it is a manifold factor, and since we are in a 4-manifold we might hope to be able to find a nearby genuine $D^2 \times S^1$.

More self-intersections in the accessory disks at each stage give rise to parallel Whitehead doubles. The limiting 3-manifold is a solid torus minus $Wh \times$ Cantor set. The quotient $D^2 \times S^1/(Wh \times \text{Cantor set})$ is also a manifold factor, and for the same reason, since the shrinking construction in the proof of Theorem 2.10 can be applied component by component.

3.10. **Why Grope?** Remember gropes? They're back.

Suppose S is a framed surface in V with a pair of self-intersections, and W is an immersed Whitney disk in $V - S$. We can push S over W and eliminate the self-intersections to create S'. Of course, S' might have new self-intersections, two for every self-intersection of W.

If we are determined to replace S with an embedded framed surface, we can do it — providing we're prepared to raise the genus. Push a finger of S along one boundary arc α of ∂W, and do framed sum of S with itself. Let's call the result R. The surface R has two fewer self-intersections than S, but its genus is one higher. Notice that the meridian of the finger bounds an immersed Whitney disk W' in $V - S$. The boundaries of the disks ∂W and $\partial W'$ intersect transversely in a single point in R; i.e. they form an embedded symplectic basis for a $\mathbb{Z}^2 \subset H_1(R)$.

The Whitney disk W is still there in $V - R$, after nudging it slightly off α. It spans a longitude of the new handle of R. And there's a *dual* Whitney disk W', spanning the meridian of the handle (it's a tiny transverse disk to α). The disks W and W' are on opposite 'sides' of R; this only makes sense if we are implicitly thinking of R in a three-dimensional submanifold of V. Moreover, the interiors of W and W' are disjoint, and their boundaries intersect transversely in a single point. The union $R \cup W \cup W'$ is said to be a 1-stage *capped* grope. The grope is said to be *S-like*. R is the *body* and $W \cup W'$ the *caps*.

We can recover the original configuration $S \cup W$ by doing a Whitney move across W'; or we can recover $S' \cup W'$ by doing a Whitney move across W. But there is a third move called *compression* in which we cut out a tubular neighborhood of $W \cup W'$ and replace it by a single disk shaped somewhat like a saddle. This new disk is made from two copies of each of W and W' together with a square which is the neighborhood of $\partial W \cap \partial W'$ in R.

Compression can be used to get rid of unwanted intersections of W and W' with other surfaces. Suppose we have A intersecting W and B intersecting W'. We can push A by a finger move off W and into the gap between the new parallel copies of W', while pushing B by a finger move off W', into the gap between the new parallel copies of W. This creates two new intersections of A with B, but removes the two intersections we started with.

Formally, an n-stage grope E in V consists of the following:

(1) A proper framed embedded compact surface with boundary R_1;
(2) Let $3 \leq j \leq n$. Then R_j is a union of framed compact surfaces with boundary, properly and disjointly embedded in $V - \cup_{i<j} R_i$, so that the attaching circles ∂R_j form a symplectic basis for $H_1(R_{j-1})$. For $j = 2$, we only require that ∂R_j form a symplectic basis for some subspace of $H_1(R_1)$.

If S is obtained by compressing R_1 along ∂R_2, we say the grope is *S-like*. We are typically interested in disk-like gropes, those for which ∂R_2 form a complete symplectic basis for $H_1(R_1)$. The union $E := \cup_j R_j$ is the *body*.

A *capped* n-stage grope in V is the same as E together with immersed Whitney disks W (the *caps*) for a symplectic basis of R_n properly embedded in $V - E$. If E is disk-like, the caps certify that E is π_1-null. Finally, we say a (framed) grope is *properly immersed* if the body is embedded and the caps are immersed interior disjointly from the body.

By the way, the number of accessory disks attached at each stage of a Casson handle might increase, but it doesn't have to. The simplest Casson handle has exactly one accessory disk with exactly one self-intersection at each stage. But the number of surfaces in a grope necessarily grows exponentially. See Figure 15.

This proliferation of horizontal complexity, so to speak, lets a grope do more with less vertical complexity. We shall see an important example of this in § 3.11.

3.11. **Transverse gropes.** If E is a properly immersed capped grope, another properly immersed sphere-like capped grope E_t is *transverse* if there is one point where the bottom stage of E intersects the bottom stage of E_t transversely, and all other intersections are between caps. If we totally contract E_t, it becomes a sphere, and intersects the body of E only once in the first stage.

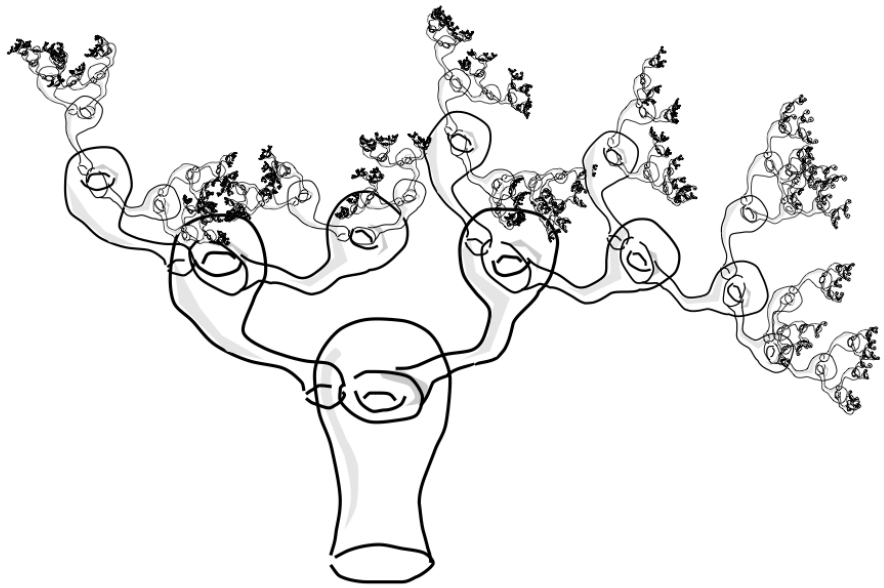


FIGURE 15. The simplest nine stage grope.

A capped grope E of height at least 2 can be thought of as being obtained from the first stage surface by attaching capped gropes $E^{\pm}$ of height at least 1 to a standard meridian-longitude basis of the first stage surface. Note that $E^{\pm}$ are always disk-like, whether E is or not.

Lemma 3.8 (transverse grope)**.** *Let E be a capped grope of height at least 2. Then E^+ admits a transverse grope E_t^+ of height at least 1.*

Proof. We take two parallel copies of E^-, one slightly in the past, the other slightly in the future. Note that these parallel copies are disjoint from each other and from E except in their caps. Push the boundary of E^- down, and take a product with a timewise interval to make an annulus, which glues up with the two parallel copies to make E_t^+. Since E^- is disk-like, E_t^+ is sphere-like. Only the annulus intersects E^+ transversely in one point in the first stage. □

Example 3.9. Let W be a Whitney disk for intersections of A and B. Let p be a point of $A \cap B$. Then the intersection of $A \cup B$ with a nearby sphere is a Hopf link, and the complement of this Hopf link is a torus that intersects W transversely in one point. The torus can be capped by normal disks for A and B. The result is a capped sphere-like grope W_t transverse to W.

3.12. **Height raising for Gropes.** In Freedman's original proof of the Poincaré Conjecture, one of the key steps is the Reimbedding Theorem, which says that inside the first 6-stages of a Casson handle, you can always find a 7-stage one (and then by induction, an n-stage one for any n). The proof is very difficult. This is a place where gropes show their advantages over Casson handles. The analogous result for gropes is called height raising.

Proposition 3.10 (Grope height raising)**.** *Suppose E is a properly immersed capped grope in V. Suppose further that E^+ has height at least 1. Then E can be extended to an n-stage capped grope in any neighborhood N of E, whose first stage agrees with that of E.*

Proof. The hypothesis says in the extreme case that E^+ is a (possibly disconnected) capped surface, while E^- consists entirely of caps. We express this by saying $E^\pm$ has height $(1, 0)$ (this is sometimes expressed colloquially by saying that E has height 1.5). We ignore the bottom stage of E, since it plays no role. It starts disjoint from everything, and it will stay that way throughout the argument.

For simplicity we assume E^+ consists of a genus 1 surface R with caps A, B and E^- consists of a single cap W. Note that A, B, W are all framed, but immersed, and possibly there are intersections of all three. We improve the situation in a series of steps.

1. Construct F, a sphere-like capped grope of height 1 as in Lemma 3.8 from two copies of E^+ and an annulus transverse to W at one point. The body of F intersects W in exactly one point and is disjoint from R, but the caps of F might intersect each of W, A, B in many points.

2. Create a large number of parallel copies of F, and contract their top stages one by one. This produces a sequence of framed immersed spheres $F'_1, F'_2, \cdots$, each of which intersects W transversely in a single point, and are otherwise disjoint from everything else; i.e. E^+, F, other F'_j, and so on.

3. Slide intersections of the caps of F with W over some F'_i. The new F is disjoint from W except at one point. Slide intersections of A, B with W over some F'_j. The new A, B are disjoint from W. Thus W intersects only itself.

4. Slide self-intersections of W over parallel copies of F. At the end of this process, W is replaced by a capped grope of height 1.

Thus we can go from $(1, 0)$ to $(1, 1)$. Repeat this with the roles of $+$ and $-$ alternately exchanged. The heights of $E^\pm$ grow like

$$(1, 0) \to (1, 1) \to (2, 1) \to (2, 3) \to (5, 3) \to \cdots$$

in particular, they grow without bound (and at a geometric rate). □

We indicate how to reach the hypothesis of Proposition 3.10 from the hypothesis of Theorem 3.6.

