NOTES ON 3-MANIFOLDS

DANNY CALEGARI

Abstract. These are notes on 3-manifolds, with an emphasis on the combinatorial theory of immersed and embedded surfaces. These notes follow a course given at the University of Chicago in Winter 2014.

Contents

1. Dehn's Lemma and the Loop Theorem

The purpose of this section is to give proofs of the following theorem and its corollary:

Theorem 1.1 (Loop Theorem). *Let M be a 3-manifold, and B a compact surface contained in ∂M. Suppose that N is a normal subgroup of $\pi_1(B)$, and suppose that N does not contain the kernel of $\pi_1(B) \to \pi_1(M)$. Then there is an* embedding *$f : (D^2, S^1) \to (M, B)$ such that $f : S^1 \to B$ represents a conjugacy class in $\pi_1(B)$ which is not in N.*

Corollary 1.2 (Dehn's Lemma). *Let M be a 3-manifold, and let $f : (D, S^1) \to (M, \partial M)$ be an embedding on some collar neighborhood A of ∂D. Then there is an* embedding *$f' : D \to M$ such that f' and f agree on A.*

Proof. Take B to be a collar neighborhood in ∂M of $f(\partial D)$, and take N to be the trivial subgroup of $\pi_1(B)$. The Loop Theorem produces an embedding $f' : (D, S^1) \to (M, B)$ whose boundary is nontrivial in $\pi_1(B)$. The only embedded essential loop in an annulus is its core, so the boundary is isotopic to $f(S^1)$ in B, and after an isotopy (and change of orientation if necessary) we can assume that $f' = f$ on a collar neighborhood of S^1. □

Dehn's Lemma was first stated by Dehn in 1910 [7], with an erroneous proof (the error was not detected until 1929 by Kneser), and the first rigorous proof was given by Papakyriakopoulos in 1957 [23]. The version of the Loop Theorem stated above was first

Date: November 18, 2014.

formulated by Stallings, and we give Stallings' proof [38], using Papakyriakopoulos' *tower construction*.

The proof of these, and many other theorems in 3-manifold topology, depend on combinatorial arguments; in the smooth category, such arguments depend on first putting a surface (or some other object) into *general position*; in the PL category, such arguments depend on *simplicial approximation*. To apply either tool one must first know that the domain and target have (unique) smooth or PL structures; for 2-manifolds this fact is due to Radó, and for 3-manifolds it is due to Moise [20].

Before beginning the proof of the Loop Theorem we make a few remarks about general position and simplicial approximation, especially as it applies to maps from surfaces to 3-manifolds.

1.1. **General position.** It is often very convenient in low dimensional topology to replace an arbitrary map between manifolds with a nearby (i.e. C^0 close) smooth map that is as "generic as possible". We are generally interested in one of the following two cases:

(1) a map $\gamma : L \to S$ from a 1-manifold L to a surface S
(2) a proper map of pairs $\Gamma : S, \partial S \to M, \partial M$ from a surface S to a 3-manifold M

It is straightforward to show that any map from a 1-manifold L to a surface S may be C^0 approximated by a smooth map in *general position* — i.e. for which the map is an immersion with finitely many isolated double points, and no other singularities. What about maps from surfaces to 3-manifolds? Any immersion $S \to M$ can be perturbed to be in general position — i.e. with finitely many transverse double curves, meeting in finitely many triple points, and no other singularities. But when can an arbitrary map of pairs be approximated by an immersion? And if the map is already an immersion on the boundary, when can this immersion of the boundary be extended to an immersion of the interior? It turns out one more kind of singularity is necessary.

Example 1.3 (Order 2 branch point). Let $f : D^2 \to D^2 \times [-1, 1]$ be given in polar coordinates by $(r, \theta) \to (r, 2\theta, r\sin(\theta))$. The map is an immersion away from the point $0 \in D^2$, and an embedding away from a proper arc $J' \subset D^2$ with 0 as its midpoint, which is folded at 0 onto its image arc J. See Figure 1.

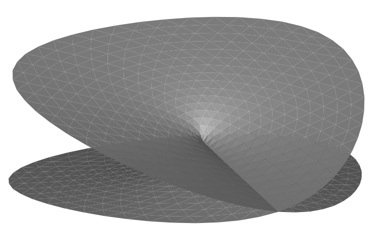


FIGURE 1. An order 2 branch point joined by an arc of singularity to a transverse double point of the boundary circle.

If $f : (S, \partial S) \to (M, \partial M)$ is a proper immersion in general position, let L denote the singular subset of $f(S)$ and L' its preimage in S. Then L' is an immersed 1-manifold, whose components map in pairs to the double curves and double arcs of L. Each arc of L' has 2 endpoints on ∂S, and since arcs of L' map to L in pairs, the number of singular points on ∂S must be divisible by 4. These points map in pairs to double points of $f(\partial S)$; hence we deduce that $f(\partial S)$ must have an *even* number of double points.

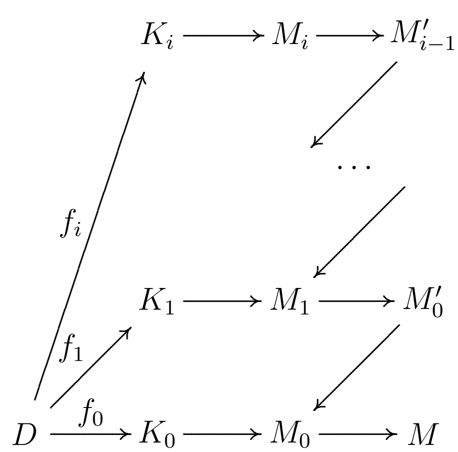
In general, any proper map f restricting to a generic immersion $f : \partial S \to \partial M$ may be C^0 approximated by a map with finitely many transverse double curves meeting in finitely many triple points, and order 2 branch points; and (as we have just shown) the number of order 2 branch points has the same parity as the number of double points of $f(\partial S)$.

1.2. **Simplicial approximation.** If K and L are compact simplicial complexes and $f : K \to L$ is any map, then after subdividing K sufficiently many times, the map f may be replaced by a homotopic simplicial map. For, if we subdivide K sufficiently many times, each simplex of K is taken entirely into the star of some simplex of L, and then we may just take each vertex v of K to a vertex of L closest to $f(v)$, and extend the result linearly.

If L is a polyhedron in a 3-manifold M, a *regular neighborhood* of L is a closed neighborhood N of L, homeomorphic to a 3-manifold (usually with boundary unless L is empty or all of M), which contains L as a "spine" to which N deformation retracts. For example, we may take N to be the union of the stars of simplices in L in a barycentric subdivision. The manifold N is well-defined up to homeomorphism and up to isotopy in M.

1.3. **Tower construction.** We now begin the proof of the Loop Theorem. By hypothesis we start with a proper map $f : (D, S^1) \to (M, B)$ for which the conjugacy class of $f(S^1)$ in $\pi_1(B)$ is not contained in N. We then replace f with a simplicial map f_0, and let M_0 denote a closed regular neighborhood of the image K_0. We also let B_0 denote $B \cap M_0$, so that B is a surface in ∂M_0, and we let N_0 be the normal subgroup of $\pi_1(B_0)$ mapping to $\pi_1(B)$. Note that the conjugacy class of $f_0(S^1)$ in $\pi_1(B_0)$ is not contained in N_0.

Suppose that M_0 admits a nontrivial double cover (equivalently: a connected double cover) M_0'. We can lift f_0 to $f_1 : (D, S^1) \to (M_0', B_0')$. Let K_0' denote the preimage of K_0 in M_0', and let $K_1 = f_1(D^2) \subset K_0'$. Then define M_1 to be a closed regular neighborhood of K_1. We repeat this construction inductively, whenever M_i admits a nontrivial double cover, thus obtaining a *tower* of maps:



The first step is to prove that the tower is finite, and that every component of ∂M_i is a sphere.

Lemma 1.4. *The tower is finite. That is, there is some finite i for which M_i does not admit a nontrivial double cover.*

Proof. Define the complexity of f_i to be the number of simplices of D minus the number of simplices of K_i. This complexity is non-negative, and by construction it must decrease under each covering step. □

Lemma 1.5. *Suppose some component of ∂M_i is not a sphere. Then M_i admits a nontrivial double cover.*

Proof. Double covers are parameterized by $H^1(M_i;\mathbb{Z}/2\mathbb{Z})$. If this group vanishes, so must $H_1(M_i;\mathbb{Z}/2\mathbb{Z})$ and $H_2(M_i,\partial M_i;\mathbb{Z}/2\mathbb{Z})$ (by Poincaré duality). But in the long exact sequence of homology we have

$$H_2(M_i,\partial M_i;\mathbb{Z}/2\mathbb{Z}) \to H_1(\partial M_i;\mathbb{Z}/2\mathbb{Z}) \to H_1(M_i;\mathbb{Z}/2\mathbb{Z})$$

so that $H_1(\partial M_i;\mathbb{Z}/2\mathbb{Z}) = 0$ and therefore every component of ∂M_i is a sphere. □

Now, at every stage of the tower (including the last stage) the conjugacy class of $f_i(S^1)$ in $\pi_1(B_i)$ is not contained in N_i. But B_i is a planar surface, and therefore its fundamental group is normally generated by boundary loops. It follows that some boundary component of B_i is not contained in N_i. This boundary component is an embedded loop bounding an embedded disk in the corresponding sphere component of ∂M_i. So we obtain an embedding $g_i : (D, S^1) \to (M_i, B_i)$ with $g_i(S^1)$ not contained in N_i.

The crux of the matter is to investigate the composition of g_i with the covering projection from $M_i \subset M'_{i-1}$ to M_{i-1}; call this composition $g'_i : (D, S^1) \to (M_{i-1}, B_{i-1})$. Since the covering has degree 2, after perturbing it to be in general position, we may assume that g'_i is an immersion whose only singularities are arcs and loops of double points; in other words there are *no triple points*.

We claim that we can modify the map g'_i to replace it by an embedding. This is done by inductively simplifying g'_i by surgering it along double arcs or double circles. First we consider a double curve J. The preimage of this double curve in D is J', which is either a single curve mapping to J by a degree 2 map, or two curves mapping to J by degree 1 maps. In the latter case, the two components of J' might bound disjoint interiors, or be nested. See Figure 2.

In the first case J' is a single loop. We cut along this loop, twist the boundary by angle π, and then reglue. The result is a new map of a disk with the same image, except that now we can perturb the map to eliminate the loop J'.

In the second or third cases, assume that one of the disks is innermost. Cut the other disk out and replace it with a copy of the innermost disk. In the third case this eliminates at least two circles of intersection. In the second case, we can perturb the result and eliminate at least two circles of intersection. So after finitely many moves of this kind we can eliminate every double circle.

Next we consider a double arc J. The preimage of this arc in D is J', a pair of embedded, disjoint proper arcs. There are two possibilities, depicted in Figure 3.

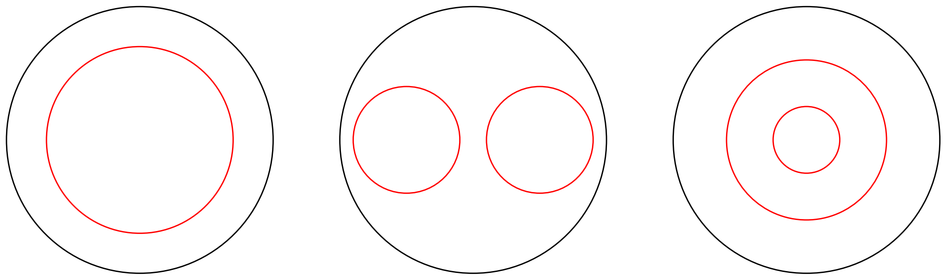


FIGURE 2. Three possible configurations of double curves on D with the same image

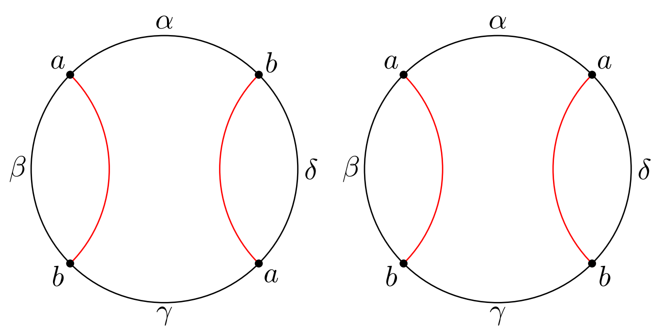


FIGURE 3. Two possible configurations of double arcs on D with the same image

As before, assume one of the lunes is innermost. We form two disks: one obtained by gluing the two lunes together, and the other by replacing one lune by the innermost one and pushing them apart. In the notation implied by the figure, the homotopy classes of these boundary loops are $\alpha\beta\gamma\beta^{\pm}$ and $\beta^{\mp}\delta$ (without loss of generality). Since their product $\alpha\beta\gamma\delta$ is not in N_{i-1} (by hypothesis), at least one of these simpler loops is not in N_{i-1}.

Thus after finitely many simplifications, we obtain an *embedding* $g_{i-1} : (D, S^1) \to (M_{i-1}, B_{i-1})$ with $g_{i-1}(S^1)$ not contained in N_{i-1}. Induct until $i = 0$, in this way obtaining an embedded disk and proving the Loop Theorem and Dehn's Lemma.

1.4. **Incompressible surfaces.** One crucial application of the Loop Theorem is the following proposition, called Kneser's Lemma:

Lemma 1.6 (Kneser's Lemma)**.** *Let S be a 2-sided embedded surface in a 3-manifold M. If S is not π_1-injective, there is some properly embedded disk (D, S^1) in $(M - S, \partial(M - S))$ whose boundary is essential in S.*

Proof. Let γ be a loop in S which is essential in S, and inessential in M. Then γ bounds a disk $f : D \to M$. Since S is 2-sided, γ can be pushed off S on one side; since S is embedded, we can perturb f so that the intersection of the interior of D with S is a system of disjoint simple curves in D. If an innermost curve has boundary inessential in S, we can swap it for a disk in S and push it off, eliminating a curve of intersection. Thus eventually we find a disk D whose interior maps to $M - S$ and whose boundary is essential in S. Cut open M along S and apply the Loop Theorem. □

Let S be a surface satisfying the hypothesis of Kneser's Lemma, and let D be an embedded disk as promised by the conclusion. The surface S can be *compressed* along D, cutting S open along ∂D and gluing in two disjoint parallel copies of D. The result is a new embedded surface S' (possibly disconnected even if S was), each of whose components has smaller genus than S. Informally, one calls a surface *incompressible* if it can't be compressed. It is convenient to extend this definition to spheres, for which the analog of a compression is to simply throw away a sphere that bounds a 3-ball:

Definition 1.7. A connected embedded surface S in a 3-manifold is *incompressible* if one of the following conditions holds:

(1) S is a sphere which does not bound a ball; or
(2) S is not a sphere, and no essential simple closed curve on S bounds an embedded disk in $M - S$.

Kneser's Lemma says that a 2-sided incompressible (embedded) surface is π_1-injective. As we have seen, 2-sidedness is essential for the proof, since otherwise we can't arrange for D to have a collar neighborhood of the boundary disjoint from S.

Example 1.8 (Lens space). For coprime (p, q) with $0 \leq q < p$ let P be a regular p-gon, and let $Q(p)$ be the polyhedron which is the join of P with an interval. Then $Q(p)$ is topologically a 3-ball, an looks like a polyhedral "lens", with p triangles on the "top" side and p triangles on the "bottom". See Figure 4 for an example with $p = 6$.

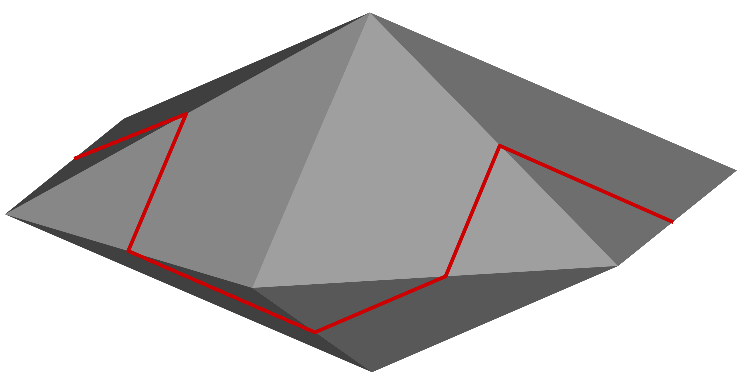


FIGURE 4. A polyhedral lens, obtained as the join of a 6-gon with an interval

The Lens space $L(p, q)$ is obtained from $Q(p)$ by gluing each triangle on the top to the triangle on the bottom after rotating q units about the central axis. The peripheral circle P covers a loop in $L(p, q)$ with degree p, and the interval I closes up to form a complementary loop; the lens space is the union of two solid torus neighborhoods of these loops, glued up along their boundary in such a way that the meridian of one solid torus becomes the (p, q) curve on the boundary of the other. The space $L(p, q)$ has (cyclic) fundamental group $\mathbb{Z}/p\mathbb{Z}$, as can be seen by taking p copies of $Q(p)$ and gluing the top of one to the bottom of the next (in cyclic order) with a twist of q units; the result is evidently S^3, and admits a natural covering map to $L(p, q)$.

If p is even, we can draw an embedded loop γ on $\partial Q(p)$ (indicated in red in Figure 4) which bounds a disk D in $Q(p)$. Now, let $q = 1$. When we glue up $Q(p)$ to $L(p, 1)$, the boundary of this disk is glued to itself, and the result is a nonorientable 1-sided surface S with Euler characteristic $1-p/2$. Compress S as much as possible. Each compression produces a surface with at least one component S' which is still 1-sided and nonorientable. Since $\pi_1(L(p,1)) = \mathbb{Z}/p\mathbb{Z}$, the end result can't be π_1-injective unless S' is a projective plane. But a 1-sided projective plane has a neighborhood whose boundary is a separating sphere; by Seifert van-Kampen this sphere gives rise to a decomposition of $\pi_1(L(p,1))$ as a free product $G * \mathbb{Z}/2\mathbb{Z}$ for some G. But this is impossible for p even and bigger than 2. Thus S' is embedded and incompressible but not π_1-injective.

2. Stallings' theorem on ends, and the Sphere Theorem

The purpose of this section is to prove the following theorem of Stallings on the structure of groups with more than one end:

Theorem 2.1 (Ends of groups)**.** *A finitely generated group G has more than one end if an only if it is a nontrivial amalgamated free product or HNN extension over a finite subgroup; equivalently, if G admits an action without global fixed points on a simplicial tree, with finite edge stabilizers and without edge inversions.*

From this theorem and the Loop Theorem, we will deduce Papakyriakopoulos' Sphere Theorem:

Theorem 2.2 (Sphere Theorem)**.** *Let M be a compact 3-manifold with $\pi_2(M)$ nontrivial. Then M contains a 2-sided embedded sphere or projective plane whose fundamental class in π_2 is nontrivial in M.*

2.1. Groups acting on trees.

We summarize elements of the theory of group actions on trees that we need in the sequel. The main reference is [33].

Let T be a (simplicial) tree, and let G act on T simplicially. An element g acts with an *edge inversion* if there is some edge e of T so that g takes e to itself with the opposite orientation — i.e. it exchanges the two endpoints. If there are no such elements, we say G acts without edge inversions. This can always be achieved by first subdividing the edges of T if necessary.

We usually make the simplifying assumption that G acts minimally — i.e. that it does not preserve any proper nonempty subtree. Thus, for every vertex v the convex hull of the orbit Gv is equal to T. Equivalently, for any two vertices v and w there are elements $g, g' \in G$ so that w is on the unique embedded path from gv to $g'v$.

Suppose G acts on T without inversions. Then the quotient T/G is a simplicial graph. It is often convenient to simplify T and the action so that the quotient is as simple as possible.

Lemma 2.3. *A group G admits a nontrivial decomposition as an amalgamated product $G = A *_B C$ or HNN extension $G = A*_B$ if and only if it acts minimally on a nontrivial tree T without inversions in such a way that $B = G_e$, the stabilizer of some edge e.*

Proof. First suppose G acts on T minimally and without inversions. Let e be an open edge, and define T' to be the quotient of T where every component of $T - Ge$ is collapsed to a point. Then T' is a tree, and T'/G is a graph with exactly one edge. Thus T'/G is either an interval, or a loop. Let u, v be the vertices of e in T'. In the first case, u and v map to the two distinct vertices of T'/G; otherwise there is some g with $gu = v$. If $A = G_u$, $C = G_v$ and $B = G_e$ then G has a decomposition of the desired form.

To see that this decomposition is nontrivial, suppose otherwise, so that $G_v = G_e$ (say). Since G acts without inversions, and the quotient has a single edge, the group G_v must act *transitively* on the edges incident to v. But G_v fixes e and therefore e must be the *unique* edge incident to v. This already contradicts the case that T'/G is a loop; if T'/G is an interval, then all vertices which are neighbors of u are translates of v, so they are 1-valent; but then u is fixed by G, contrary to the hypothesis that G acts minimally on T (and therefore also on T').

Conversely, suppose G admits the structure of a nontrivial amalgamated product $A *_B C$. By Seifert van-Kampen we can build a $K(G, 1)$ from a $K(A, 1)$, $K(C, 1)$ and $K(B, 1)$ by attaching the $K(B, 1)$ to the other factors by the mapping cylinders associated to the inclusion monomorphisms. The universal cover contains copies of the universal covers of the $K(A, 1)$ and $K(C, 1)$ factors, separated by copies of the universal covers of the $K(B, 1)$ factors. The pattern of attachment in the universal cover is a tree. The case of an HNN extension is similar. □

Example 2.4 ($\mathrm{PSL}(2, K)$). The quintessential example of a group acting on a tree is $\mathrm{PSL}(2, K)$ when K is a field with a discrete valuation $v : K^* \to \mathbb{Z}$. In K we find the ring $\mathcal{O}$ (consisting of elements with non-negative valuation), which is a local ring with maximal ideal $\mathfrak{m}$ the set of elements with positive valuation, and quotient field $k := \mathcal{O}/\mathfrak{m}$. Note that $\mathfrak{m}$ is a principal ideal, and any element π with $v(\pi) = 1$ is a generator (π is called a *uniformizer*).

There is a tree T whose vertices correspond to *projective equivalence classes* of $\mathcal{O}$-modules $\Lambda \subset K^2$ isomorphic to $\mathcal{O}^2$, where two lattices Λ, Λ' are equivalent if there is $\alpha \in K^*$ with $\alpha\Lambda = \Lambda'$. Two equivalence classes of lattices Λ and Λ' are joined by an edge if there is some $\alpha \in K^*$ so that $\alpha\Lambda \subset \Lambda'$ and $\Lambda'/\alpha\Lambda = k$. The group $\mathrm{PSL}(2, K)$ acts on T, and the stabilizer of each vertex is isomorphic to $\mathrm{PSL}(2, \mathcal{O})$.

If we identify $\Lambda/\pi\Lambda = k^2$, equivalence classes of lattices Λ' joined by an edge to Λ correspond to lines in the plane k^2. Thus, the set of neighbors of each vertex is identified with the projective line over k, and the action of the stabilizer $\mathrm{PSL}(2, \mathcal{O})$ on this projective line is by its image in $\mathrm{PSL}(2, k)$ under the quotient map $\mathcal{O} \to k$.

Thus $\mathrm{PSL}(2, K)$ acts transitively on the neighbors of each vertex, and the quotient of T is an interval, giving rise to a decomposition of $\mathrm{PSL}(2, K)$ as an amalgamated product

$$\mathrm{PSL}(2, K) = \mathrm{PSL}(2, \mathcal{O}) *_P \mathrm{PSL}(2, \mathcal{O})^B$$

where superscript denotes conjugation by the matrix $B := \left(\begin{smallmatrix} \pi & 0 \\ 0 & 1 \end{smallmatrix}\right)$, and P is the "parabolic" subgroup of $\mathrm{PSL}(2, \mathcal{O})$ consisting of matrices congruent to $\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$ mod $\mathfrak{m}$. The noncompact tree can be compactified by adding a Cantor set of "ends", corresponding to the points in the projective line over $K_{\mathfrak{m}}$, the $\mathfrak{m}$-adic completion of K.

See Serre [33] for details.

As an application of Lemma 2.3 we deduce the Kurosh Subgroup Theorem:

Theorem 2.5 (Kurosh Subgroup Theorem). *Suppose B is a subgroup of a free product $A_1 * A_2$. Then B is a free product $B = B_0 * B_1 * B_2 * \cdots$ where B_0 is free, and for all $i > 0$, the group B_i is conjugate into one of A_1 or A_2.*

Proof. Associated to the decomposition $G := A_1 * A_2$ there is a minimal action of G without edge inversions on a tree T such that T/G is an interval, and such that for some edge e of T the vertex stabilizers are A_1 and A_2, and the edge stabilizer is trivial. Now let B act on T. Edge stabilizers are contained in edge stabilizers of G and are therefore trivial. Vertex stabilizers are contained in vertex stabilizers of G and are therefore conjugate into one of the A_i. The quotient $\Gamma := T/B$ is a graph, and let Γ' be a maximal tree in Γ. The graph of groups decomposition associated to Γ' is a group B' exhibited as a free product of the vertex groups (since all edge groups are trivial). The graph of groups decomposition of B is obtained from B' by iterated HNN extensions associated to the edges of $\Gamma - \Gamma'$; each of these has trivial amalgamating subgroup, so contributes a free generator to B. Thus $B = F * B'$ where $F = \pi_1(\Gamma)$ and B' is a free product of subgroups B_j each conjugate into some A_i. □

2.2. **Ends of groups.** We now define ends of groups, and give the proof of Theorem 2.1, following [38], § 4.B.

If X is a locally finite graph, define the set of ends $\mathcal{E}(X)$ as the inverse limit

$$\mathcal{E}(X) := \lim_{\leftarrow} \pi_0(X - K)$$

taken over the (directed) system of all compact subsets K of X. Note that each $\pi_0(X - K)$ is itself a finite set, so this inverse limit can be naturally topologized as a compact totally disconnected space.