Proposition 3.11 (Immersed disk to small grope)**.** *Let V be a simply-connected 4-manifold with boundary, and suppose some $\alpha \in H_2(V, \partial V)$ represented by an immersed framed proper disk A with embedded boundary has an algebraic dual β in $H_2(V)$ with $w_2(\beta) = 0$.*

Then there is a properly immersed capped disk-like grope E in V with the same boundary as A, and whose complete compression is in the class of α, for which $E^\pm$ have height $(1, 0)$. Furthermore, there is a geometric dual S_A for the first stage of E.

Proof. First, Proposition 3.5 says we can find a new framed immersed disk A' representing α with the same boundary as A with a geometrically dual framed sphere S_A (so that $|A' \cap S_A| = 1$).

We can transform A' into a framed embedded surface R_0 by pairing self-intersection points of opposite sign, and tubing one side to the other by a finger move. Because A' has a geometrically dual sphere S_A it's π_1-negligible, and the same is evidently true of R_0: a Whitney disk for each pair of intersections stays interior disjoint from R_0. So we can find framed immersed caps for R_0. We have successfully constructed a grope E of height 1 to replace A.

Let W and W' be Whitney disks for each half of a symplectic basis for $H_1(E)$. For simplicity, let's assume each consists of a single immersed disk. The interiors of W, W' are disjoint from R_0, but they might intersect themselves, each other, and S_A.

We need to replace W by a capped surface whose body is disjoint from W' and from the interiors of its own caps. The induction step in the proof of Theorem 3.6 implies that we can replace W by a new W which admits a geometrically dual sphere S_W with $|S_W \cap W| = 1$. We can push intersections of S_A with W over copies of S_W to make S_A disjoint from W. Then we can push intersections of W' with W down into R_0 and over S_A so that W' is disjoint from W. Finally we can tube pairs of self-intersections of W to make a new embedded surface R_1. Immersed caps for R_1 can be framed (with a boundary twist if necessary) then pushed off R_1 onto S_W and off R_0 onto S_A so that the resulting grope is properly immersed. □

Putting these propositions together we conclude:

Corollary 3.12 (Immersed disk to tall grope)**.** *Let V be a simply-connected 4-manifold with boundary, and suppose some $\alpha \in H_2(V, \partial V)$ represented by an immersed framed proper disk A with embedded boundary has an algebraic dual β in $H_2(V)$ with $w_2(\beta) = 0$.*

Then there is a properly immersed capped disk-like grope E in V with the same boundary as A, and whose complete compression is in the class of α, of any desired height. Furthermore, there is a geometric dual S_A for the first stage with $|S_A \cap E| = 1$.

Proof. Apply Proposition 3.11 to produce a grope E of height $(2, 1)$ with a geometric dual S_A for the first stage, and then apply Proposition 3.10 to grow E as high as we want plus one stage.

The geometric dual S_A for the first stage of E might *a priori* intersect higher stages. Here's how to fix this up. Every body stage R^+ of E above the first comes with a pair R^- on the other side of the same height. Taking two copies of this pair and an annulus, we can construct a dual grope R_t^+ for every stage, with caps intersecting only the caps of E. Any intersection of S_A with R^+ can be piped over copies of R_t^+. Eventually this replaces S_A with a new (not necessarily properly) immersed grope, whose body is disjoint from the body of E except for a single point at the first stage. Completely contract this grope to obtain a new geometric dual S_A intersecting the body of E in only one point in the first stage, then contract the top stage of E so that the caps of E are disjoint from S_A too. □

3.13. **Towers.** Let D be a standard properly embedded D^2 in D^4. We can create two points of self-intersection by doing a finger move, producing an immersed disk D'. In more detail, we let α be an embedded path in D^2 between two points, and let β be a proper arc in $D^4 - D$ between the same two points, so that $\alpha \cup \beta$ bounds an embedded disk A.

Push D at one endpoint of α along β until it crosses itself. The finger move creates an embedded model Whitney disk W. The disk A is called an *accessory disk*, and intersects

W in one point. The operation that produces D' (together with W and A) from D is called a *clean finger move*. A *model cleanly immersed grope cap* is a disk obtained from an embedded D by doing finitely many clean finger moves as above. The *model tower caps* are the Whitney and accessory disks that arise. Notice that a model cleanly immersed grope cap is the same as an ordinary immersed grope cap, except that it's not allowed to intersect other grope caps as it might in a properly immersed grope.

A *tower* has *stories* which are capped gropes, and whose caps are cleanly immersed grope caps. A tower is S-like if the grope in the bottom story is S-like, and all the gropes in higher stories are disk-like. A model one-story capped tower is obtained from a model capped grope by doing finitely many clean finger moves in each grope cap and attaching Whitney and accessory disks. The $(n+1)$st story of a tower is obtained by replacing the Whitney and accessory disks of the nth story by one-story capped towers.

The *body* of a capped tower is everything except for the Whitney and accessory disks in the top story. Notice that the body includes the cleanly immersed grope caps in each story. The Whitney and accessory disks in the top story are the *tower caps*. A tower is *properly immersed* if the body is embedded, and the tower caps are immersed interior disjointly from the body. Note that the grope caps at each stage in a properly immersed tower are cleanly immersed but not literally embedded.

3.14. **Height Raising for Towers.** Heights and stories can be raised in towers as easily as in gropes. For technical reasons (explained in § 3.15) we do not assume the ambient 4-manifold V is simply-connected. For this reason, it is not so useful to talk about π_1-negligibility, rather we focus on the somewhat orthogonal property of π_1-nullity. A subset X of Y is π_1-*null* if the image of $\pi_1(X)$ in $\pi_1(Y)$ is trivial. If X is a finite subcomplex of a manifold Y then π_1-nullity of X is the same as π_1-nullity of a regular neighborhood.

Lemma 3.13 (Adding tower caps)**.** *Let E be a properly immersed capped grope of height at least 2 which is π_1-null and has a geometric dual sphere S for its first stage intersecting E in exactly one point. Then after adjusting the caps, E can be extended to a properly immersed capped one-story tower T and S adjusted so that it intersects T in exactly one point.*

Proof. The first step is to get the caps disjoint from each other. Construct sphere-like properly immersed gropes $E_t^\pm$ of height at least 1 which are geometric duals for the first stage of $E^\pm$ but possibly intersect the caps of E. We construct an enormous number of parallel copies of $E_t^\pm$ and then contract their top stages one by one. This might introduce new self-intersections of the caps of E, but at the end of this process we get arbitrarily many *disjoint* parallel spheres $S_i^\pm$, each of which is a geometric dual for the first stage of $E^\pm$ and is otherwise disjoint from E, even from its caps. Push self-intersections of the caps of E down each other and over $S_i^\pm$. The result is a new properly immersed grope E, but now each cap at the top layer intersects only itself (i.e. it is cleanly immersed).

Since the new E is contained in a neighborhood of the old, it is still π_1-null. Thus we can find immersed Whitney and accessory disks for the caps of E. These might intersect the body of E but we can push these intersections down and over copies of S.

This gives us T, a properly immersed capped one-story tower. S still does not intersect the capped grope E, but it might intersect the caps of T. Now, at this stage, the caps of T

might intersect our $S_i^\pm$. So we throw these $S_i^\pm$ away, construct a new collection of parallel sphere-like transverse gropes $E_t^\pm$, then contract their top stages one by one. This produces new $S_i^\pm$ which are disjoint from S and from T except for one point in the body. Now we can push intersections of S with the caps of T down to the second stage and over disjoint $S_i^\pm$. □

Proposition 3.14 (Tower height raising)**.** *Suppose T is a properly immersed n-story disk-like capped tower in V and suppose that the grope in the first story has height at least 3. Then for any N the top story grope caps can be changed so that the tower extends to T', a properly immersed $(n+1)$-story disk-like capped tower, whose top story has grope height at least N.*

Proof. Note that $\pi_1(T)$ is generated by intersections amongst the tower caps of the top story of T.

Let E be the grope in the first story of T, and let $E^\pm$ be the gropes beginning at the second stage of E. If we create a transverse sphere-like grope E_t^+ from two copies of E^- and an annulus, then E_t^+ intersects the first stage of E^+ transversely. But the grope caps of E_t^+ intersect the grope caps of E^- because these caps are not literally embedded, but only cleanly immersed.

If this were the top story, the intersection points of the grope caps would have (immersed) Whitney and accessory disks. Otherwise, they would have properly immersed capped towers in place of these disks. Nevertheless, we can still do a 'grope Whitney move' by cutting out a neighborhood of one Whitney arc and gluing in two copies of the next story. Continue this inductively until the top of the top story, to obtain $(S')^\pm$.

It is slightly inconvenient that the $(S')^\pm$ are transverse to the second stage of E and not the first. We solve this problem by letting V' be equal to V minus a tubular neighborhood of the first stage of E; then $E^\pm$ become disk-like capped gropes at the bottom stage of capped towers $T^\pm$, and now the $E^\pm$ each have geometric duals to their *first* stage.

Throw away the copies of the very top tower caps of $T^\pm$ in $(S')^\pm$ to obtain $S^\pm$. Since these top tower caps carry π_1, it follows that $S^\pm$ are π_1-null (in V'). Contract the top stage of $S^\pm$ so $|S^\pm \cap T^\pm| = 1$. Notice $S^\pm$ are still π_1-null in V'.

Now apply grope height raising (Proposition 3.10) so S^+ has height at least $N+2$. Push many parallel copies of S^+ off itself and contract the top stories to get arbitrarily many disjoint parallel height $N+1$ properly immersed sphere-like capped gropes S_i^+, each intersecting T^+ only in single points in the first stage of E^+. These new gropes are contained in a neighborhood of S^+ so they are all π_1-null.

Take each self-intersection point in the tower caps (i.e. the Whitney and accessory disks) on the $+$ side and push them down to the first stage of E^+, then pipe them off with the new parallel gropes. The result has height $N+1$ on the $+$ side but does not yet have tower caps, and furthermore the grope caps in the top story might intersect each other. Contract the top story of these gropes. They now have height N, and the top story caps are disjoint from each other. Because we got rid of intersection points in the tower caps, and piped them into disjoint π_1-null capped gropes S_i^+, the new caps are π_1-null.