Suppose G is a finitely generated group, and let S be a finite generating set. Form the Cayley graph $C_S(G)$ and define $\mathcal{E}(G)$, the space of ends of G, to be $\mathcal{E}(G) := \mathcal{E}(C_S(G))$. It is straightforward to see that $\mathcal{E}(G) = \mathcal{E}(X)$ for any proper connected graph X on which G acts properly and cocompactly, so this definition is independent of the choice of generating set S.

We can also give a homological definition of the *number* of ends, as follows: if C^* denotes (simplicial) cochains on X with values in $\mathbb{Z}/2\mathbb{Z}$, and C^*_f denotes cochains with finite support (which form a complex, because X is locally finite) then define C^*_e by the exact sequence

$$0 \to C^*_f \to C^* \to C^*_e \to 0$$

and then the number of ends is the dimension of $H^0_e(X)$. There is an exact sequence in cohomology:

$$H^0_f(X) \to H^0(X) \to H^0_e(X) \to H^1_f(X) \to H^1(X)$$

If X is connected and infinite, then $H^0_f(X) = 0$ and $H^0(X) = \mathbb{Z}/2\mathbb{Z}$. Thus X has more than one end if and only if $H^1_f(X) \to H^1(X)$ has nontrivial kernel.

If X is a graph, we consider $\mathbb{Z}/2\mathbb{Z}$-valued cochains on X. A subset V of vertices is identified with the support of a unique 0-cochain; by abuse of notation we also write this cochain as V. Thus δV is a 1-cocycle, which is also identified with its support, which is a union of edges. With this language, if X is the Cayley graph of G with respect to any finite generating set, then if we define A to be the G-module consisting of all subsets of G modulo finite subsets (where subsets are identified as above with 0-cochains) then $H^0_e(X) \cong H^0(G; A)$; see e.g. Epstein [8] for details. This gives another way to see that the number of ends is independent of the choice of generating set.

Let $Q(X)$ denote the set of subsets of vertices (equivalently 0-cochains) V for which δV is finite. A set $V \in Q(X)$ is said to be *connected* if the complete subgraph it spans is connected. It is *nontrivial* if δV is not equal to δF for any *finite* F. If V is in $Q(X)$, so is its complement V^*. Notice that as 0-cochains, there are formulae $V^* = 1 - V$ and $V \cap W = VW$. For $V \in Q(X)$, we say that an end $\eta \in \mathcal{E}(X)$ is contained in V if for each compact set K containing δV, the component of $X - K$ corresponding to η is contained in V. Thus, a nontrivial V determines a nontrivial partition of $\mathcal{E}(X)$ into $V \cap \mathcal{E}(X)$ and $V^* \cap \mathcal{E}(X) = (V \cap \mathcal{E}(X))^*$; notice that both of these subsets are closed.

A nontrivial $V \in Q(X)$ is *narrow* if $|\delta V|$ is minimal; call this minimal value k the *width* of X. It is easy to verify that any narrow V is connected. Let $\cdots V_n \subset \cdots \subset V_2 \subset V_1$ be an infinite decreasing family of narrow sets, and suppose $W = \cap_n V_n$ is nonempty. Then every edge in δW is contained in δV_i for all sufficiently large i, and therefore all of δW is contained in δV_n for some n. On the other hand, $W \cap \mathcal{E}(X) = \cap_n (V_n \cap \mathcal{E}(X))$ is nonempty, whereas $V_n^* \cap \mathcal{E}(X) \subset W^* \cap \mathcal{E}(X)$ for any n; thus W is nontrivial and contained in $Q(X)$, and is therefore narrow. It follows that for any vertex $v \in X$ we may find a *minimal* narrow V containing v.

The key lemma is the following:

Lemma 2.6 (Nested narrow sets)**.** *Let V be a minimal narrow set containing v, and let W be any other narrow set. Then at least one of VW, V^*W, VW^*, V^*W^* is finite.*

Proof. We estimate

$$|\delta VW| + |\delta V^*W| + |\delta VW^*| + |\delta V^*W^*| \leq 4k$$

since each of the boundary components is made up of pieces of the boundary of V and W, and each piece occurs on at most two sides. If $|\delta VW| < k$ (say) then $\delta VW = \delta F$ for some finite F. But then $VW + F = 1$ or $VW + F = 0$; the first case is impossible because VW is contained in V, and V^* is infinite; thus $VW = F$ and we are done. So we deduce that each of the four sets above is narrow. But then one of VW or VW^* contains v and is properly contained in V, contrary to the hypothesis that V is minimal. □

Now let V be any minimal narrow set containing v, and consider the family of its translates gV for $g \in G$. Each of these translates is narrow and minimal for gv, so this system of sets induces a *nested* partition of $\mathcal{E}(X)$. We may therefore build a tree T whose edges are equivalence classes of $g\delta V$, where $g\delta V$ and $h\delta V$ are equivalent if they separate $\mathcal{E}(X)$ in the same way. The group G acts on T minimally.

If X has more than 2 ends, then since X is locally finite, any two translates $g\delta V$ and $h\delta V$ which are not sufficiently close have at least one end "between" them, and are therefore inequivalent, so that edge stabilizers are finite. If X has 2 ends, then it is straightforward to see (and was anyway known by Freudenthal) that G is virtually cyclic. In either case this proves the theorem.

2.3. **3-manifold groups acting on trees.**

Proposition 2.7 (Action on tree)**.** *Let M be a compact 3-manifold with fundamental group G, and suppose G acts minimally on a tree T without inversions. Let B be a conjugacy class of edge stabilizer. Then M contains a 2-sided essential embedded surface S whose fundamental group injects into B.*

Proof. As above we can assume that G acts transitively on the edges of T. We can write G as $A *_B C$ or $A*_B$ where B is an edge stabilizer, and build a space K by gluing a $K(A,1)$ and $K(C,1)$ along a $K(B,1)$ (in the first case), or attaching a $K(B,1)$ in two ways to a $K(A,1)$ (in the second case). We therefore obtain a homotopy class of map $f : M \to K$; choose a map in this homotopy class which is transverse to $K(B,1)$, so that the preimage of $K(B,1)$ is an embedded 2-sided surface S in M. If some component of S is not π_1-injective, we can repeatedly compress it by applying Kneser's Lemma (i.e. Lemma 1.6); this compression can be achieved by a homotopy of the map, and reduces the complexity of S, so after finitely many such compressions, we can assume that $\pi_1(S)$ injects into B. □

2.4. **Proof of the Sphere Theorem.** We now prove the sphere theorem, following Stallings. Suppose that M is a compact 3-manifold with $\pi_2(M)$ nontrivial. If ∂M contains a sphere or projective plane which is contractible in M, then M is itself contractible, contrary to hypothesis; any other sphere or projective plane component of ∂M satisfies the conclusion of the theorem, so we assume there are no such components. If some component of ∂M is not injective, then by the Loop Theorem we can compress it to produce a new, simpler manifold. Since M is homotopy equivalent to a wedge of the form $M' \vee M''$ or $M' \vee S^1$, we deduce that $\pi_2(M')$ is nontrivial for at least one factor if $\pi_2(M)$ is. So after finitely many compressions we obtain $M'' \subset M$ with incompressible boundary, no component of which is a sphere or projective plane, and $\pi_2(M'') \neq 0$.

Thus if $\tilde{M}''$ denotes the universal cover of M'', every component of $\partial\tilde{M}''$ is contractible, and therefore

$$\pi_2(M'') = \pi_2(\tilde{M}'') = H_2(\tilde{M}'') = H_2(\tilde{M}'', \partial\tilde{M}'') = H^1_c(\tilde{M}'')$$

On the other hand, $H^1(\tilde{M}'') = 0$ because $\tilde{M}''$ is simply connected, and $H^0_c(\tilde{M}'') = 0$ because $\tilde{M}''$ is noncompact, and therefore $\tilde{M}''$ has more than one end. So we deduce that $\pi_1(M)$ acts nontrivially on a tree without inversions, and with *finite* edge stabilizers.

By Proposition 2.7, M admits a 2-sided essential embedded surface S whose fundamental group injects into the edge stabilizer; but this implies that S is a sphere or projective plane.

3. Prime and free decompositions, and the Scott Core Theorem

Definition 3.1. If M, M' are connected, oriented 3-manifolds, the *oriented connect sum* $M\#M'$ is obtained by removing a small open ball from both M and M', and gluing the resulting boundary spheres by an orientation-reversing homeomorphism.

A priori it might appear that connect sum depends on the choice of the small balls we remove, and on the choice of the homeomorphism used to glue the resulting boundary spheres. However, there is no ambiguity, since firstly any two balls in the same component are isotopic, and secondly there is only one orientation-reversing homeomorphism of the 2-sphere up to isotopy.

Example 3.2. Connect sum with S^3 gives back the same manifold; i.e. $M\#S^3 = M$.

Definition 3.3. Connect sum is *nontrivial* if neither of the factors is S^3. A 3-manifold is *prime* if it cannot be written as a nontrivial connect sum. A *decomposition* of M is a factorization $M = M_1\#M_2\#\cdots\#M_n$ where no M_i is S^3 unless $i = 1$ and $M = S^3$.

Example 3.4. If ∂M has sphere components, we denote by $\hat{M}$ the result of capping off these components. Then for any 3-manifold, $M = \hat{M}\#^n B^3$ where n is the number of sphere components of ∂M.

The first purpose of this section is to prove the Prime Decomposition Theorem:

Theorem 3.5 (Prime Decomposition Theorem)**.** *Let M be a closed, oriented 3-manifold. Then M admits a finite decomposition into prime factors. Furthermore this decomposition is unique (up to order of the factors).*

The existence of a finite decomposition is due to Kneser, and the uniqueness is due to Milnor. Along the way we explain the connection between connect sum decompositions of manifolds and free decompositions of (fundamental) groups, and prove Grushko's Theorem and Kneser's Conjecture, following Stallings. This group-theoretic work in turn lets us give a proof of the Scott Core Theorem [31]:

Theorem 3.6 (Scott Core Theorem)**.** *Let M be a (possibly noncompact) 3-manifold with $\pi_1(M)$ finitely generated. Then there is a compact 3-submanifold $C \subset M$, called a* Scott core*, such that the inclusion induces an* isomorphism $\pi_1(C) \to \pi_1(M)$*. In particular, $\pi_1(M)$ is finitely presented.*

3.1. **Grushko's Theorem.** By Seifert–van Kampen, if $M = M_1\#M_2$ then $\pi_1(M) = \pi_1(M_1) * \pi_1(M_2)$. If G is a finitely generated group, the *rank* of G is the minimal number of generators for G. Grushko's Theorem says the following:

Theorem 3.7 (Grushko's Theorem)**.** *Let G be a finitely generated group, and suppose $G = A * B$. Then* $\text{rank}(G) = \text{rank}(A) + \text{rank}(B)$*.*

A group with rank 0 is the trivial group; thus Grushko's Theorem immediately implies that in any decomposition of a 3-manifold *with factors that are not simply connected* the number of factors is bounded by the rank of $\pi_1(M)$. Thus existence of a prime decomposition follows from Grushko's Theorem once one knows:

Theorem 3.8 (Poincaré Conjecture)**.** *Let M be a closed, simply-connected 3-manifold. Then $M = S^3$.*

The Poincaré Conjecture is a theorem of Perelman [24, 25, 26], and is proved using PDE methods — i.e. Ricci flow — together with a very subtle analysis of the finite and infinite time singularities. On a Riemannian manifold, both the metric tensor g and the Ricci curvature R are symmetric 2-forms; i.e. they are sections of the same tensor bundle. So it makes sense to evolve the metric by $\partial_t g = -2R$. This has the effect of spreading out the manifold in directions where it is negatively curved, and shrinking it where it is positively curved. With arbitrary initial data, one has short time existence and uniqueness, and a singularity can occur in finite time only by the manifold shrinking to a point (in which case the metric becomes asymptotically round and the manifold is exhibited as a quotient of S^3 by a finite group of isometries) or by the curvature blowing up to $+\infty$ along some embedded sphere which becomes *pinched*. In the first case one sees that the manifold is finitely covered by S^3, so if it is simply-connected it must be S^3. In the second case, before the singularity occurs, one can cut this sphere and plug in rounded balls, exhibiting the original manifold as a (possibly trivial) connect sum of new manifolds on which the Ricci flow can be continued. Only finitely many singularities of the second kind can develop in finite time, and if π_3 of the original manifold is nontrivial, the same is true for each of the factors that is produced. A 3-manifold with nontrivial π_3 admits a nontrivial "minimax sweepout" by immersed spheres, and in [26] it is shown that the diameter of the critical sphere must go to zero at a definite rate. Thus each of the factors shrinks to a point in finite time, and the original manifold is certified as a connect sum of finitely many copies of S^3.

It would be perverse to invoke the Poincaré Conjecture to prove the Prime Decomposition Theorem, and we eschew this strategy. But we give a proof of Grushko's Theorem following Stallings [35] in order to explain its close relation to 3-manifold topology.

Proof. Every subgroup of a free group is free, and the rank of a free group is equal to the rank of H_1, which is additive under free product. Thus Grushko's theorem is true for free groups.

Evidently $\text{rank}(G) \le \text{rank}(A) + \text{rank}(B)$, so it suffices to find a surjection $\phi : F \to G$ where F is a free group with $\text{rank}(F) = \text{rank}(G)$, and a decomposition $F = F_1 * F_2$ where $\phi(F_1) = A$ and $\phi(F_2) = B$.

A *handlebody* is a regular neighborhood of a graph Γ in S^3 (for a precise definition, look ahead to Definition 3.20). The double of a handlebody is a closed manifold M. Note that M retracts onto Γ, and this retraction induces an isomorphism on π_1, so $\pi_1(M) = F$ is free of prescribed rank. Choose a surjection $\phi : F \to G$ where $\text{rank}(F) = \text{rank}(G)$. Let $K = K(A,1) \vee K(B,1)$, and build a map $f : M \to K$ realizing ϕ. As in the proof of Proposition 2.7, we can make this map transverse to the basepoint p of the wedge, and compress $f^{-1}(p)$ until it consists of a separating union S of 2-sided spheres. Suppose the collection S has as few components as possible. We claim S consists of a single component. This will prove the theorem, since S will exhibit M as $M = M_1 \# M_2$, and by construction, $\phi : \pi_1(M_1) \to A$ and $\phi : \pi_1(M_2) \to B$ will be surjective.

So suppose S contains at least two spheres. The union S is separating, so let α be an arc properly contained in $M - S$ from one component to a different one, and such that $f(\alpha)$ is a loop in K. Let β be a loop in M ending at the endpoint of α, so that if $\gamma = \alpha * \beta$, then $f(\gamma)$ is homotopically trivial in K. Make γ transverse to S and with the minimal number of intersections, so that it is decomposed into segments γ_i. By minimality, each γ_i is essential in M rel. endpoints, and $f(\gamma_i)$ is a loop in K on one side. Since an alternating product of nontrivial elements in a free product is nontrivial, some $f(\gamma_i)$ is trivial in A or B; if the preimage starts and ends on the same component of S, it represents a loop in M in the kernel of $F \to G$, and we may cut the arc γ_i out and reduce the number of intersections. So at the end we obtain an arc γ properly contained in $M - S$ from one component of S to a different one, and such that $f(\gamma)$ is homotopically trivial in $K(A,1)$ (say). A tubular neighborhood C of γ has boundary an annulus D running from component S_0 to S_1 (say); if D' is obtained by pushing the interior of D slightly into C, then $D \cup D'$ is a torus bounding a solid torus (just a thickened neighborhood of ∂C in C). Let C' be the neighborhood of γ bounded by D'. Define a new map $f' : M \to K$ which agrees with f outside C, and such that $f'(C') = p$. The meridian of the torus $D \cup D'$ maps to a homotopically trivial loop in $K(A,1)$, so we can extend f' over a compressing disk for this meridian. Then we can extend f' over the remaining 3-ball of the solid torus, using the fact that $\pi_2(K(A,1)) = 0$. The result is a new map f' for which the preimage of p has one fewer component, and one component is $C' \cup S_0 \cup S_1$; now we can push the interior of C' into B to produce f'' transverse to p and such that the preimage of p is a union of strictly fewer spheres. So we are done by induction. □

Essentially the same argument proves Kneser's Conjecture, as was observed by Stallings [35]:

Theorem 3.9 (Kneser's Conjecture)**.** *Let M be a 3-manifold, and suppose $\pi_1(M) = G_1 * G_2$. Then there is a connect sum decomposition $M = M_1 \# M_2$ where $\pi_1(M_i) = G_i$.*

Proof. Build $K = K(G_1, 1) \vee K(G_2, 1)$ and $f : M \to K$ and $S = f^{-1}(p)$ consisting of a nonseparating union of spheres as above. If S has more than one component, then by the same argument as above we can replace f by f' so that $(f')^{-1}(p)$ consists of fewer spheres. □

3.2. **Alexander's Theorem.** Since $M = M \# S^3$, to have any hope of proving the Prime Decomposition Theorem (without assuming the Poincaré Conjecture!) we must be sure that S^3 itself does not admit a nontrivial decomposition.

An embedded sphere in S^3 is necessarily separating, since $H_2(S^3; \mathbb{Z}/2\mathbb{Z}) = 0$. In fact, Alexander duality shows that H_1 of the complement must vanish. On the other hand, if one allows arbitrary (i.e. wild) embeddings, the complement of the sphere can be very complicated.

Example 3.10 (Alexander's horned sphere). The sphere indicated in Figure 5 is smooth away from a Cantor set, and does not bound a ball on the "outside"; in fact, the fundamental group of the exterior is not finitely generated.

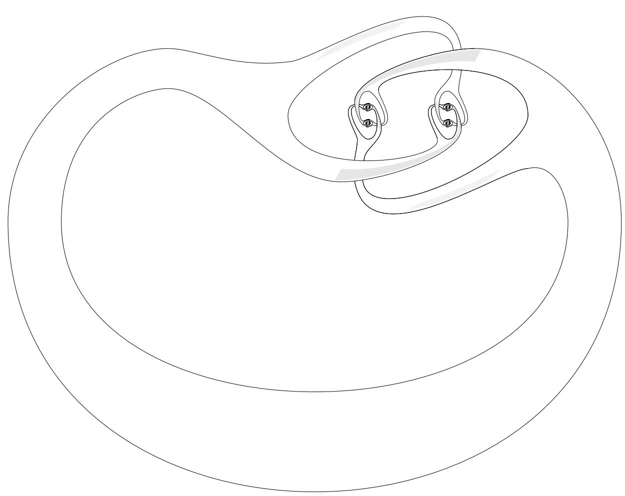


FIGURE 5. Alexander's horned sphere does not bound a ball on one side; the construction can be modified so that the sphere does not bound a ball on either side.

The spheres arising in a direct sum decomposition are locally flat (i.e. they have an I-bundle collar neighborhood) and locally flat surfaces in 3-manifolds may be taken to be smooth (or PL). So it is just as well that every smooth (or PL) sphere in S^3 is standard:

Theorem 3.11 (Alexander)**.** *Every smooth embedded sphere in S^3 bounds a ball (on either side).*

Proof. By removing a point from a complementary region, it suffices to show that every smooth embedded sphere in $\mathbb{R}^3$ bounds a ball on one side. By an isotopy, put the sphere S in general position with respect to the foliation of $\mathbb{R}^3$ by horizontal planes. This means that there are finitely many nondegenerate critical points, and these occur at distinct levels, and away from the critical points, the level sets of S are finite unions of circles.

There are three kinds of critical points — local maxima, local minima, and saddles. Maxima and minima contribute 1 to χ and saddles contribute -1. We prove the theorem by induction on the number of saddles. If there are no saddles, there must be exactly one maximum and one minimum, and each intermediate level set consists of a single circle bounding a disk in its level plane; the union of these disks is a ball, and the theorem is proved.

Suppose some nonsingular level set consists of a union of circles. Choose an innermost circle γ which bounds a disk D in its level set. We can compress S along γ to produce two new (disjoint) smooth spheres S_1, S_2 each with the same critical points as γ, except for a new minimum and maximum coming from D. So either S_1 and S_2 both have fewer saddles than S (in which case they are both standard by induction), or else S_1 (say) has no saddles; but then S_1 bounds a ball, and we see that S_2 is isotopic to S and with the same number of saddles, and we have eliminated the circle γ. If S has at least two saddles, we can find some circle α with at least one saddle on both sides. By eliminating innermost circles as above we can find ourself (after finitely many steps) at a point where α is innermost. Then compressing along α produces two spheres with definitely fewer saddles.

So we are reduced to the case of a single saddle. Just above this saddle (without loss of generality) S intersects the level set in two circles γ_1, γ_2, which bound disks D_1, D_2 in their level set, and disjoint disks E_1, E_2 in S each with a single maximum. Each of the spheres $D_1 \cup E_1$ and $D_2 \cup E_2$ thus bounds a ball; suppose $D_2 \cup E_2$ is innermost. Then the ball bounded by $D_2 \cup E_2$ is disjoint from S and we can push E_2 across this ball to D_2, and cancel the maximum with the saddle. This eliminates the saddle, and proves the theorem. □

Alexander's Theorem shows that S^3 is prime; in fact it is *irreducible*, which means that every (smoothly) embedded sphere bounds a ball. The difference between prime and irreducible is that for a prime manifold, one only knows that every (smoothly) embedded *separating* sphere bounds a ball. The following proposition shows that nonseparating spheres all arise for the same reasons:

Proposition 3.12. *Let M closed and oriented contain a nonseparating sphere. Then there is a (possibly trivial) decomposition $M = M' \# S^2 \times S^1$. Thus, a prime (closed, oriented) 3-manifold is irreducible unless it is $S^2 \times S^1$.*

Proof. Let S be a nonseparating sphere, and let γ be an embedded circle which meets S transversely in one point. Set N to be a regular neighborhood of $S \cup \gamma$. Then N is homeomorphic to $S^2 \times S^1 - B^3$, and ∂N is a separating sphere which exhibits the desired connect sum decomposition of M. □

3.3. **Existence of prime decomposition.** We now prove the existence of a prime decomposition, following Kneser. In fact, we will show that there is a (readily computable) quantity $k(M)$ so that any decomposition of M into more than $k(M)$ factors has some S^3 summand. In fact, if M admits a triangulation with t 3-simplices, we may take $k(M) = 6t + \text{rank}(H_1(M;\mathbb{Z}/2\mathbb{Z}))$.

The method of proof is also important, since it proceeds by taking a collection of disjoint spheres, and simplifying the intersection with the 2-skeleton of a triangulation. We shall see in the sequel that many important theorems can be deduced from properties of the intersection of some 2-dimensional object with a triangulation which minimizes a suitable complexity.

Suppose we have a nondegenerate decomposition $M = M_1 \# M_2 \# \cdots \# M_n$, and let S be a disjoint union of spheres witnessing the decomposition. Fix a triangulation τ of M, and arrange S so that $S \cap \tau$ is in general position. This means that S is disjoint from the vertices of τ, that it intersects the edges of τ in finitely many isolated points, and that it intersects the faces of τ in finitely many isolated loops and proper arcs. Define a lexicographic complexity (e, f) where e is the number of points of $\tau^1 \cap S$, and f is the number of components of $\tau^2 \cap S$, and suppose that we have found a disjoint union of $(n-1)$ spheres S defining a nondegenerate decomposition, for which the complexity is minimal.

Note that the M_i are obtained from the components of $M - S$ by capping off boundary spheres with balls; thus the decomposition is nondegenerate providing no component of $M - S$ is homeomorphic to a multiply-punctured ball. We will modify the isotopy class of the collection S in the course of the argument, while preserving the number of components and the nondegeneracy.

First we eliminate loops of $\tau^2 \cap S$. If γ is an innermost loop of $\tau^2 \cap S$, then it bounds a disk D in τ^2 with interior disjoint from S, and we can compress the component of S meeting γ along D to produce $S_1 \cup S_2$. At least one of the S_i does not bound a (punctured) ball on either side; throwing the other sphere away gives a new nondegenerate decomposition with the same number of spheres and with smaller f. Thus for (e, f) minimal, there are no loops of $\tau^2 \cap S$.

Next we eliminate arcs of $\tau^2 \cap S$ with both endpoints on the same edge. Such an arc bounds a bigon D in τ^2 and we can push S over D to reduce e by 2. Thus for (e, f) minimal, every arc of $\tau^2 \cap S$ has endpoints on distinct edges; such arcs are called *normal*.

Next we arrange that for every 3-simplex σ of τ, every component of $S \cap \sigma$ is a disk. For, otherwise, some innermost (non-disk) component R has one boundary component which bounds an innermost disk in $\partial\sigma$. This disk can be pushed into σ to a disk D with boundary on ∂R, with the interior of D disjoint from S (by the assumption that R is innermost among non-disk components) and then we can compress S along D and throw away one component, reducing e. Thus for (e, f) minimal, every component of $S \cap \sigma$ is a disk for every 3-simplex σ of τ.

Now consider a simplex σ. Its boundary is decomposed by loops of $S \cap \partial\sigma$ into surfaces, whose Euler characteristics sum to 2. Every disk component must contain at least one vertex (since its boundary is made of normal arcs), so there are at most 4 of these. Thus there can be at most 2 more components with negative Euler characteristic, and every other component is an annulus containing no vertex; call these annuli *good*. Thus: there are at most 6 bad pieces, and all the rest are good annuli. Each good annulus bounds two disks in $\sigma \cap S$, and the union bounds an I-bundle over a disk. Any component of $M - S$ made up of these I-bundles is itself an I-bundle; every other component has at least one bad piece in some simplex, so there are at most $6t$ such components, where t is the number of 3-simplices of τ. Every I-bundle has a sphere as a boundary component, so it is either $S^2 \times I$ or a punctured $\mathbb{RP}^3$ (i.e. a twisted I-bundle over $\mathbb{RP}^2$). By hypothesis, the splitting is nondegenerate, so there are no $S^2 \times I$ complementary components. But every punctured $\mathbb{RP}^3$ adds a nontrivial $\mathbb{Z}/2\mathbb{Z}$ summand to $H_1(M; \mathbb{Z}/2\mathbb{Z})$. Thus we obtain Kneser's bound $n \le 6t + \text{rank}(H_1(M; \mathbb{Z}/2\mathbb{Z}))$, and prove the existence of a prime decomposition.