Furthermore, their self-intersections come in pairs with (immersed) Whitney disks, because they were obtained by push off. Find immersed accessory disks using π_1-nullity.

Correct the framings by twists, then push self-intersections and new intersections with the grope caps down to where they can be piped off with copies of S_i^+.

Finally, switch the role of $+$ and $-$ to do the same on the $-$ side. □

3.15. **Squeezing towers.** It seems that towers are as easy to produce as gropes but it is not yet clear why we want the extra combinatorial complexity that comes with interrupting our gropes every so often for a layer of cleanly immersed caps.

We want our gropes or towers to converge geometrically, so that their closure in the ambient manifold V is homeomorphic to their end compactification. We're going to arrange this via *squeezing*: if X is a subspace of Y and N is a neighborhood of X we say Y can be *squeezed* into N if there is an isotopy $Y \times [0, 1] \to Y$ starting at the identity, and with $Y \times 1$ lying in N.

If S is a compact surface with boundary, then S squeezes into a neighborhood of a graph, consisting of the circles in a symplectic basis together with a finite number of arcs to connect it. If S is the first stage of a grope E, then we can perform this squeezing repeatedly, surface by surface. Finally, if T is a capped tower, we can squeeze all of T into a graph Γ, consisting of circles generating π_1 in the tower caps, together with trees connecting up components.

Now in a 4-manifold, homotopy implies isotopy for graphs, and for their tubular neighborhoods. So if Γ is π_1-null, we can squeeze all of $N(\Gamma)$ — and therefore all of T — into a ball in V of radius ϵ.

This procedure works well for a finite story tower, but it's important to be able to performing a relative squeezing for infinite story towers, where each successive story is squeezed tighter and tighter, leaving previous stages alone. Actually, once we've put the top story components of a finite tower inside tiny disjoint balls, it's more convenient to grow successive stories from these components, using Proposition 3.14.

Let's summarize this with a Lemma:

Lemma 3.15 (Squeeze the top stories)**.** *Let T be a properly immersed capped tower with at least 2 stories. Then for any $\epsilon > 0$, after possibly adjusting the tower caps, we can squeeze the components of the top story into disjoint embedded balls, all of radius at most ϵ.*

Proof. First, grow T to at least 3 stories using Proposition 3.14. If T' is a component of the third story, then we can throw away the caps of T' to produce a (componentwise) π_1-null capped grope E, and then completely contract E to get new tower caps for the top story of the original T. These new tower caps have the following two properties:

(1) because they are obtained by contraction, although they may have self-intersections they do not intersect each other; and
(2) because components of E were π_1-null, the new tower caps, and hence the new tower itself, is π_1-null.

Squeeze each component of the top story of the new T into a neighborhood of a graph, and then using π_1-nullity squeeze each graph neighborhood into a separate ball of radius $< \epsilon$. The rest of T moves by an ambient isotopy. □

Note that although the squeezed tower is π_1-null in the ϵ-ball, lower stories might also intersect that ball, and the squeezed tower might not be π_1-null in the complement of them. But towers can be grown with no hypothesis on π_1 of the ambient manifold, and by throwing away the caps of an extra story and contracting, we can achieve π_1-nullity, and iterate the procedure, all while staying in the ϵ-ball.

If we try to do the same thing just with gropes, we can certainly use π_1-nullity in a simply-connected V to squeeze most of a grope into an ϵ-ball, and then continue to grow the grope inside that ball, but we can never be sure that the result will be π_1-null in the ball in the complement of the lower stages. Thus we get stuck after squeezing once, with no obvious way to squeeze again. Towers by contrast carry their own local certificate of π_1-nullity in the top story caps.

3.16. **Convergent infinite towers.** Putting together the conclusions of the previous few sections we obtain the following:

Corollary 3.16 (Flexible handles exist)**.** *Let V be a simply-connected 4-manifold with boundary, and suppose some $\alpha \in H_2(V, \partial V)$ represented by an immersed framed proper disk A with embedded boundary has an algebraic dual β in $H_2(V)$ with $w_2(\beta) = 0$.*

Then there is an embedded infinite tower T with the same boundary as A, and a (relatively closed) neighborhood N of T so that the closure $F := \overline{N}$ is homeomorphic to the end compactification of N, and whose frontier ∂F is equal to the end compactification of ∂N, and is furthermore homeomorphic to S^3.

Proof. By Corollary 3.12 we can find a properly immersed disk-like capped grope E of any desired height, with a geometric dual S for the first stage. By Proposition 3.13 we can grow E to a 1-story properly immersed capped disk-like tower T.

By iterating Proposition 3.14 we can change the tower caps and extend T to a properly immersed disk-like capped tower with 3 stories.

By iterating Lemma 3.15 and Proposition 3.14 we can push the components of the top two stories into smaller and smaller disjoint balls and then continue alternately to grow them and to squeeze them. The result is an embedded infinite tower F whose closure is homeomorphic to its end compactification, and the same is true for a suitable (relatively closed) neighborhood N which gets thinner and thinner (i.e. it tapers) as it gets closer to the ends.

Now, by the small print in Proposition 3.14 all this can be achieved in such a way that the story heights $N_1, N_2, N_3, \cdots$ grow as fast as we desire. In particular, Theorem 2.12 implies that we can choose this sequence to grow fast enough that the frontier of N is the complement of a shrinkable decomposition in S^3. Since F is the end compactification of N, its frontier is the end compactification of ∂N, which is S^3. □

In these notes we'll call F as above a *flexible handle*. This was Casson's original terminology for his handles, so you might think I'm asking for trouble. But in the first place no one uses the term 'flexible' any more to refer to Casson handles, and in the second place F can stand for Freedman (or, if you like, for Frank Quinn, who introduced the simplifying technology of gropes in this context).

4. Proof of the 4 dimensional Poincaré Conjecture

Corollary 3.16 shows us that where we want a properly embedded disk, we can find a flexible handle. By Tower Height Raising (Proposition 3.14) where there is one flexible handle there are many: if F is a flexible handle, and $T \subset F$ is a finite subtower with at least two stories, we can grow T inside any neighborhood $N(T)$ into another flexible handle F'.

4.1. **Jigsaw Puzzles.** Both D^4 and F have boundary S^3. We let Λ denote the set of limit points of F. Thus Λ is a wild Cantor set in S^3, and F is the end-compactification of $F - \Lambda$, which is a closed (but not compact) regular neighborhood of a (noncompact) infinite story tower T.

Both D^4 and F have boundary S^3. It is a rather surprising, but very useful fact that there is a surjective map $\pi : D^4 \to F$ which is a homeomorphism on the boundary. In fact, something stronger is true:

Proposition 4.1 (Collapse a wild collar)**.** *There is a wild $S^3 \times [0, 1]$ collar of ∂D^4 and a Cantor set $\Lambda \subset S^3$ so that F is equal to the quotient of D^4 by collapsing each point $\times I$ in the collar to its endpoint.*

Corollary 4.2 (Collared F is standard)**.** *Attach an $S^3 \times [0, 1]$ collar to ∂D^4 and to ∂F to create $\mathbf{D}$ and $\mathbf{F}$. Then π extends by the identity on the collars to a surjective $\pi : \mathbf{D} \to \mathbf{F}$ which is ABH. In particular, $\mathbf{F}$ is homeomorphic to $\mathbf{D}$ which is diffeomorphic to D^4.*

Proof. Shrink the fibers of the original π by pushing down endpoints in the new collar. □

We shall prove Proposition 4.1 in § 4.2 and 4.3. The strategy is to cut up $F - \Lambda$ into *jigsaw pieces* J_i, and then reassemble the pieces inside a standard handle so that the complement is $\Lambda \times [0, 1]$. In a sense to be made precise, this will be achieved by turning handles inside-out: in § 4.2 the description is of an increasing union of 1-handles, whereas in § 4.3 it is of the complement of a nested intersection of (families of) 2-handles.

For the sake of simplicity, we'll assume every grope stage has genus 1, and every cap stage has a single Whitney/accessory pair, although we really won't use this simplifying assumption in any meaningful way. The jigsaw piece associated to a genus 1 grope is indicated by Figure 13 and the jigsaw piece associated to a Whitney/accessory pair is indicated by Figure 12.

4.2. **$F - \Lambda$ as an increasing union of 1-handles.** The *inside view* expresses $F - \Lambda$ as an increasing union of pieces V_i, each obtained from a 4-ball by attaching a collection of 1-handles. The jigsaw pieces are the differences of successive stages in the union.

Let V_0 be a neighborhood of the attaching circle at the base of F, and for each n let V_n denote a neighborhood of the union of the first n stages of the tower (some of these stages are surfaces, some are cleanly immersed caps). Since every V_n has the homotopy type of a graph, it's homeomorphic to a boundary connect sum $\natural_m S^1 \times D^3$s for some m. Likewise $\partial V_n = \sharp_m S^2 \times S^1$ and $\pi_1(V_n)$ is free on m generators.

Each inclusion $V_{n-1} \to V_n$ corresponding to a grope stage is homologically trivial; it corresponds to the map $F_{m(n-1)} \to F_{m(n)}$ sending each generator of the first group to the product of commutators of generators in the second group. However, each inclusion corresponding to a cap stage is trivial on π_1, since the (immersed) Whitney and accessory disks kill the generators of the previous stages.

Let $J_n :=$ closure of $V_n - V_{n-1}$. It arises as a neighborhood of the nth stage of the tower, and is therefore typically disconnected for $n > 1$. Each component is just a union of 1-handles, and therefore homeomorphic to a boundary sum $\natural_i S^1 \times D^3$ for some i (the is can differ from component to component of the same J_n. The inclusion $J_n \to V_n$ induces injections on π_1 of each factor, to distinct conjugacy classes of subsets of the generators of V_n. Figures 12 and 13 show how the core circle of a generator of π_1 of V_n includes into V_{n+1} in a cap stage and a grope stage respectively. The jigsaw pieces are the components of the various J_n. We call the first kind of jigsaw piece a *Whitney piece* and the second kind a *Bing piece*.

4.3. $D^4 - (F - \Lambda)$ **as a nested intersection of 2-handles.** The *outside view* expresses the complement of $F - \Lambda$ in a 4-ball as a nested intersection of pieces Λ_i, each a disjoint union of 2-handles properly contained in the previous ones.

Attaching a 1-handle to D^4 gives the same result as drilling out an unknotted 2-handle. We can think of each V_i stage therefore as D^4 minus a tubular neighborhood of a collection of properly embedded unknotted D^2s; write this collection of 2-handles as Λ_i. Now, each component of $\partial\Lambda_i$ decomposes into $\partial^\pm\Lambda_i$, where each $\partial^+\Lambda_i$ is a solid torus contained in ∂D^4, and each $\partial^-\Lambda_i$ is a (linking) solid torus in the interior.

Each $\Lambda_{i+1} \subset \Lambda_i$ sits inside Λ_i in the following way: in the solid torus $\partial^+\Lambda_i$ there is a finite collection of solid tori $\partial^+\Lambda_{i+1}$, whose cores are either parallel Bing doubles (at a surface stage) or Whitney doubles (as a cap stage) of the core of $\partial^+\Lambda_i$. The components of $\partial^+\Lambda_{i+1}$ get thinner and thinner, as their diameters get smaller and smaller in S^3, but each Λ_{i+1} must crash through the co-core of Λ_i, since the Bing or Whitney cores are essential links in the previous solid tori.

Let J_0' be the difference $D^4 - \Lambda_0$. This is a solid $S^1 \times D^3$ and can evidently be identified with J_0. For each i, let J_{i+1}' be the difference $J_{i+1}' := \Lambda_i - \Lambda_{i+1}$. Each component of J_{i+1}' is a 2-handle from which we have drilled out a collection of 2-handles, the attaching solid tori of which sit inside the attaching solid torus of the big 2-handle as neighborhoods of Bing or Whitney doubles.

The components of $\cap_{i\le n}\Lambda_i$ attach to the boundary S^3 in smaller and smaller circles, limiting to points of Λ. But each 2-handle in Λ_i must extend like a very spiky icecream cone all the way down into the core of Λ_1. Taking a limit, the components of $\cap_i\Lambda_i$ become thinner and thinner, and limit to intervals, and the entire intersection is equal to a product $\Lambda \times [0,1]$. One set of endpoints $\Lambda \times 1$ is the familiar $\Lambda \subset S^3$; the other set lie on a wild Cantor set somewhere in the interior of D^4. This wild $\Lambda \times [0,1]$ lies in a wild collar $S^3 \times [0,1]$; this can be seen by inductively pushing the free boundary components of successive jigsaw pieces deeper and deeper into the interior of D^4 so that their union sits on a wild interior S^3.

This completes the proof of Proposition 4.1 and Corollary 4.2.

Remark 4.3. Proposition 4.1 by itself is not enough to deduce that F is standard.

Let B denote a crumpled cube, and let $F = B \times I$. Then $\partial F = S^3$ by Bing's double Theorem 2.9. We may exhibit this $B \times I$ as a sort of flexible handle made entirely from grope stages. Dually, we may exhibit it as the complement of a nested sequence of 2-handles. This shows how to realize $B \times I$ as a quotient of D^4 by a $\Lambda \times [0,1]$ in a wild collar. But the interior of $B \times I$ is not even simply-connected! This example is one reason why we must insert cap stages infinitely often into our towers.

4.4. **A plugged Design in D^4.** We shall now construct the Design. This is a common subspace of a standard handle and a flexible handle; we denote it $\mathbf{\Gamma}$. First we'll obtain a common subspace $\mathbf{\Gamma}'$ of D^4 and F called a *plugged* Design. Then $\mathbf{\Gamma}$ will be obtained from $\mathbf{\Gamma}'$ by removing countably many disks — the plugs.

What actually is $\mathbf{\Gamma}'$? It's an infinite dyadic rooted tree of flexible handles. Let's denote this tree by $\mathfrak{T}$.

The root and the edges of $\mathfrak{T}$ are neighborhoods of 1-story towers. The tree $\mathfrak{T}$ has a (middle third) Cantor set of ends, one for each infinite embedded path starting at the root. Each such path corresponds to a flexible handle built by assembling the neighborhoods of the 1-story towers corresponding to its edges. Crucially, handles corresponding to different paths have disjoint ends both in F and in $\mathbf{\Gamma}'$.

We shall explain how to find a copy of $\mathbf{\Gamma}'$ in D^4 and in F. The jigsaw pieces in D^4 and in F act like a set of coordinates to ensure that the construction is the same on both sides. As before there is an inside and an outside view of the tree. The inside view as seen from F is to grow the tree, branch by branch and tower by tower. The outside view as seen from D^4 is to carve out the branches like a sculpture from a block of wood.

4.4.1. *Growing $\mathbf{\Gamma}'$ in F.* Where can we find a dyadic tree T of flexible handles in F? Well, we already have one flexible handle, namely F itself. The entire flexible handle F is obtained as a tapered neighborhood in W of a 2-complex — an infinite story convergent tower. To grow new handles inside F we need to make room. Leave the first story neighborhood alone, then push the second and higher stories of the tower slightly into itself so that it attaches to the first story neighborhood on the left side. We can then grow a disjoint tower on the right side by repeated application of Proposition 3.14 and Lemma 3.15. Notice that Lemma 3.15 ensures that the ends of the left and the right flexible handle are disjoint in the original handle. We write the first story of F as B, and the two new flexible handles as F_L and F_R. Notice that $F_L - B$ is a finite union of flexible handles, each of which has as its first story a component of the second story of F_L.

We now have a very stubby dyadic rooted tree. There is a root corresponding to the first story, then two edges corresponding to the two infinite towers we've constructed. We now repeat the process: for each of two towers, we push it slightly into itself so that it attaches on the left side, and grow a disjoint tower on the right side in the space that this opens up. We let B_L and B_R denote the union of B with the second story tower neighborhoods of F_L and F_R respectively, and we denote the flexible handles that begin with B_L (resp. B_R) by B_{LL} and B_{LR} (resp. B_{RL} and B_{RR}).

Repeat the process. Here's the inductive step. We have a dyadic rooted tree of depth n with 2^n leaves, indexed by strings σ of length n in the alphabet $\{L, R\}$. There is a flexible handle F_σ for each σ, and the ends of these flexible handles are all disjoint. If τ is the prefix of σ obtained by removing the last letter, then F_σ is contained in F_τ and shares the bottom $(n-1)$-story tower neighborhoods with it. $F_{\tau L}$ is a copy of F_τ whose nth and higher story neighborhoods are pushed slightly into F_τ, but the combinatorics of the tower stages of $F_{\tau R}$ might differ considerably above the $(n-1)$st story. We write $B_\tau = F_{\tau L} \cap F_{\tau R}$, a neighborhood of the $(n-1)$-story tower at the bottom of F_τ. Now for each F_σ we push F_σ above its nth stories over to the left, and grow a new flexible handle on the caps of the nth story tower over on the right. After countably many steps we've built $\mathfrak{T}$ and $\mathbf{\Gamma}'$.

Notice that every end of $\mathfrak{T}$ corresponds to a Cantor set of ends of $\mathbf{\Gamma}'$, those coming from the associated flexible handle. The map from ends of $\mathfrak{T}$ to Cantor sets in F is continuous in the Hausdorff topology.

4.4.2. *Whittling* $\mathbf{\Gamma}'$ *out of* D^4. The construction of $\mathbf{\Gamma}'$ in D^4 is likewise rather painless. We describe it inductively in terms of the picture of nested 2-handles.

We start with $\Lambda_0 \subset D^4$. The first gap g_0 is a D^4, the co-core of Λ_0. Now, Λ_1 is a union of 2-handles in Λ_0, and g_0 intersects each of these 2-handles in its co-core. So removing g_0 from Λ_1 turns it into a union of $S^1 \times D^3$s. The collection Λ_1 is a union of $S^1 \times D^3$s, but we can proceed component by component, and therefore we temporarily suppress this point for the sake of simplicity. Write (each component of) Λ_1 as a product $S^1 \times [0,1] \times D^2$. Then g_1' is a (union of) $S^1 \times [1/3, 2/3] \times D^2$ sitting inside $S^1 \times D^3$; removing it (them) splits Λ_1 into two copies: an 'inner' and an 'outer' one. Each of these corresponds to the tower (actually, union of towers) corresponding to the components of F_L, F_R minus B. so we construct a $\Lambda_{1,L}$ in the inner Λ_1 and a $\Lambda_{1,R}$ in the outer Λ_1, cut out their intersections with g_0 and g_1', and then let g_2' be the (union of) middle third solid annuli that bisect them radially. And so on.

Notice that the combinatorics of how successive collections of 2-handles $\Lambda_{n,\sigma}$ sit inside $\Lambda_{n-1,\tau}$ depends on the precise combinatorics of $F_\sigma - B_\tau$, which depends on precisely how F_σ was grown in F.

The gap g_0 is a 4-disk, but every component of every g_i' is a solid $S^1 \times D^3$. Now, $\mathbf{\Gamma}'$ is not yet the design, and the components of g_i' are not yet the gaps. Each gap will be obtained from a component of a g_i' by plugging the core with a D^2, making it thereby contractible. We shall find these disk plugs as a subset of $\mathbf{\Gamma}'$ in § 4.6.

4.5. **Pairing up gaps.** What's the complement of $\mathbf{\Gamma}'$ in F? We claim it consists of a collection of gaps h_0 and h_i' in bijection with g_0 and g_i'. This is obvious, since there is one gap in D for every component of the frontier of $\mathbf{\Gamma}'$; the analogous component of the frontier of the $\mathbf{\Gamma}'$ in F therefore bounds its own corresponding gap.