3.4. **Uniqueness of prime decomposition.** We give the proof of the uniqueness of prime factorization, following Milnor [19], thus completing the proof of Theorem 3.5. We prove the theorem for closed, oriented 3-manifolds. Explicitly, we show that if M admits two prime decompositions as $M = M_1 \# M_2 \# \cdots \# M_n$ and $M = M_1' \# M_2' \# \cdots \# M_m'$ then $m = n$, and the factors are the same up to permutation. We may further assume that m and n are both bigger than 1, and that no M_i or M_j' is S^3.

The first step is to match $S^2 \times S^1$ factors on both sides. Note that a closed surface S embedded in a 3-manifold is nonseparating if and only if the class of $[S]$ is essential in $H_2(M; \mathbb{Z}/2\mathbb{Z})$. For, a separating surface bounds on either side, whereas a nonseparating surface intersects some transverse essential loop γ in 1 point, so that $[S] \cap [\gamma] = 1$.

Lemma 3.13. *Let M be a closed, oriented 3-manifold, and let $M = M_1 \# M_2 \# \cdots \# M_n$ be a decomposition into prime factors. Then one of the factors is an $S^2 \times S^1$ if and only if M contains a nonseparating embedded sphere.*

Proof. If one of the factors is $S^2 \times S^1$, then there is an embedded $S^2 \times S^1 - B^3$ in M, and there is a nonseparating sphere in this $S^2 \times S^1 - B^3$.

Conversely, suppose M contains a nonseparating sphere S. Let T be the collection of embedded spheres decomposing M into its prime factors (minus balls); each component of T is separating, and therefore homologically trivial. Let S intersect T transversely in as few circles as possible. If S is disjoint from T, it is contained in some punctured M_i; but then M_i contains an $S^2 \times S^1$ summand, and is therefore already equal to $S^2 \times S^1$. If $S \cap T$ is nonempty, consider a loop of intersection γ bounding an embedded disk D in T disjoint from S. Let γ decompose S into E_1 and E_2. Then there are two new spheres $E_1 \cup D$ and $E_2 \cup D$ such that $[E_1 \cup D] + [E_2 \cup D] = [S]$ in $H_2(M; \mathbb{Z}/2\mathbb{Z})$ (so that at least one is nonseparating), and such that each intersects T in fewer pieces than S (after an isotopy). The lemma follows. □

Lemma 3.14. *Let M be a closed, oriented 3-manifold and let S, S' be nonseparating embedded spheres. Then there is a self-homeomorphism of M taking S to S'.*

Proof. If the spheres are disjoint, let $N = M - (S \cup S')$, so that N is either connected with 4 boundary spheres, or disconnected into two pieces with 2 boundary spheres. Cap off boundary spheres with balls, and find an isotopy in each component which interchanges the balls in pairs. Now remove the balls, and extend the result of the isotopy to a self-homeomorphism of M.

If the spheres intersect, let γ be a circle of intersection bounding a disk D in S disjoint from S', and decomposing S' into E_1 and E_2. Then as above, one of the $E_i \cup D$ is a nonseparating sphere disjoint from S' and intersecting S in fewer circles that S' does. Then there is a homeomorphism taking S' to $E_i \cup D$, and by induction another homeomorphism taking $E_i \cup D$ to S. □

It follows that we may match up $S^2 \times S^1$ factors of the two decompositions of M, and by splitting off these summands one by one, we may reduce to the case that every embedded sphere in M is separating.

So, let S be a sphere bounding $M_1 - B^3$, and let T be a union of spheres realizing the second decomposition, so that $M - T$ is a union of punctured M_i'. We would like to realize S disjoint from some such collection T; if we can do this, it will be contained in some punctured M_j', and we will therefore realize this punctured M_j' as $M_1 - B^3$ union a punctured ball, so that $M_j' = M_1$ and we can split off the two factors, and the theorem will follow by induction. So suppose $S \cap T$ has the minimal number of pieces where S and T are as above. An innermost disk D of S is contained in some punctured M_j' and its boundary decomposes some component T_i of T into E_1 and E_2. As before, consider the two spheres $E_1 \cup D$ and $E_2 \cup D$. These are contained in a punctured M_j', and must therefore both bound punctured balls B_1 and B_2. If the punctured balls were disjoint, their union would be all of the punctured M_j', thereby exhibiting M_j' as S^3, which is absurd; thus $B_1 \subset B_2$ (say). But then $E_2 \cup D$ bounds a punctured M_j' (with possibly fewer punctures than before) on the other side, so we can replace the component T_i by $E_2 \cup D$, obtaining a new collection T' meeting S in fewer components. This completes the proof.

3.5. **Scott Core Theorem.** We now give the proof of the Scott Core Theorem, whose most important corollary is the fact (proved independently by Shalen) that a finitely generated 3-manifold group is finitely presented.

If H is a finitely generated group, the Kurosh Subgroup Theorem (i.e. Theorem 2.5) implies that the indecomposable factors of H are unique up to order and isomorphism, and the non-$\mathbb{Z}$ factors are unique up to conjugacy (the number of each kind of factor is finite by Grushko's Theorem 3.7). Let $i(H)$ denote the number of indecomposable factors of H, and $f(H)$ the rank of the biggest free factor. Define the *complexity* of H to be the lexicographic pair $(i(H), f(H))$.

Lemma 3.15. *Let G and H be finitely generated. Let $\phi : G \to H$ be surjective, and injective on each non-$\mathbb{Z}$ indecomposable factor. Then either ϕ is an isomorphism, or the complexity of H is less than that of G.*

Proof. By the Kurosh Subgroup Theorem, for each non-$\mathbb{Z}$ indecomposable factor G_i of G, the group $\phi(G_i)$ is conjugate into some non-$\mathbb{Z}$ factor $H_{j(i)}$ of H. Let G' be the normal subgroup of G generated by the nonfree factors of G, and let H' be the normal subgroup of H generated by the $H_{j(i)}$ where the i runs over nonfree factors. Then $G/G' \to H/H'$ is surjective. Furthermore, G/G' is free of rank $f(G)$, and by Grushko's Theorem, $f(G) \geq i(H/H') \geq f(H)$. Furthermore, $i(H) = i(H/H') + d$ where d is the number of distinct indices in $j(i)$; thus $i(G) \geq i(H)$, and $f(G) = f(H)$ if and only if the $j(i)$ are all distinct, and $G/G' \to H/H'$ is an isomorphism. But in this case by Kurosh, ϕ

must be an isomorphism of free factors and an isomorphism of each indecomposable factor individually, and hence ϕ is an isomorphism. □

Lemma 3.16. *Let R be a freely indecomposable finitely generated group of finite rank $n > 1$, and suppose that every finitely generated subgroup of R of strictly smaller rank is finitely presented. Then there is a finitely presented group G and a surjection $\phi : G \to R$ which does not factor through $G \to K_1 * K_2 \to R$ for K_i both nontrivial, and where $G \to K_1 * K_2$ is surjective.*

Proof. Let $\mathcal{C}$ denote the class of finitely presented groups of rank n which surject onto R by homomorphisms which are injective on each non-$\mathbb{Z}$ indecomposable factor. The class $\mathcal{C}$ is nonempty, since it contains a free group of rank n. Each group in $\mathcal{C}$ gets a complexity $(i(G), f(G))$, and we choose such a G of minimal complexity. If $\phi : G \to R$ is an isomorphism then evidently we are done. Otherwise there is a nontrivial element g in the kernel. Define H to be the quotient of G by the normal closure of g, so that H is finitely presented, $G \to H$ is surjective, and ϕ factors as $G \to H \to R$.

Suppose $\psi : H \to R$ factors through a surjection $H \to K_1 * K_2$ where both K_i are nontrivial. Let L_i denote the image of K_i in R. By Grushko's Theorem, each K_i has rank strictly less than n, and therefore so does L_i, so by hypothesis, each L_i is finitely presented. But $G \to L_1 * L_2$ is surjective; if some non-$\mathbb{Z}$ indecomposable factor G_j of G is not conjugate into some L_i, then by Kurosh G_j does not map injectively, and then it would not be injective when further mapped by $L_1 * L_2 \to R$. It follows from Lemma 3.15 that the complexity of $L_1 * L_2$ is less than that of G. Moreover, $L_1 * L_2$ is finitely presented, since each L_i is, and of rank n since it surjects onto R. Finally, each L_i maps injectively to R, so $L_1 * L_2$ is in the class $\mathcal{C}$, contrary to the choice of G. Thus $\psi : H \to R$ does not factor through a surjection, and the lemma is proved. □

We now prove the Scott Core Theorem for $\pi_1(M)$ freely indecomposable of finite rank.

Proposition 3.17. *Let M be a 3-manifold with $\pi_1(M)$ finitely generated. Then $\pi_1(M)$ is finitely presented. Moreover, if $\pi_1(M)$ is indecomposable, there is a compact submanifold Q of M which is a Scott Core.*

Proof. For rank 0 we can take the empty set as a core, and for rank 1 we can take an embedded loop. So we can assume (by induction) that the rank is $n > 1$, and that every finitely generated subgroup of $\pi_1(M)$ of rank less than n is finitely presented. By Lemma 3.16 there is a finitely presented group G and a surjective $\phi : G \to \pi_1(M)$ such that ϕ does not factor through a surjection of G to a nontrivial free product. Let K be a compact 2-complex with $\pi_1(K) = G$, and let $f : K \to M$ induce ϕ on π_1. Let Q be a compact 3-manifold containing $f(K)$ whose boundary surface has least complexity subject to these conditions. Since the surjection ϕ factors through $\pi_1(Q)$, the inclusion $\pi_1(Q) \to \pi_1(M)$ is surjective. Furthermore, if every boundary component of Q is incompressible, then $\pi_1(Q)$ injects into $\pi_1(M)$, by Kneser's Lemma, and Seifert-van Kampen, so Q would satisfy the conclusion of the proposition.

So suppose ∂Q is compressible. If it compresses to the outside, we can enlarge Q by adding a thickened compressing disk, obtaining some new Q' whose boundary has smaller complexity, contrary to the definition of Q. If it compresses to the inside, we can split

Q along this compressing disk to obtain Q_1, Q_2 (if separating) or Q_1 (if not), and exhibit $\pi_1(Q) = \pi_1(Q_1) * \pi_1(Q_2)$ or $\pi_1(Q) = \pi_1(Q_1) * \mathbb{Z}$. By the hypothesis of rank, $\pi_1(Q_1)$ is nontrivial. By the defining property of G, we may assume that $\phi(G)$ is conjugate into Q_1, so we may replace f by a homotopic $f' : K \to Q_1$, so that Q_1 also contradicts the definition of Q. Thus, ∂Q is incompressible, and Q is a Scott Core. □

Proposition 3.17 already proves that every finitely generated 3-manifold group is finitely presented, which is one of the most useful corollaries of the Scott Core Theorem; it further shows that for every M with $\pi_1(M)$ finitely generated there is some *compact* 3-manifold N with $\pi_1(N) = \pi_1(M)$ (just take a connect sum of Cores for the indecomposable factors); but it remains to show that we can take N to be a compact submanifold of M. When $\pi_1(M)$ is freely decomposable the proof is substantially more complicated.

Lemma 3.18. *Let M be a 3-manifold with $\pi_1(M)$ finitely generated, and suppose we write $\pi_1(M) = G_1 * G_2 * \cdots * G_n * F$ where each G_i is freely indecomposable and noncyclic, and F is free. Then there is a compact submanifold N which is the union of compact submanifolds N_i and disjoint 1-handles, and satisfying*

1. *each N_i has indecomposable fundamental group and incompressible boundary;*
2. *$\pi_1(N) \to \pi_1(M)$ is a split epimorphism; and*
3. *each $\pi_1(N_i)$ is mapped isomorphically to a conjugate of G_i.*

Proof. First, let K be a compact 2-complex with $\pi_1(K) = \pi_1(M)$ (such a K exists, by Proposition 3.17), and let $f : K \to M$ induce an isomorphism on π_1, and let N be a regular neighborhood of the image. Note that $\pi_1(N) \to \pi_1(M)$ is a *split* surjection (because it can be precomposed with $\pi_1(K) \to \pi_1(N)$ in such a way that the composition $\pi_1(K) \to \pi_1(M)$ is an isomorphism). The fact that this surjection is split depends on the fact that $\pi_1(M)$ is known to be finitely presented, and is the only place in the rest of the argument where we use this.

Compress ∂N as much as possible to obtain a new submanifold consisting of a disjoint union of compact incompressible submanifolds N_i of M, and a collection of disjoint 1-handles so that the union (which we relabel as N) is compact, connected, and $\pi_1(N) \to \pi_1(M)$ is a split surjection. Next, let L be a simplicial $K(\pi_1(M), 1)$ obtained as a tree of $K(G_i, 1)$'s and $K(\mathbb{Z}, 1)$'s, one for each factor in a free decomposition of $\pi_1(M)$, and let $g : M \to L$ be transverse to the midpoints of the edges, so that the preimage F is a 2-sided embedded surface, each of whose component has trivial image in $\pi_1(M)$. Make F transverse to N, and compress the components, so that $F \cap N$ is a union of disks transverse to the 1-handles, together with incompressible surfaces contained in the N_i. Since each N_i has incompressible boundary, each incompressible surface of $F \cap N_i$ injects into $\pi_1(M)$, and is therefore a 2-sphere (because $M \to L$ induces an isomorphism on π_1). Cut the N_i along these 2-spheres if necessary, and call the new pieces N_i. By Kneser's Conjecture (Theorem 3.9), we can further decompose each N_i along essential spheres and disks into the indecomposable summands of $\pi_1(N_i)$. Now replace each such essential sphere or disk with a 1-handle, to get a new union (which again we relabel as N) consisting of the disjoint union of N_i and 1-handles, where each $\pi_1(N_i)$ is freely indecomposable in $\pi_1(N)$, where $\pi_1(N) \to \pi_1(M)$ is a split surjection, and where each $\pi_1(N_i)$ is mapped *injectively* (by incompressibility of the boundary) into a conjugate of some indecomposable G_j.

On the other hand, since the surjection is split, each indecomposable G_j maps injectively by the splitting map into some $\pi_1(N_i)$, and thus after relabeling and discarding some N_j if necessary, we can assume $\pi_1(N_i)$ maps isomorphically to a conjugate of G_i, for each i. □

Without loss of generality, we assume that $\pi_1(M)$ has some factor G_1 which is not $\mathbb{Z}$. Ignoring basepoints, each $\pi_1(N_i)$ maps to the conjugacy class of G_i. If we choose a basepoint e on N_1, then we can choose an embedded arc γ_i from e to N_i (*not* assumed to be disjoint from the N_j) and for each free $\mathbb{Z}$ factor of $\pi_1(M)$ choose an embedded loop δ_j, so that the γ_i and δ_j are disjoint except at e, and we have isomorphisms $\pi_1(N_i \cup \gamma_i, e) \to G_i$ for each i, and $\pi_1(\delta_j, e) \to \mathbb{Z}$ an isomorphism onto $\mathbb{Z}$ factors of $\pi_1(M)$.

Thus, we ultimately obtain N, the union of the N_i with 1-handles, and $\pi_1(N) = \pi_1(M) * F_r$ for some r. Actually, it is convenient (for the sake of an inductive argument) to allow N to be the union of N_i with a thickened neighborhood of some graph (which might contain vertices and loops), and with $\pi_1(N) = \pi_1(M) * F_r$, and $\pi_1(N) \to \pi_1(M)$ surjective. Recall that a 3-manifold obtained by thickening a graph is called a handlebody (see Definition 3.20). If F is the surface witnessing the free factorization of $\pi_1(M)$ (as in the proof of Lemma 3.18) then we can further assume that F meets the handlebody part of N transversely in s meridian disks. Note that each collection C of components of $N - F$ contained in a component of $M - F$ is associated to a factor G_j or $\mathbb{Z}$ of $\pi_1(M)$, and in the latter case C consists of a union of handlebodies, while in the former case C consists of some handlebodies together with with N_j union some 1-handles. Let h denote the number of nontrivial handlebody pieces contained in the C associated to all G_j factors, and define a complexity to be the ordered pair $(r + s, h)$. Now choose N as above minimizing $(r + s, h)$.

First we prove a lemma about homomorphisms from the fundamental groups of handlebodies to $\mathbb{Z}$.

Lemma 3.19. *Let H be a handlebody, and let $\phi : \pi_1(H) \to \mathbb{Z}$ be a homomorphism with nontrivial kernel (for instance, if the genus of H is at least 2). Then there is a handlebody $H' \subset H$ of smaller genus than H such that $\pi_1(H') \to \pi_1(H) \to \mathbb{Z}$ has the same image as ϕ.*

Proof. Let Γ be a core graph of H. If some embedded loop γ in Γ is conjugate into the kernel of $\pi_1(\Gamma) \to \mathbb{Z}$, then if p is a point in an edge of γ, any arc in Γ passing over p can be replaced with an arc going around γ the other way. Thus if H' is obtained by cutting H along the disk dual to p, it satisfies the conclusions of the lemma. If H has genus 1, the core circle is in the kernel, and we are done. If the genus of H is at least 2, and γ_1, γ_2 are two embedded loops in a core rose Γ for H, then either one of the γ_i maps to 0 in $\mathbb{Z}$, or they map to $a, b \in \mathbb{Z}$; sliding γ_1 over γ_2 replaces a by $a \pm b$, so after finitely many such slides, we may assume that one of the γ_i maps to 0, and we are done in this case too. □

If $\pi_1(N) \to \pi_1(M)$ is not an isomorphism, there is some kernel, and we can find a nontrivial embedded loop $\beta \subset N$ conjugate into the kernel, and intersecting F in the least number of points. There are two cases to consider.

(Case 1:) β is disjoint from F. In this case, β is contained in some submanifold of $N - F$. Let C be the union of pieces of $N - F$ in some component of $M - F$. If C is associated to a $\mathbb{Z}$ factor, it is a union of handlebodies, and for some handlebody H the map $\pi_1(H) \to \mathbb{Z}$

has a nontrivial kernel. By Lemma 3.19, one of these handlebodies can be replaced by a simpler handlebody (thereby reducing r) without affecting the image of $\pi_1(N)$.

So C is associated to some G_j, and is a disjoint union of handlebodies and N'_j, which is N_j together with some 1-handles. Since $\pi_1(N_j) \to \pi_1(G_j)$ is an isomorphism, if β is contained in N'_j there must be at least one 1-handle attached to N_j in C. But if $\alpha' \subset N'_j$ based at a point $e_j \in N_j$ runs over the core of this 1-handle, then since the image of $\pi_1(N'_j, e)$ is contained in G_j, which is equal to the image of $\pi_1(N_j, e)$, it follows that there is some $\alpha \subset N_j$ also based at e_j such that α and α' have the same image in G_j. Thus we can remove the 1-handle from N'_j to obtain a new N with smaller r, and still surjecting onto $\pi_1(M)$.

Finally, if β is contained in some handlebody H in C, we can join H to N'_j by a 1-handle (which increases r by 1), then compress away all the 1-handles of H as above. Since β is contained in H, it follows that H has positive genus, so the net result of this modification does *not* increase r (or s); however it *does* decrease h. So the complexity is reduced.

(Case 2:) β intersects F. In this case, let D be an immersed disk mapping to M with $\partial D = \beta$, intersecting F transversely. By homotopy we can eliminate loops of intersection, so that $D \cap F$ is a system of proper arcs, and by hypothesis we assume this system is nonempty. An innermost arc σ' of $D \cap F$ cobounds a bigon of D with an arc $\beta' \subset \beta$ contained in some component C of $N - F$. If the endpoints of β' are contained in the same component of $C \cap F$, let α be an arc in $C \cap F$ with the same endpoints as β'. Then $\beta' \cup \alpha$ and $(\beta - \beta') \cup \alpha$ are loops in N, each homotopically trivial in M (because every loop in F is inessential in M) and each intersecting F in fewer points than β after an isotopy, contrary to the definition of β. Thus we may assume that the endpoints of β' are contained in *different* components of $C \cap F$. These different components are the ends of two thickened arcs σ_1, σ_2 contained in 1-handles of N. We modify N by cutting out σ_1 (say) and replacing it with σ', which we then push off into the side of F not containing C, and perturb to be an embedded arc σ; effectively, we "push" σ_1 over the immersed bigon bounded by β' and σ', and then perturb the result to be embedded. The new N has isomorphic fundamental group to the old one, and fewer intersections with F. Moreover, $\pi_1(N)$ still surjects onto $\pi_1(M)$, since any loop in the old N which ran over σ_1 can be homotoped across the bigon into σ. This completes the proof of the Scott Core Theorem.

3.6. **Heegaard splittings.** In this section we give another proof of the existence of prime decomposition; this depends on a theorem of Haken on the interaction of reducing spheres with Heegaard splittings, and we give an elementary proof of this theorem due to Jaco [15]. Along the way we introduce the notion of a *Heegaard splitting*, one of the fundamental ways to build and present 3-manifolds.

Definition 3.20. A *handlebody* M is a 3-manifold obtained by taking a regular neighborhood of a connected graph Γ in S^3; it depends up to homeomorphism only on the homotopy type of Γ, which is to say on its Euler characteristic. If $\pi_1(M) = \pi_1(\Gamma)$ is free of rank g, we say M has *genus* g. Note in this case that ∂M is a closed, oriented surface of genus g.

Note that a handlebody can be realized as a submanifold of S^3 with connected complement, and therefore it is irreducible, by Alexander's Theorem.

Definition 3.21. A *Heegaard splitting of genus* g of a closed, oriented 3-manifold M is a decomposition $M = H_1 \cup_\psi H_2$, where the H_i are handlebodies of genus g, and they are glued by an orientation-reversing homeomorphism $\psi : \partial H_1 \to \partial H_2$.

Heegaard splittings were introduced by Heegaard in 1898, although he did not prove any significant theorems about them. Heegaard made the following elementary observation:

Proposition 3.22. *Every closed, oriented 3-manifold admits a Heegaard splitting of some genus.*

Proof. Let τ be a triangulation of M, and let τ' denote the dual cell decomposition. Then a neighborhood of the 1-skeleton $N(\tau^1)$ is a handlebody, and $M - N(\tau^1) = N((\tau')^1)$ is another handlebody, thus exhibiting the desired structure. □

Example 3.23. There is only one way to glue a sphere to itself, so the only 3-manifold with a Heegaard splitting of genus 0 is S^3. If $M = T_1 \cup_\psi T_2$ is a Heegaard splitting of genus 1, the meridian of T_2 is attached to the (p,q) curve on the torus ∂T_1 for some coprime p,q and after attaching a disk along this curve there is only one way to attach the remaining 3-ball. Thus we obtain the *Lens space* $L(p,q)$ (see Example 1.8) which is covered by S^3 with deck group $\mathbb{Z}/p\mathbb{Z}$ when $p \geq 1$, and is equal to $S^2 \times S^1$ if $p = 0$.

Another way to prove the existence of Heegaard splitting is via Morse theory. If M is a 3-manifold, and f is a Morse function in which the index of the critical points is nondecreasing (one calls this an *ordered Morse function*), then some level set S of f separates the critical points of index 0 and 1 from the critical points of index 2 and 3. The subset H^- of M below this level set is made from 0 and 1 handles; thus its core is a graph, and the subset is a handlebody. Conversely, the subset H^+ of M above this level set is made from 2 and 3 handles; thus its cocore is a graph, and it too is a handlebody.

Cerf showed (see Cerf [6], Chap. V or Laudenbach [17] for an elementary proof) that any two ordered Morse functions may be joined by a path of ordered functions which are Morse except at finitely many times when a pair of critical points with adjacent indices is created or canceled (one calls this a "birth" or "death" process). At the level of Heegaard splittings, the cocore of a 1-handle is a disk D_1 in H^- with boundary a compressing loop γ_1 on S, and the core of a 2-handle is a disk D_2 in H^+ with boundary a compressing loop γ_2 on S. The pair can be canceled if γ_1 and γ_2 intersect transversely in one point, thus eliminating a pair of critical points and reducing the genus of the splitting by one. The inverse operation is called *stabilization* of a Heegaard splitting. There is only one way to stabilize a Heegaard splitting (up to homeomorphism), because the creation of a canceling pair of handles is a local operation; but inequivalent splittings of the same manifold *can* become equivalent after stabilization. Thus Cerf's theorem implies the Theorem of Reidemeister–Singer, which says that any two Heegaard splittings of a fixed 3-manifold become equivalent after finitely many stabilizations. We shall not prove this theorem here.

Let $M = H_1 \cup_\psi H_2$ be a Heegaard splitting of M of genus g. A *Heegaard reducing sphere* is a separating sphere S which intersects each H_i in a proper essential disk, the two disks meeting in an essential simple loop in the boundary surface. The sphere realizes M as a connect sum $M = M_1 \# M_2$ (not necessarily nontrivial), and each M_i gets an induced Heegaard splitting of genus g_i, where $g_1 + g_2 = g$. If the splitting is a nontrivial stabilization, with γ_1, γ_2 curves as above bounding disks on the two sides and meeting transversely in a point, then if α is the boundary of a regular neighborhood of the union $\gamma_1 \cup \gamma_2$ in the splitting surface, then α is the intersection of the Heegaard surface with a transverse sphere. If α is essential in the surface, this is a reducing sphere; otherwise the genus $g = 1$. Thus: a Heegaard splitting of genus $g > 1$ which is a nontrivial stabilization admits a (compressible) reducing sphere.

We introduce the idea of boundary compression:

Definition 3.24. A properly embedded surface $(S, \partial S) \subset (M, \partial M)$ is *boundary incompressible* if one of the following conditions holds:

1. S is a disk which does not cobound a ball with a disjoint disk in $\partial M - \partial S$; or
2. S is not a disk, and no essential simple proper arc α on S cobounds an embedded bigon D in $M - S$ with a proper arc β in $\partial M - \partial S$.

A properly embedded surface is boundary compressible if it is not boundary incompressible. Such a surface can be *boundary compressed*: in the first case, the disk can be pushed across the ball and eliminated; while in the second case, we can cut the surface S open along α and glue in two disjoint parallel copies of D. The result is a new properly embedded surface S' (possibly disconnected even if S was connected), each of whose components has bigger Euler characteristic than S. Note that if S is incompressible, the result S' of a boundary compression is also incompressible, since any compressing disk for S' can be properly isotoped so that its boundary is in $S \cap S'$ and its interior is disjoint from S. It follows that any properly embedded surface S in M can be repeatedly compressed and then boundary compressed until it is both compressible and boundary incompressible.