OK, this is kind of unsatisfactory. It would be nice to have a concrete picture of how $\mathbf{\Gamma}'$ embeds in F in such a way that we could *see* the complementary regions, and see how they biject with the g_i'. We can actually achieve this as follows.

Let's think of F as a subset of D. Rather than think of it as $D - \Lambda \times I$ we think of it as being obtained by 'pushing' in the solid tori nesting to Λ to lower and lower levels. Thus we can let $K_1 \subset S^3$ be a solid torus neighborhood of the attaching core $\partial^- J_0$, then successively let $K_n \subset K_{n-1}$ be the solid tori of the attaching handles of $\partial^- J_1$, and so on. Think of D^4 as a product $D^3 \times [-1,1]$, but now reimbed $D^3 \times [0,1]$ in a different way so that it becomes part of the wild collar. Push each $K_n \times 1$ vertically in the wild collar from height 1 down to height $1/n$. Then $\Lambda := \cap_i K_i$ is pushed all the way down to height 0. Let $K^n := \cup_{i \le n} K_i \times [1/i, 1]$. Then $K^\infty := \cup_n K^n$ is the region we push S^3 over to deformation retract D to F. In particular, K^∞ is a *handle gap*.

The gap corresponding to g_0 is obtained from a ball in F by attaching a handle gap K^∞. The gaps corresponding to the solid tori g_1' are obtained from solid tori by attaching partial handle gaps $K^\infty - K^1$, and each successive gap correpsonding to a solid torus in g_n' is obtained from a solid torus by attaching partial handle gaps $K^\infty - \cup_{i \le n} K^n$. These budge up against each other in the way indicated in Figure 16. Notice how you can really see how $\Lambda \times$ Cantor set is stretched out into the interior of F. The complement of the gaps in both cases is $\mathbf{\Gamma}'$.

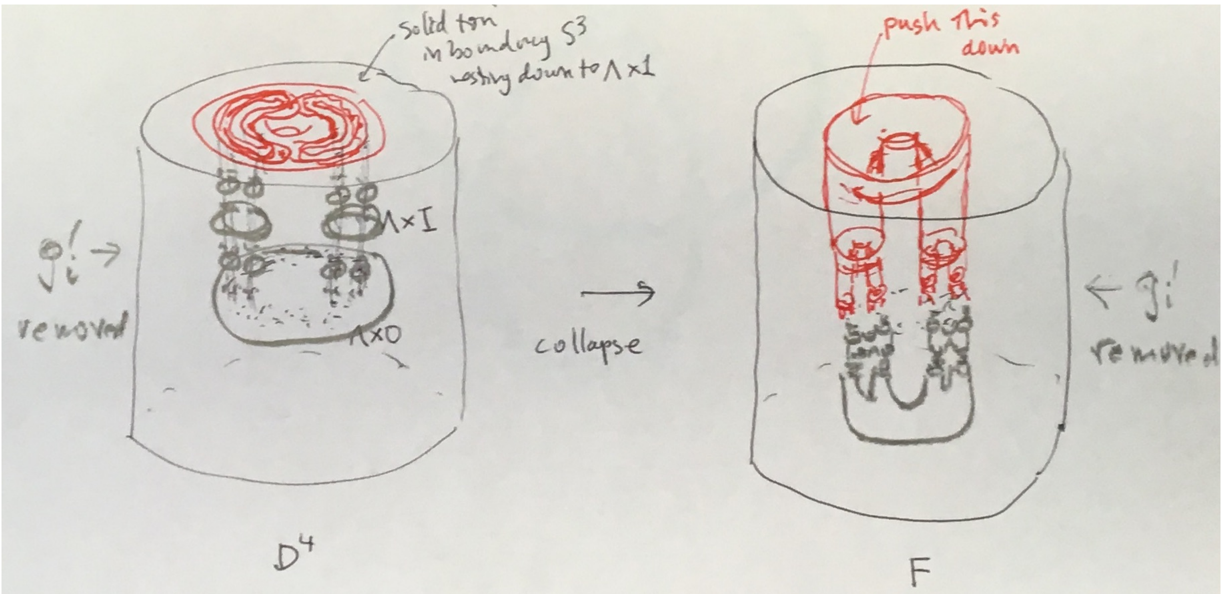


FIGURE 16. The gaps in D^4 and in F.

4.6. **Finding disk plugs.** The Design $\mathbf{\Gamma}$ is obtained from the plugged Design $\mathbf{\Gamma}'$ by removing the plugs. These plugs are a family of D^2s which plug the holes in the $S^1 \times D^3$ gaps. Now, each $S^1 \times D^3$ gap in g_k' intersects some $D^3 \times t$ level where $t \in I_k$ in a solid torus which is a component of K_n. This K_n is a solid torus corresponding to a cap stage in the nth story, so it is the Whitehead double of some parent solid torus. Thus it bounds an obvious immersed disk in the parent torus, which intersects itself in an arc at the clasp and a couple of meridian disks. Push one arc in the positive direction and the other in the negative, so that they end at an interior point level of the vertical Cantor set. At these levels the meridian disks lie in the complement of the gaps, so we have successfully embedded a plug for the solid torus gap in $\mathbf{\Gamma}'$. Remove a countable union of such plugs from $\mathbf{\Gamma}'$ in D^4 and in F. The result is the Design $\mathbf{\Gamma}$.

Remark 4.4. Freedman calls a plugged torus a *red blood cell*. The terminology *platelet* is sometimes also used.

4.7. **The Common Quotient.** The union $\mathbf{\Gamma}$ together with a collar $S^3 \times I$ embeds in both $\mathbf{D}$ and $\mathbf{F}$. The complementary regions are the gaps. Since the gaps correspond, there is a common quotient Q and projections

$$\alpha : \mathbf{D} \to Q \text{ and } \beta : \mathbf{F} \to Q$$

by crushing the gaps on either side.

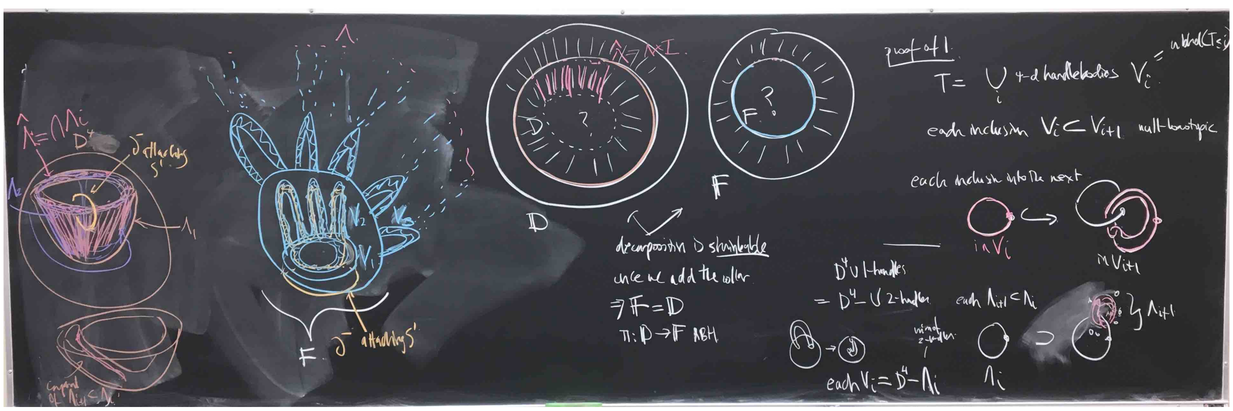
In **D** these gaps consist of countably infinitely many plugged solid tori, together with one D^4. We don't know what these gaps are in **F**, and perhaps we'll never know. Notice that the collar stays embedded on both sides.

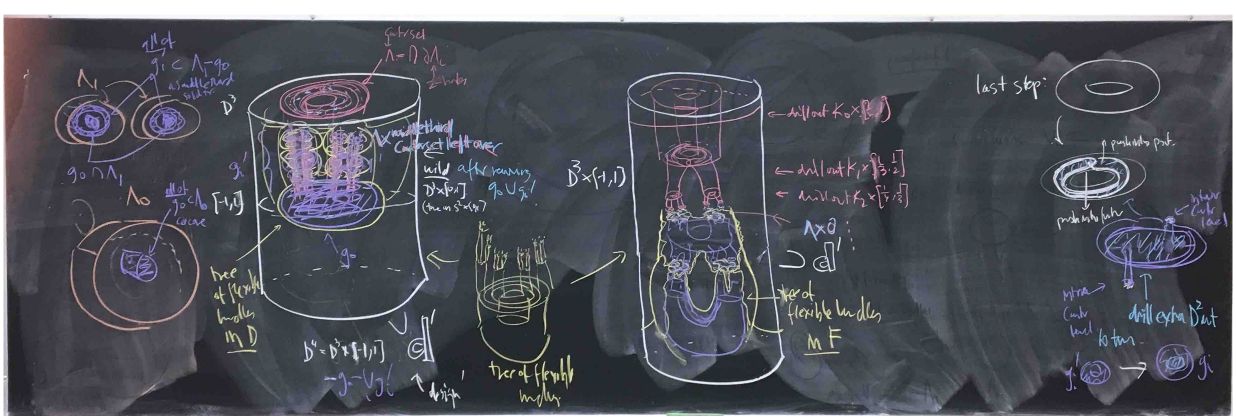
Proposition 4.5. *Both α and β are ABH. Further, the homeomorphisms can be taken to be the identity on the respective collars.*

Proof. Virtually all the work is already done. The decomposition α is shrinkable because it's null and birdlike equivalent of depth 2. The depth 2 pieces are the plugged tori: the plug is a disk and therefore starlike equivalent, and the quotient is obtained by rotating a 2d icecream cone through a torus in SU(2), and is therefore starlike equivalent. Now apply Theorem 2.14 twice. We conclude α is ABH, and therefore Q is homeomorphic to D^4.

It's obvious that the decomposition for β has nowhere dense image: there are open balls in $\mathbf{\Gamma}$ arbitrarily close to each decomposition gap. Since **F** is also a D^4, and since $\beta : \mathbf{F} \to Q$ takes collars to collars by the identity, we can apply the Ball to ball theorem; i.e. Theorem 2.16 and conclude that β is ABH. □

Corollary 4.6 (Flexible handles are standard)**.** *Let F be a flexible handle. Then F is homeomorphic to an ordinary 2-handle $D^2 \times D^2$ relative to a neighborhood of the attaching circle in its boundary.*





4.8. **Poincaré.** With Corollary 4.6 available, we can promote immersed Whitney disks to embedded ones, and we obtain a topological form of the h-Cobordism theorem in dimension 4.

Theorem 4.7 (h-Cobordism)**.** *Suppose M, N are smooth simply-connected 4-manifolds and W is a smooth h-Cobordism between them. Then W is homeomorphic to a product.*

Proof. Start with a regular Morse function on W for which M and N are level sets. As in the proof of Theorem 1.4 we cancel 0 handles with 1 handles, then trade 1 handles for 3 handles. Likewise, we cancel 5 handles with 4 handles then trade 4 handles for 2 handles. Since W is a homology product, the cellular chain map from 3 handles to 2 handles is given by an invertible integer matrix. By handle slides, we can change bases so that this matrix is the identity.

Let X be the middle level of W. Note that X is simply-connected because M is, and W has no 1-handles. The descending manifolds of the 3 handles and the ascending manifolds of the 2 handles give smooth embedded spheres; let α and β be homology classes representing two of these spheres that span a hyperbolic subspace of H_2. By Proposition 3.4 we can find framed immersed spheres A, B representing α and β each with a geometric dual sphere S_A, S_B. Actually, since A and B start out embedded, if we're careful we can arrange for all the spurious intersections to take place in the dual spheres, and keep A and B embedded.

Immersed Whitney disks for pairs of intersections of A with B can be constructed and pushed off A and B into S_A or S_B. Then by Corollary 3.16 and Corollary 4.6 we can find *topological* framed embedded Whitney disks to replace them, and use these to find a new pair of framed embedded spheres A', B' intersecting geometrically in one point. Note we can assume A' and B' are smooth near this point. Since the spheres are geometrically dual, we can cancel a 2–3 handle pair. Once all the handles are cancelled W is revealed as a product. □

Theorem 4.7 is especially useful in view of the following two theorems of Wall:

Theorem 4.8 (Wall; stable diffeomorphism)**.** *Let M and N be simply-connected smooth closed 4-manifolds, and suppose M and N are homotopy equivalent. Then for some k the manifolds $M\#_k S^2 \times S^2$ and $N\#_k S^2 \times S^2$ are diffeomorphic.*

Proof. Without loss of generality we can pass to the case that M and N are connected. Since M and N are homotopy equivalent they have the same signature, and therefore there is *some* smooth cobordism W from M to N. Represent a generator of $\pi_1(W)$ by the core of an embedded $S^1 \times D^4$. Attaching a (6-dimensional) 2-handle to this $S^1 \times D^4$ cuts it out and replaces it with a $D^2 \times S^3$; after finitely many surgeries we obtain a simply-connected cobordism W.

Since M, N and W are simply-connected, we may trade 1-handles for 3-handles and 4-handles for 2-handles. Let X be the 4-manifold intermediate between the 2 and 3-handles. Since M and N are simply-connected, the attaching circles for the 2-handles are trivial in M, and likewise for the 3-handles; thus X is diffeomorphic to $M\#_k S^2 \times S^2$ on one hand, and $N\#_l S^2 \times S^2$ on the other. Since M and N are homotopic, their homology has the same rank, so $k = l$. □

Theorem 4.9 (Wall; h-Cobordism exists)**.** *Let M and N be simply-connected smooth closed 4-manifolds, and suppose M and N are homotopy equivalent. Then M and N are h-Cobordant.*

Proof. We give the sketch of a proof.

Let W be a simply-connected cobordism from M to N with only 2 and 3 handles, and let X be as in the proof of Theorem 4.8. By creating new cancelling 2–3 handle pairs, we may replace X by $X\#_k S^2 \times S^2$ for any k. If $k > 0$ then the intersection form $Q(X)$ is indefinite, and is therefore isomorphic to $n(1) \oplus m(-1)$ or $n(\pm E_8) \oplus mH$ for some positive integers n and m, with the constraint that n is even if $Q(X)$ is even, by Rochlin's Theorem.

Now, by Theorem 4.8 it follows that if we stabilize enough, the middle level becomes diffeomorphic either to $\#_n \mathbb{CP}^2 \#_m \overline{\mathbb{CP}}^2$ or $\#_{n/2} K3 \#_{m-3n/2} S^2 \times S^2$ according to whether $Q(M)$ is odd or even.

We may cut along the middle level and then by an explicit construction, reglue by a diffeomorphism in such a way that the 3 and 2 handles cancel algebraically. The resulting W is an h-Cobordism. ∎

Corollary 4.10. *Let M and N be smooth simply-connected closed 4-manifolds which are homotopy equivalent. Then M and N are homeomorphic. In particular, a smooth homotopy 4-sphere is homeomorphic to S^4.*

Proof. By Theorem 4.9 there is a smooth h-Cobordism between M and N. By Theorem 4.7 this cobordism is a topological product. ∎

4.9. **Poincaré Again.** Corollary 4.10 doesn't yet prove the topological Poincaré Conjecture. What if M is a homotopy 4-sphere which is not smooth, and perhaps admits no smooth structure?

Freedman reasoned as follows. If M is a homotopy 4-sphere, then at the very least $M' := M - p$ is contractible. Therefore by obstruction theory it should admit *some* smooth structure.

We explain this. If X is a topological manifold, the germ of a neighborhood of the diagonal in $X \times X$ is a kind of topological substitute for the tangent bundle, and is classified by a homotopy class of map from X to a space called BTOP. If X is smoothable, then the map $X \to$ BTOP lifts to BO, the classifying space for metrizable vector bundles. The converse is known to hold by the work of Kirby–Siebenmann *either* when the dimension of X is at least 5 *or* when the dimension of X is 4 *but* X is noncompact. Since M' is contractible, there's no obstruction to lifting its classifying map to BO, and since it's noncompact, it's smoothable.

So: M' is a smooth 4-manifold, and one can check it's proper homotopy equivalent to $\mathbb{R}^4$. A careful analysis proves a proper version of Wall's Theorem 4.9 and constructs a smooth proper h-Cobordism W from M' to $\mathbb{R}^4$. Then one needs to show the following:

Theorem 4.11 (Proper h-Cobordism)**.** *Let M, M' be smooth simply-connected 4-manifolds, and let W be a proper smooth h-Cobordism between them. Then W is homeomorphic to a product.*

The proof is basically the same as the proof of Theorem 4.7 except one must be careful, when constructing topologically embedded spheres, that they do not pile up on top of each other going out the end, or else the topological foliation of W by flowlines might be incomplete and fail to exhibit it as a product. One needs to be able to promote an immersed 2-handle to a flexible 2-handle without going too far away. This can be done by proving controlled versions of the various steps in the proof of Corollary 4.6. The details are not especially difficult, and are spelled out explicitly in [9].

Applying Theorem 4.11 we finally conclude:

Theorem 4.12 (Freedman, 4d Poincaré Conjecture)**.** *A closed 4-manifold M homotopic to S^4 is homeomorphic to S^4.*

4.10. **Big exotic $\mathbb{R}^4$s.** In this section we explain how to construct exotic $\mathbb{R}^4$s: smooth structures on the topological manifold $\mathbb{R}^4$ that are not diffeomorphic to the standard structure. These exotic $\mathbb{R}^4$s have the following remarkable feature: they contain compact sets K that are so big that there is no smooth S^3 in this $\mathbb{R}^4$ separating K from infinity (i.e. K can't be smoothly engulfed). These exotic $\mathbb{R}^4$s are 'big' — they don't smoothly embed in ordinary smooth $\mathbb{R}^4$. In the next section we give a 'small' example which does.

Example 4.13 (Kirby). Start with $W = \mathbb{CP}^2 \#_{10} \overline{\mathbb{CP}}^2$. Then $Q(W) = (1) \oplus 10(-1)$. The classification of indefinite forms says that *every* odd indefinite form is $n(1) \oplus m(-1)$ for some positive n and m, so we can also write $Q(W) = -E_8 \oplus (-1) \oplus H$. We know using flexible handles that the hyperbolic H is represented by a pair of embedded framed spheres, intersecting transversely in one point. Let N be a topological neighborhood of this pair of spheres. Then N is a smooth open manifold, homeomorphic to $S^2 \times S^2 - \text{point}$. Now, since open Flexible handles are diffeomorphic to subspaces of smooth handles, we can *smoothly* embed N as a subset of $S^2 \times S^2$, with complement the compact set K. The complement R of the Flexible $S^2 \vee S^2$ in $S^2 \times S^2$ is a smooth open manifold homeomorphic to $\mathbb{R}^4$, and contains K in its interior.

Suppose there were a smooth S^3 in R separating K from infinity. Then because this S^3 lies in N, we can transplant it back to W, and cut it out to produce W'. Then glue back in a smooth D^4 to produce W''. This would be a closed, smooth simply-connected 4-manifold with $Q(W'') = -E_8 \oplus (-1)$ which is negative definite. But Donaldson showed that if M is any simply-connected closed 4-manifold, then if $Q(M)$ is definite it is isomorphic to $n(1)$ or $n(-1)$. This contradiction shows that no smooth S^3 in R exists, so R is an exotic $\mathbb{R}^4$.

The next example is algebraically more involved, but the input from smooth topology is much simpler.