The following proposition is due to Haken [11], and is generally known as *Haken's Lemma*. The proof we give is due to Jaco:

Proposition 3.25 (Haken's Lemma)**.** *Let M be reducible. Then every Heegaard splitting of M is reducible.*

Proof. Let F be a Heegaard surface, splitting M into handlebodies $H^\pm$. Let S be a nontrivial reducing sphere in M with the property that its intersection $S \cap F$ is in general position, and with the least number of components. We will show that $S \cap F$ consists of exactly one component, which proves the proposition.

First, the number of components is positive, because handlebodies are irreducible. Next, observe first that S^+ is incompressible in H^+. For, otherwise, a compression replaces S with two new spheres, each of which intersects F in fewer components than S, and at least one of which is nontrivial.

Suppose R is a component of S^+. Then R is a planar surface. We would like to show that R is a disk. Suppose not. Choose a maximal collection of disjoint nonparallel compressing disks $\mathcal{D}$ for F in H^+ (these could be dual to the edges of a rose embedded in H^+ to which H^+ deformation retracts). By an isotopy, we can make $R \cap \mathcal{D}$ consist of a collection of essential arcs in R. Performing a collection of boundary compressions in these arcs (across bigons in the disks of $\mathcal{D}$) reduces R to a collection of incompressible pieces, each contained in a ball of $H^+ - \mathcal{D}$; since the pieces are incompressible, they are disks. Thus boundary compression along the system of arcs $R \cap \mathcal{D}$ reduces R completely to disks. Choose a minimal family of arcs A from $R \cap \mathcal{D}$ which cut R into disks. Each A either reduces b_1

or separates. Compressing along a nonseparating arc decreases the number of boundary components by 1, while compressing along a separating arc increases it by 1. We claim that there will be strictly more nonseparating compressions than separating ones, and prove this by induction. The base case is the annulus, in which every essential arc is nonseparating, so this is obviously true. If we do a separating compression, we produce two new surfaces R_1, R_2, and for each by induction we will have an excess of nonseparating compressions; thus the excess is at least 2, and subtracting 1 for the compression producing the R_i from R completes the induction step and proves the claim. But now the result of all these boundary compressions isotopes S to S' which intersects F in fewer pieces. So R was a disk after all.

But if every component of S^+ is a disk (and similarly for S^-) then S is a union of spheres, one for each component of $S \cap F$, so this intersection has exactly one component, and S is a Heegaard reducing sphere for the splitting. □

Thus if M is reducible, we can write $M = M_1 \# M_2$ nontrivially so that each of the M_i has strictly smaller Heegaard genus than M. It follows that this process terminates after finitely many steps (and shows that the number of prime summands is bounded by the minimal Heegaard genus for M).

In general, the collection of Heegaard splittings of an irreducible 3-manifold is very mysterious. But at least in the particular case of the 3-sphere, the situation is as simple as it could be, as shown by Waldhausen [41]:

Proposition 3.26 (Waldhausen). *Every Heegaard splitting of S^3 is* standard*; i.e. it is obtained by stablizing the unique splitting of genus* 0*. Thus there is exactly one Heegaard splitting of S^3 of each genus $g \geq 0$.*

Proof. When $g = 0$ there is nothing to prove, and when $g = 1$ this is just the observation that a knot in S^3 with a neighborhood whose boundary compresses to the outside is the unknot (which follows from Dehn's Lemma). So we can assume the genus $g > 1$.

The following argument is due to Rieck [28], simplifying an argument developed by Rubinstein–Scharlemann [29], and depending on a Theorem of Casson–Gordon whose proof we shall defer until a later section. Let $x : S^3 \to [-1, 1]$ be a Morse function with a single maximum and minimum, and level sets all spheres S_x, and let $y : S^3 \to [-1, 1]$ be a "sweepout", for which $y^{-1}(1)$ is the core of the "upper" handlebody, $y^{-1}(-1)$ is the core of the "lower" handlebody, and each other level set Σ_y is isotopic to the splitting surface. Cerf theory (see [6] and [29]) says that we can find such functions x and y such that the subset Γ of $[-1, 1] \times [-1, 1]$ for which the level sphere of x and the level surface of y are not transverse is a 4-valent graph Γ (with isolated vertices on the boundary); moreover, on points on the edges of Γ the surfaces S_x and Σ_y intersect in a single nondegenerate critical point (a circle or saddle), and at vertices of Γ the intersections have two critical points.

Give a complementary region R to Γ the label I if the components of $\Sigma_y \cap S_x$ are inessential in Σ_y for all (x, y) in R, and give it the label E otherwise. Then there is either a chain of adjacent E regions joining $y = -1$ to $y = 1$, or a chain of I regions joining $x = -1$ to $x = 1$ either adjacent or meeting at a vertex.

For each nonsingular (x, y), the regions of $\Sigma_y - S_x$ are "left" or "right" according to which side of S_x they are on. If (x, y) is in an I-region (equivalently, if $\Sigma_y \cap S_x$ consists only of inessential loops in Σ_y) then exactly one of these sides contains pieces of positive genus. Thus, I regions are either IL or IR. Near $x = -1$ all I regions are IR, and near $x = 1$ all I regions are IL. Passing over an edge of Γ cannot change an IL region to an IR region, so such regions are never adjacent. Similarly, passing through a vertex can move at most one "handle" from the L side to the R side, so it can't change an IL region to an IR region unless $g \le 1$.

We conclude therefore that there must be a chain of adjacent E regions joining $y = -1$ to $y = 1$. Now, for each nonsingular (x, y) in an E region, some loop of $\Sigma_y \cap S_x$ is essential in Σ_y, and we see that an innermost such loop bounds a compressing disk for Σ_y on at least one side, either the "positive" side or the "negative" side. Near $y = 1$ the Σ_y collapse to a graph on the positive side, so there must be positive compressing disks in E regions here; similarly, near $y = -1$ there must be negative compressing disks in E regions. Since there is a chain of adjacent E regions, either there is a single E region with both positive and negative compressing disks, or there are adjacent E regions, one with positive and the other with negative compressing disks. When we pass over an edge of Γ corresponds to passing through a single critical point; thus the loops of intersection before and after passing through this singularity can be realized disjointly in Σ (up to isotopy). Thus we conclude (finally) that there are disjoint essential loops on Σ which bound compressing disks for the handlebodies on opposite sides. Such a Heegaard splitting is said to be *weakly reducible*.

The Theorem of Casson–Gordon (which we shall prove in the sequel) says that if a closed orientable 3-manifold M admits a Heegaard splitting which is weakly reducible, then M contains a 2-sided embedded incompressible surface; such a surface is either a reducing sphere (in which case the Heegaard splitting is reducible) or else has infinite π_1; since 2-sided embedded incompressible surfaces are π_1-injective, the latter situation cannot occur in S^3. So we conclude that every Heegaard surface in S^3 of genus at least 2 is reducible.

Since the reducing sphere is inessential in S^3 by Alexander's Theorem, we see that the original Heegaard splitting is obtained from two splittings of S^3 of lower genus; by induction, each of these is standard, and therefore so is the original splitting. ☐

Note that it is not at all easy to *recognize* a standard handlebody in S^3; an example of a (geometrically) interesting genus 2 handlebody in S^3 determinining a (topologically) standard Heegaard splitting is given in Figure 6.

3.7. **Group theory and the Poincaré Conjecture.**

In [36], Stallings showed that the Poincaré Conjecture follows from a purely group theoretic statement; in [13], Jaco showed that the group theoretic statement and the Poincaré Conjecture are *equivalent*. Thus, Perelman's proof also proves this group theoretic statement.

If H is a handlebody of genus g, then ∂H is a closed oriented surface of genus g, and the map $\pi_1(\partial H) \to \pi_1(H) = F_g$ induced by inclusion is surjective. Jaco proved the somewhat surprising fact that all surjective maps between these groups arise in this way:

Lemma 3.27 (Jaco). *Let Σ_g be a closed oriented surface of genus g and Γ_h a wedge of h circles, and let $f : \Sigma_g \to \Gamma_h$ induce a surjection in π_1. Then $h \le g$, and if $h = g$ there is a homotopic map $f' : \Sigma_g \to \Gamma_g$ whose mapping cylinder is a handlebody of genus g.*

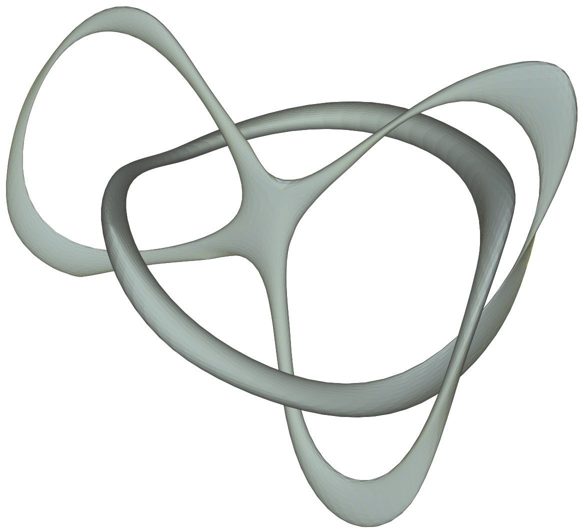


FIGURE 6. A genus 2 handlebody in S^3 with handlebody complement.

Proof. This lemma is proved by a variation on Stallings' proof of Grushko's Theorem and Kneser's Conjecture (Theorem 3.7 and Theorem 3.9) one dimension lower. Write Γ_h as $S^1 \vee \Gamma_{h-1}$, and by a homotopy replace the basepoint by an interval joining the S^1 and Γ_{h-1} factors, and let p be the midpoint of the interval. Make f transverse to p by a homotopy, so that $f^{-1}(p)$ is a collection of circles in Σ_g whose union L is homologically trivial.

If L consists of a single circle, we can compress Σ_g along this circle and cut Γ_g along p, and reduce to two subsurfaces with genus summing to g, mapping to S^1 and Γ_{h-1} respectively, surjectively on π_1. We claim that after a homotopy of f, we can always assume that L consists of a single circle.

First, L is nonempty, or else the image of $\pi_1(\Sigma_g)$ would be properly contained in $\pi_1(\Gamma_h)$, contrary to hypothesis. If L contains at least two components, there is an arc α in $\Sigma_g - L$ running from one component L_1 to a different component L_2, and by exactly the same algebraic argument as in the proof of Theorem 3.7 we can assume (after replacing α by a possibly different arc) that $f(\alpha)$ is a homotopically trivial loop on one side of the wedge.

The key difference between this situation and the proof of Theorem 3.7 is that *a priori* α might not be embedded. So let p be a point of self-intersection of α bounding an embedded arc $\sigma \subset \alpha$ on one side, ending on the loop L_1, and let σ' be a small embedded arc contained in the interior of α, and crossing the rest of α only at p. Let $\sigma' = \sigma_1'\sigma_2'$, where the σ_i are the parts on either side of p. Build a new arc $\sigma'' = \sigma_1'\sigma L_1 \sigma^{-1}\sigma_2'$; i.e. σ'' is obtained from σ' by pushing it over σ, and adding a copy of L_1 at the end. The relative homotopy class of σ'' is obtained from that of σ' by inserting L_1, but $f(L_1) = p$ so the relative homotopy class of $f(\sigma'')$ in Γ_h is the same as that of $f(\sigma')$. Moreover, by a small homotopy, we can push σ'' off σ, eliminating the point p of self-intersection. After finitely many such moves, we obtain a new *embedded* arc α' from L_1 to L_2, so that $f(\alpha)$ is a homotopically trivial loop in Γ_h on one side of p, and now the argument of Theorem 3.7 goes through one dimension lower to build a new map homotopic to f such that $f^{-1}(p)$ has one fewer point of intersection.

The argument then reduces to the case that $h = 1$; obviously then $g \geq 1$, and if $g = 1$ it is elementary that we can homotop the map so that the mapping cylinder is a solid torus. This proves the lemma. □

Theorem 3.28 (Jaco, Stallings)**.** *For any $g > 1$ let Σ_g denote the closed, oriented surface of genus g, and let F_1 and F_2 be free groups of rank g. Let $\phi : \pi_1(\Sigma_g) \to F_1 \times F_2$ be a surjective homomorphism. Then either of the following conclusions is equivalent to the truth of the Poincaré Conjecture:*

1. *some nontrivial element of $\ker(\phi)$ is represented by an essential simple closed curve on Σ_g; or*
2. *the map ϕ can be factored through an essential map of $\pi_1(\Sigma_g)$ into some nontrivial free product.*

Proof. First we show the two conclusions are equivalent. Suppose γ is an essential simple closed curve on Σ_g in the kernel. If γ is separating, then Σ_g/γ is a wedge of two surfaces of smaller genus, and ϕ factors through the fundamental group of the wedge, which is a nontrivial free product. If γ is nonseparating, let α be simple and intersect γ transversely in a single point. Then $[\gamma, \alpha]$ is simple and in the kernel, and is essential if $g > 1$. Conversely, suppose ϕ factors through a map ψ to $G_1 * G_2$. Let $K = K(G_1, 1) \vee K(G_2, 1)$ and realize ψ by $f : \Sigma_g \to K$ transverse to the basepoint p. Compress inessential components of the preimage until we get an essential embedded loop in the preimage. This shows the equivalence of the two conclusions.