Example 4.14. Start with $W = \mathbb{CP}^2 \#_8 \overline{\mathbb{CP}}^2$. Let α be the homology class in $H_2(W)$ representing the vector $(3, 1, 1, \cdots, 1)$ in the obvious basis. Then $\alpha^2 = 1$ and $\alpha^\perp$ is isomorphic to $-E_8$. One can see this explicitly: take as a basis for $\alpha^\perp$ the vectors

$$\beta_i := (0^i, 1, -1, 0^{7-i}) \text{ for } i \le 7, \quad \beta_8 := (-1^4, 0^4)$$

where superscripts means repeat a vector entry. Evidently all β_j are in $\alpha^\perp$. We compute

$$\beta_i^2 = -2 \text{ for all } i, \quad \beta_i \cdot \beta_{i+1} = 1 \text{ for } i < 7, \quad \beta_3 \cdot \beta_8 = 1$$

and all other $\beta_i \cdot \beta_j = 0$. This is a standard basis for $-E_8$.

Now, α is represented by a smoothly embedded torus. But as is well-known (as the first non-trivial case of the Thom Conjecture), α isn't represented by a smoothly embedded

S^2 with framing 1, or else a neighborhood of this S^2 would be a punctured $\mathbb{CP}^2$, and by cutting it out and gluing in a D^4 we'd get a smooth closed simply-connected manifold with form $-E_8$. This would already contradict Rochlin's Theorem.

On the other hand, we know α is represented by a topologically embedded sphere S with framing 1, and a smooth neighborhood N of this sphere smoothly embeds in $\mathbb{CP}^2$. The complement of S in $\mathbb{CP}^2$ is a smooth manifold R, homeomorphic to $\mathbb{R}^4$, and containing N as the complement of a compact set K. As in the previous example, if the smooth structure on R were standard, we could separate K from S with a smooth S^3, and get a contradiction as in the previous paragraph.

4.11. **Small exotic $\mathbb{R}^4$s.** If U is any open subset of standard $\mathbb{R}^4$, and K is any compact subset of U, there is a smooth sphere in U separating K from infinity. Thus, the exotic $\mathbb{R}^4$s constructed in § 4.10 do not smoothly embed in standard $\mathbb{R}^4$.

Example 4.15 (Casson–Freedman). The input to this example is a compact counterexample to the smooth 5-dimensional h-cobordism theorem, due to Donaldson [5]. He shows that the two manifolds $\mathbb{CP}^2\#_9\overline{\mathbb{CP}}^2$ and the 'Dolgachev surface' $L(2,3)$ are not diffeomorphic. Nevertheless, they are homotopy equivalent, and therefore smoothly cobordant by Wall.

Let's describe these manifolds. Take two general cubics $f=0$ and $g=0$ in $\mathbb{CP}^2$ and blow up the nine points of intersection. This is $\mathbb{CP}^2\#_9\overline{\mathbb{CP}}^2$, and it comes with a map to $\mathbb{CP}^1$ defined by f/g. The fibers are cubic curves $f-\lambda g=0$, exhibiting $\mathbb{CP}^2\#_9\overline{\mathbb{CP}}^2$ as an elliptic surface. The Dolgachev surface $L(2,3)$ is obtained by *logarithmic transforms* on two smooth fibers with multiplicities 2 and 3. This is a 4-dimensional operation, rather analogous to a Dehn surgery in 3-manifold topology.

A cobordism Y between these two manifolds exists by Wall. Since the manifolds are simply-connected, we can arrange for all critical points to have index 2 or 3. In fact, it turns out Y can be obtained by attaching exactly one 2-handle and exactly one (algebraically) cancelling 3-handle to L. The middle level is therefore diffeomorphic both to $L\#S^2\times S^2$ and (turning handles upside down) to $\#_2\mathbb{CP}^2\#_{10}\overline{\mathbb{CP}}^2$.

In this middle level X the core and the co-core of the two handles are a pair of smooth embedded two-spheres A and B, each with self-intersection number zero, which intersect each other algebraically once, and geometrically $2k+1$ times for some positive k. We can make $X-A-B$ simply-connected at the cost of increasing the number of intersections of A with B. Thus the linking tori at the points of intersection are dual to a collection of embedded Flexible handles, each spanning a bigon between a successive pair of intersection points. Let W be an open neighborhood of $A\cup B$ together with these Flexible handles. Since the Flexible handles are topologically standard, W is homeomorphic to an open neighborhood of $S^2\vee S^2$ in $S^2\times S^2$. Let Z be the part of the cobordism intersecting the middle level in W; i.e. the union of the smooth gradient flowlines, together with the critical levels. Since Z is obtained topologically by attaching 3-handles to the S^2s in $S^2\times S^2$, it's homeomorphic to a product $\mathbb{R}^4\times I$ with $Z_0:=\mathbb{R}^4\times 0\subset L$ and $Z_1:=\mathbb{R}^4\times 1\subset\mathbb{CP}^2\#_9\overline{\mathbb{CP}}^2$.

It turns out that Z smoothly embeds in $S^4\times I$. To see this, build a Morse function on $S^4\times I$ with one cancelling pair of 2-3 handles, then perturb the core/co-core spheres in the middle level so they intersect combinatorially in the same was as A and B, and span the bigons with smooth Whitney disks. Now use the fact that Flexible handles smoothly embed in standard smooth handles. We now have two smooth manifolds Z_0, Z_1 homeomorphic to $\mathbb{R}^4$ and embedded in smooth S^4 and therefore also in smooth $\mathbb{R}^4$. We show that these are (small) exotic $\mathbb{R}^4$s.

Suppose Z_0 is standard. Then for a big enough smooth compact ball $B_0 \subset Z_0$ the gradient flow is nonsingular on $S^3_0 := \partial B_0$ and gives us a smooth embedded $S^3_1 \subset Z_1 \subset \mathbb{CP}^2 \#_9 \overline{\mathbb{CP}}^2$ bounding a topological ball. The ball B_0 is a smooth standard ball. Every smooth standard ball in S^4 has a smooth standard ball as complement, since we can shrink B radially down to a standard ball in a chart. Therefore $S^4_0 - B_0$ is a smooth standard ball, and by the gradient flow we see that the region outside S^3_1 is a smooth standard ball in S^4_1. Thus the region inside S^3_1 is a smooth standard ball B_1 in Z_1, and the cobordism Y gives a diffeomorphism between $L - B_0$ and $\mathbb{CP}^2 \#_9 \overline{\mathbb{CP}}^2 - B_1$. By Hatcher's proof of the Smale Conjecture, every orientation-preserving diffeomorphism of S^3 is smoothly isotopic to the identity, so we get a diffeomorphism between L and $\mathbb{CP}^2 \#_9 \overline{\mathbb{CP}}^2$, contrary to Donaldson's theorem. This contradiction shows that Z_0 is exotic.

4.12. **Homology 3-spheres bound contractible 4-manifolds.** We have seen that closed simply-connected smooth 4-manifolds are classified up to homeomorphism by their intersection forms Q. By now many restrictions are known on which forms can be realized by smooth manifolds. However in the topological world, we have:

Theorem 4.16 (Freedman, realization of forms)**.** *Let Q be a nondegenerate integral symmetric bilinear form. Then there is a closed simply-connected 4-manifold W with $Q(W) = Q$.*

This follows from:

Theorem 4.17 (Freedman, homology 3-spheres bound)**.** *Every homology 3-sphere Σ bounds a contractible 4-manifold V.*

We give the proof of Theorem 4.16 assuming Theorem 4.17.

Proof. Suppose Q has rank n and let α_i be a basis for $\mathbb{Z}^n$. Build a compact simply-connected 4-manifold W' by attaching 2-handles to knots K_i in $S^3 = \partial D^4$ in such a way that the framing of K_i is equal to $Q(\alpha_i, \alpha_i)$ and the linking number of K_i and K_j is $Q(\alpha_i, \alpha_j)$. The same diagram represents a Dehn surgery recipe for a closed 3-manifold Σ that arises as the boundary of W'. Each 2-handle can be coned to the center of D to form a 2-sphere in W' representing an α_i, and $H_2(W') = \mathbb{Z}^n$ in such a way that the intersection pairing on $H_2(W')$ is equal to Q. Now for a 4-manifold with boundary, Lefschetz duality gives a perfect pairing between $H_2(W')$ and $H_2(W', \partial W')$, so it must be that $H_2(W') \to H_2(W', \partial W')$ is surjective. Since $\pi_1(W') = 1$ we must have $H_1(\partial W') = H_1(\Sigma) = 0$, and by Poincaré duality, Σ is a homology 3-sphere.

By Theorem 4.17 there is a contractible V with $\partial V = \Sigma$. Gluing V to W' along Σ gives W with $Q(W) = Q$. □

The proof of Theorem 4.17 is reasonably straightforward, modulo two things: the existence of an embedded $S^2 \times S^2$ representing a hyperbolic pair, and a technical result due to Kirby about isolated non-locally flat points in codimension one. We give a self-contained proof of Kirby's result in Appendix A.

We now begin the proof of Theorem 4.17

Proof. The first step is to build a simply-connected homology cobordism W from Σ to itself. This is a compact simply-connected 4-manifold W for which the inclusion of Σ into W is a homology isomorphism.

Since Σ is 3 dimensional, it's smooth, and so is $\Sigma \times I$. Let γ_i be a collection of disjoint knots in Σ generating π_1, and push them into the interior. Since Σ is a homology sphere, it makes sense to choose longitudinal framings of the γ_i corresponding to the boundaries of Seifert surfaces R_i'; push these (disjointly) into the interior too. The R_i' are framed in Σ, and therefore also in $\Sigma \times I$. Attach (5-dimensional) 2-handles to $\Sigma \times I$ along the framed knots γ_i to surger $\Sigma \times I$ to V. This drills out $S^1 \times D^3$s and replaces them with $D^2 \times S^2$s. Each D^2 caps R_i' to produce a closed framed surface R_i, which intersects point$\times S^2$ transversely in exactly one point. Thus, $H_2(V)$ is a direct sum of hyperbolics generated by these R_i, S^2 pairs. These can be replaced by disjoint *topological* framed embedded S^2s meeting transversely in a single point. The boundary of a neighborhood of one of these pairs is an S^3, and if we cut out the punctured $S^2 \times S^2$ it bounds and glue in D^4 we'll get a new simply-connected W with the same homology as $\Sigma \times I$.

Now, we can glue $\mathbb{Z}$ copies of W end to end to produce W'. This 4-manifold is simply-connected and has the homology of S^3 so it is homotopic to $S^3 \times \mathbb{R}$. Furthermore, it is simply-connected at either end. Drilling out a 'horizontal' $\mathbb{R}$ from the original product structure connects up these ends, and produces a *contractible* 4-manifold W'' in which $\Sigma -$ point is properly embedded. We've already seen that a contractible 4-manifold is smoothable, and smoothly properly h-Cobordant to $\mathbb{R}^4$ (Theorem 4.11). Consequently W'' is homeomorphic to $\mathbb{R}^4$, and the 1-point compactification topologically embeds Σ in S^4. Now, this Σ is locally flat everywhere except possibly at the new point. But Theorem A.1 shows that the set at which a codimension one manifold is not locally flat has no isolated points, at least if the ambient manifold has dimension at least 4. So Σ is tamely embedded after all, and splits S^4 into two compact 4-manifolds.

Observe that the construction above can be done perfectly symmetrically: in place of W' we could take $\mathbb{N}$ copies of W glued end to end and then double along Σ. So there is an involution of S^4 that has Σ as a (tame) fixed point set! In particular both sides of Σ are homeomorphic to the same compact 4-manifold V. Now, Seifert van-Kampen implies that V is simply-connected because S^4 is, and Meyer–Vietoris implies that $H_2(V) = 0$. Thus V is homotopic to a point, and the theorem is proved. □

Example 4.18. There is a closed simply-connected 4-manifold W with $Q = E_8$. This is even, so if W were smoothable, it would be spin, and therefore by Rochlin's Theorem the signature would be divisible by 16. This is a contradiction, and shows that W is not smoothable.

5. Acknowledgments

I'd like to thank the NSF, who supported me through grant DMS 1405466. I'd also like to thank Mark Powell and Arunima Ray, who generously shared a copy of a beta version of their book [2] and pointed out a couple of errors in an earlier version of this manuscript, especially a serious error in the proof of Lemma 3.13.

Finally, I'd like to thank Mike Freedman, who taught me many fragments of the theory of flexible handles over the years without ever letting on that they were related to 4-manifold topology. I'd like to thank him for that, and for some very helpful emails while I was drafting this material. But most of all I'd like to thank him for the creation of such achingly beautiful mathematics, which is a gift to all the world forever.

Appendix A. Bad points in codimension 1

The purpose of this appendix is to give a self-contained proof of the following theorem of Kirby:

Theorem A.1 (Kirby, No isolated bad points)**.** *Let N be a manifold of dimension $n \geq 4$ and let Σ be a manifold of dimension $n-1$ topologically embedded in N. Then the set of points where Σ is not locally tame has no isolated points.*

Codimension 1 is necessary of course, since the cone in D^4 on a nontrivial knot in S^3 gives a properly embedded D^2 that fails to be tame at exactly one point.

The top level proof of Theorem A.1 is quite short and elegant, but along the way we will need to appeal, recursively, to theorems of Letscher, Letscher–Cantrell, Cantrell, and Cantrell–Edwards. It's only at the bottom of the well that the hypothesis $n \geq 4$ is used, in the following form: there is no arc in $\mathbb{R}^n$ wild only at its endpoints. The Fox–Artin arc is a counterexample to this claim if $n = 3$, and this counterexample percolates all the way back up to Kirby's Theorem: the boundary of a tapered neighborhood of the Fox–Artin arc is an embedded S^2 in $\mathbb{R}^3$ wild at exactly two points.

Let's assume for the moment the following theorem of Cantrell:

Theorem A.2 (Cantrell, No bad sphere)**.** *Let $n \geq 4$ and let a topological embedding $S^{n-1} \to S^n$ be locally tame except possibly at one point. Then it's tame at this point too, and consequently is isotopic to a round S^{n-1}.*

We defer the proof to § A.2

A.1. **Proof of Theorem A.1.** In this section we give Kirby's proof of his theorem.

Proof. Suppose $f : \Sigma \to N$ is an embedding, and $p \in \Sigma$ is an isolated point where $f(\Sigma)$ is not locally tame. Then p is in the interior of a $D^{n-1} \subset N$ so that p is the only point where $f(D^{n-1})$ is not locally tame. Now cut D^{n-1} down the middle into two disks $D_1 \cup D_2$, each a standard D^{n-1}, each containing p in the boundary.

Write $E_i = f(D_i)$ and $q = f(p)$, and restrict attention to E_1 (say), dropping the subscript. We can assume E is contained in the interior of a ball B^n, and then includer that ball into S^n. So without loss of generality, we can assume $E \subset S^n$. Now, E is locally flat in S^n except possibly at $q \in \partial E$, and therefore $F := E - q$ is locally flat in $S^n - q = \mathbb{R}^n$. The space F is homeomorphic to an $(n-1)$-ball minus a point in its boundary; this is an $(n-1)$-half space H^{n-1}. We claim that *any* locally flat proper H^{n-1} in $\mathbb{R}^n$ is standard; i.e. there is a homeomorphism of pairs $(\mathbb{R}^n, F) \to (\mathbb{R}^n, H^{n-1})$.

Since F is locally flat it's at least locally standard, so we can engulf more and more of it gradually in a nice regular open neighborhood U so that U is homeomorphic to $\mathbb{R}^n$, and $(U, F) \to (\mathbb{R}^n, H^{n-1})$. Further, we can find a slightly smaller closed neighborhood $V \subset U$ so that ∂V is homeomorphic to $\mathbb{R}^{n-1}$ and is closed in U and in $\mathbb{R}^n$. Any locally flat proper $\mathbb{R}^{n-1}$

in $\mathbb{R}^n$ is compactified to S^{n-1} in S^n locally flat except perhaps at one point, so Theorem A.2 implies that $(U, \partial V) \to (\mathbb{R}^n, \mathbb{R}^{n-1})$ and at the same time, $(\mathbb{R}^n, \partial V) \to (\mathbb{R}^n, \mathbb{R}^{n-1})$. Thus ∂V — and therefore U and therefore F — are unknotted in $\mathbb{R}^n$, proving the claim. This part of the proof, modulo Theorem A.2, is due to Letscher.

We now know that each of D_1, D_2 is individually locally flat in $\mathbb{R}^n$, and furthermore the disk $\partial D_1 \cap \partial D_2$ is locally flat in each of ∂D_1, ∂D_2. We shall show that this implies D is locally flat. Of course, the only point at which this is currently in contention is p.

Think of D as a diameter of $B := D^n \subset \mathbb{R}^n$ and introduce radial coordinates in a 2-dimensional subspace transverse to D, so that D_1 and D_2 are the codimension 1 disks at angles π and 0 respectively. To be consistent with Kirby we relabel them D_π and D_0. For any $\alpha \neq \beta$ let $W(\alpha, \beta)$ be the *positive* wedge from angle α to angle β.

Choose extensions of the f_i so that

(1) $f_2(B) \subset f_1(\mathbb{R}^n)$; and
(2) the interior of $f_2(D_0)$ is disjoint from $f_1(W(\pi/2, 3\pi/2))$.

Define $f := f_1^{-1} f_2$ (this has no relation to the previous map denoted f). By hypothesis (2), the disk $f(D_0)$ is contained in the wedge $W(3\pi/2, \pi/2)$.

(argument continues)

□

A.2. **Proof of Theorem A.2.**

Proof. □

Appendix B. Quadratic forms and Rochlin's Theorem

Theorem B.1 (Rochlin). *Let W be a smooth closed simply-connected 4-manifold with even intersection form. Then the signature of W is divisible by 16.*

References

[1] S. Akbulut and R. Kirby, *An exotic involution of S^4*, Topology **18** (1979), 75–81
[2] S. Behrens, B. Kalmár, M.-H. Kim, M. Powell and A. Ray (eds.), *The disk embedding theorem*, Oxford University Press, to appear
[3] A. Casson, based on notes of C. Gordon *Three lectures on new infinite constructions in 4-dimensional manifolds*, 201–244, À la recherche de la topologie perdue. Prog. Math. **62**, Birkhäuser Boston MA, 1986
[4] R. Denman and M. Starbird, *Shrinking countable decompositions of S^3*, Trans. AMS **276** (1983), no. 2, 743–756
[5] S. Donaldson, *Irrationality and the h-cobordism conjecture*, J. Diff. Geom. **26** (1987), no. 1, 141–168
[6] M. Freedman, *A fake $S^3 \times \mathbb{R}$*, Ann. Math. (2) **110** (1979), no. 2, 177–201
[7] M. Freedman, *The topology of four-dimensional manifolds*, J. Diff. Geom. **17** (1982), no. 3, 357–453
[8] M. Freedman, *Bing topology and Casson handles*, 2013 Santa Barbara/Bonn Lectures, available online at `https://www.mpim-bonn.mpg.de/node/4436`.
[9] M. Freedman and F. Quinn, *Topology of 4-Manifolds*, Princeton University Press Princeton NJ, 1990
[10] R. Gompf and A. Stipcisz, *4-Manifolds and Kirby Calculus*, AMS Grad. Stud. Math. **20**, AMS Providence, 1999
[11] R. Kirby, *The Topology of 4-Manifolds*, LNM **1374**, Springer Heidelberg, 1989
[12] J. Milnor, *On manifolds homeomorphic to the 7-sphere*, Ann. Math. (2) **64** (1956), 399–405

[13] J. Milnor, *Notes on the h-cobordism theorem*, Notes by L. Siebenmann and J. Sondow; Princeton University Press Princeton, 1965
[14] S. Smale, *Generalized Poincaré's conjecture in dimensions greater than four*, Ann. Math. (2) **74** (1961), 291–406
[15] J. Stallings, *The piecewise-linear structure of Euclidean space*, Proc. Cambridge Philos. Soc. **58** (1962), 481–488

University of Chicago, Chicago, Ill 60637 USA
E-mail address: dannyc@math.uchicago.edu