Now, let $\phi_i : \pi_1(\Sigma_g) \to F_i$ be two surjective homomorphisms to free groups, and let $G := \pi_1(\Sigma_g)/\ker(\phi_1) \cdot \ker(\phi_2)$. We claim that $\phi := \phi_1 \times \phi_2$ is surjective if and only if G is trivial. First, let $\psi_i : F_1 \times F_2 \to F_i$ be projection to a factor. Then $F_1 \times F_2 = \ker(\psi_1) \cdot \ker(\psi_2)$. But if ϕ is surjective, then $\pi_1(\Sigma_g) = \ker(\phi_1) \cdot \ker(\phi_2)$ so that G is trivial. Conversely, if G is trivial, $\pi_1(\Sigma_g) = \ker(\phi_1) \cdot \ker(\phi_2)$. Let $(\alpha, \beta) \in F_1 \times F_2$ be arbitrary, and let $\phi_1(\alpha') = \alpha$ and $\phi_2(\beta') = \beta$. Using $\pi_1(\Sigma_g) = \ker(\phi_1) \cdot \ker(\phi_2)$ we write $\alpha' = \alpha'' x$ and $\beta' = y\beta''$. Then $\phi(\alpha''\beta'') = (\alpha, \beta)$, so that ϕ is surjective.

We now show that the truth of the first conclusion for all $g > 1$ implies the Poincaré conjecture. Let M be a simply connected closed 3-manifold, and let $M = H_1 \cup_\phi H_2$ be a Heegaard splitting of minimal genus g. If $g = 1$ then M is a Lens space, and is not a counterexample. Otherwise, the inclusion of the splitting surface Σ_g into the two sides gives two surjective homomorphisms $\phi_i : \pi_1(\Sigma_g) \to \pi_1(H_i) =: F_i$ which are the factors of a map $\phi : \pi_1(\Sigma_g) \to F_1 \times F_2$. The hypothesis that M is simply connected implies that ϕ is surjective as above, so the first conclusion says we can find a simple loop γ on Σ_g in the kernel of both ϕ_i. Dehn's Lemma implies that γ bounds *embedded* disks in either handlebody, and therefore we can exhibit M as a nontrivial connect sum $M = M_1 \# M_2$ where each M_i has a Heegaard splitting of strictly smaller genus. Thus by induction, the Poincaré Conjecture is true.

Conversely, suppose the Poincaré Conjecture is true, and let $\phi : \pi_1(\Sigma_g) \to F_1 \times F_2$ be surjective, so that each factor $\phi_i : \pi_1(\Sigma_g) \to F_i$ is surjective. Let Γ_i be a wedge of g circles, and identify $\pi_1(\Gamma_i)$ with F_i. By Lemma 3.27 we can represent the ϕ_i by $f_i : \Sigma_g \to \Gamma_i$ whose mapping cylinders are handlebodies, and the union of the two mapping cylinders is a 3-manifold M with a Heegaard splitting of genus g. By the argument above, surjectivity of ϕ implies that M is simply-connected; thus M is homeomorphic to S^3. Then Proposition 3.26

implies that the Heegaard splitting of M is reducible, and the reducing sphere intersects the splitting surface Σ_g in an essential simple closed curve. □

4. Haken manifolds

Definition 4.1. A compact orientable irreducible 3-manifold is *Haken* if it contains a properly embedded 2-sided essential surface.

Remark 4.2. The term *sufficiently large* is sometimes used as a synonym for "Haken". One also sometimes considers a non-orientable generalization of Haken 3-manifolds, in which case one insists that they should contain no 2-sided projective plane.

If M is Haken, and S is a 2-sided essential surface in M, we may cut M along S to produce a new manifold M'. If S' is a 2-sided essential surface in M', we may cut M' along S', and so on. Note that if M is irreducible, so is M', since if Σ were an essential 2-sphere in M' that bounded a ball B in M but not in M', the surface S would necessarily be contained in the interior of B, which is absurd since S is essential.

Definition 4.3. If M is a Haken 3-manifold, a *partial hierarchy* for M is a sequence

$$M = M_0 \xrightarrow{S_0} M_1 \xrightarrow{S_1} \cdots \xrightarrow{S_{n-1}} M_n$$

where

1. each S_i is a properly embedded 2-sided essential surface in M_i which is not boundary parallel; and
2. each M_{i+1} is obtained from M_i by cutting along S_i.

If further M_n is a disjoint union of 3-balls then we call the sequence a *hierarchy* for M.

The main significance of Haken manifolds is that they all admit hierarchies, and this opens up the possibility of proving theorems about Haken manifolds by induction.

4.1. Manifolds with boundary.

Suppose M is irreducible with nonempty boundary. If some component of ∂M is compressible, then a compressing disk is an essential surface in M, so that M is Haken and the result of this compression is a new irreducible 3-manifold whose boundary has smaller complexity. So we can find a partial hierarchy for M by splitting inductively along disks, until all boundary components are incompressible. If every boundary component is a sphere, then since M is irreducible, we are left with a union of balls. Otherwise, some boundary component is incompressible of positive genus.

We now show that if M is Haken with boundary, then it contains a *nonseparating* properly embedded essential surface. First we prove a lemma relating the homology of M and ∂M:

Lemma 4.4 (Lagrangian subspace)**.** *Let M be a compact oriented 3-manifold. Then the kernel L of the inclusion homomorphism*

$$L \to H_1(\partial M) \to H_1(M)$$

is a Lagrangian subspace of $H_1(\partial M)$ with its intersection pairing. In particular, it has dimension equal to half of the dimension of $H_1(\partial M)$.

Proof. Integral elements α, β in L are of the form ∂A, ∂B for integral classes A and $B \in H_2(M, \partial M)$. These classes are dual to classes in $H^1(M)$ which are represented by homotopy classes of maps from M to S^1. A generic preimage is a proper 2-sided embedded surface, so we get surfaces S_A, S_B representing A, B, and whose boundaries represent α, β. If we put S_A and S_B in general position, they intersect in a family of oriented circles and intervals; each interval runs from a positive to a negative crossing in $\partial S_A \cap \partial S_B$, and all points of intersection arise this way, so $\alpha \cap \beta = 0$. Thus we have shown that the intersection pairing is trivial on L.

Now, by Poincaré duality, the intersection pairing of $H_1(M)$ with $H_2(M, \partial M)$ is nondegenerate. The quotient $H_1(\partial M)/L$ can be identified with its image in H_1, so every nonzero class β in $H_1(\partial M)/L$ pairs nontrivially with some $A \in H_2(M, \partial M)$, represented by a surface S_A. But then ∂S_A represents $\alpha \in L$ which pairs nontrivially with β. Thus L surjects onto the dual of $H_1(\partial M)/L$, and its dimension is at least half of $H_1(\partial M)$. So we conclude that L is Lagrangian, as claimed. □

Proposition 4.5. *Suppose M is Haken with nonempty boundary. Then M contains a nonseparating properly embedded essential surface.*

Proof. Every sphere component of ∂M bounds a ball, so some component is not a sphere. But then the kernel of $H_1(\partial M) \to H_1(M)$ is nonempty, and comes from some class $A \in H_2(M, \partial M)$ represented by a nonseparating properly embedded essential surface S_A. □

Using results from previous sections, we can give a characterization of Haken manifolds in terms of their fundamental groups:

Proposition 4.6. *An irreducible oriented 3-manifold M is Haken if and only if $\pi_1(M)$ splits as a nontrivial amalgamated free product or HNN extension; equivalently, if and only if $\pi_1(M)$ acts minimally on a nontrivial tree T without inversions.*

Proof. The equivalence of the two conclusions is Lemma 2.3.

Going in one direction, if M is Haken, and S is a 2-sided essential surface properly embedded in M, then by Kneser's Lemma, S is π_1-injective. Either M is closed and then we can split along S, or M has boundary, and by Proposition 4.5 we can find a nonseparating properly embedded essential surface. In either case, we can assume the splitting of $\pi_1(M)$ coming from S is nontrivial (the case that S is a compressing disk is possible, of course).

Conversely, suppose $\pi_1(M)$ acts on a tree T. Then M contains a 2-sided essential properly embedded surface S whose fundamental group injects into an edge stabilizer, by Proposition 2.7. □

4.2. **Kneser-Haken finiteness.** Recall Kneser's proof of the existence of a prime decomposition (i.e. Theorem 3.5), which argued by simplifying the intersection of a collection of pairwise disjoint essential spheres with the 2-skeleton of a triangulation of M. Essentially the same argument proves the following proposition, which is usually known as *Kneser-Haken finiteness*, and first proved by Haken. We do not prove this theorem in the most general possible form; in particular, we require the surfaces to be *closed*. For a proof where the surfaces are allowed to have boundary, see Jaco [15], Thm. III.20.

Theorem 4.7 (Kneser-Haken finiteness). *Let M be a compact irreducible 3-manifolds. Then there is a constant $h(M)$ so that if $S := S_1 \cup \cdots \cup S_n$ is a union of disjoint closed incompressible surfaces in M, and $n > h(M)$, then at least two of the S_i are parallel.*

Proof. The proof follows Kneser's argument from § 3.3 very closely. First, let τ be a triangulation of M, and isotop S so that its intersection with τ minimizes the complexity (e, f), where e is the number of points of intersection of S with τ^1, and f is the number of components of intersection of S with faces of τ^2.

For a minimal configuration, there are no loops of intersection of S with faces of τ, since the boundary of such a loop must be inessential in S (by the hypothesis that S is incompressible), and an innermost such loop can be eliminated by an isotopy (by the hypothesis that M is irreducible).

Similarly, for a minimal configuration, there are no arcs of intersection of S with faces of τ with both endpoints on one edge, or an innermost arc could be eliminated by isotopy, reducing the complexity of the intersection.

Finally, for σ a simplex, all components of $S \cap \sigma$ are disks. For, the boundary of a non-disk component must be inessential in S (by the hypothesis that S is incompressible), and an innermost such loop can be eliminated by an isotopy (again, because M is irreducible).

Thus, as before, for each simplex of σ the boundary $\partial\sigma$ is split into pieces, at most six of which are bad, and all the rest of which are annuli bounding an I-bundle over a disk in $\sigma - S$. Thus, if τ has t simplices, at most $6t$ components of $M - S$ can contain a bad piece, and all the rest are I-bundles. Each 1-sided I-bundle contributes a nontrivial summand to $H_1(M; \mathbb{Z}/2\mathbb{Z})$ as before, and we are done. □

4.3. **Existence of hierarchies.** One of the most important applications of Theorem 4.7 is to the existence of hierarchies. But first, it is important to impose some additional restrictions on the kind of partial hierarchies we allow. In particular, it is important to insist that the splitting surfaces S_i are not just incompressible but also *boundary incompressible*.

We already saw in the proof of Proposition 3.25 that any surface S in a handlebody which is incompressible and boundary incompressible is a disk. But handlebodies contain many interesting incompressible surfaces, and these are the origin of some very interesting phenomena.

Example 4.8. Let M be a handlebody of genus 4, and write $M = S \times I$ where S is a once-punctured surface of genus 2. Let α be an essential nonseparating embedded loop in S. For any n we can take the surfaces $S \times i/n$ for $0 < i < n$, cut each surface open along $\alpha \times i/n$, and join one side of $(S - \alpha) \times i/n$ to the opposite side of $(S - \alpha) \times (i+1)/n$ by a "vertical" annulus of the form $\alpha \times [i/n, (i+1)/n]$ (the two boundary annuli each with one component on ∂M). This produces an essential surface S_n in M of genus n.

Example 4.9. Let P be a pair of pants, and let $M = P \times I$. Then we can also write $M = T \times I$ where T is a once-punctured torus. If α is an essential embedded loop in T then $\alpha \times I$ is an incompressible (but *not* boundary incompressible) annulus in M. Cutting along $\alpha \times I$ produces a new manifold M' which is homeomorphic to $(T - \alpha) \times I) = P \times I = M$. Thus we may continue this procedure indefinitely, and obtain a partial hierarchy of any length.

From now on, we only consider partial hierarchies where the decomposing surfaces S_i are both incompressible and boundary incompressible. The next proposition shows that the phenomenon in Example 4.9 cannot occur for such hierarchies:

Proposition 4.10. *There is a constant $c(M)$ depending only on M, so that if $M := M_0 \xrightarrow{S_0} \cdots M_n$ is a partial hierarchy where every S_i is incompressible and boundary incompressible, there are at most $c(M)$ surfaces S_i which are not disks.*

Proof. For each M_i, let M_i' be the result of compressing the boundary of M_i as much as possible, and let R_i be the union of non-sphere components of $\partial M_i'$. Then either M_i is a union of handlebodies (in which case all incompressible and boundary incompressible surfaces in M_i are disks, and all M_j with $j > i$ are handlebodies or balls) or R_i is nonempty and incompressible in M_i', so that $\pi_1(R_i)$ injects in $\pi_1(M_i')$. But since M_i' is obtained from M by inductively splitting along incompressible surfaces, from Seifert van-Kampen we see that $\pi_1(R_i)$ injects in $\pi_1(M)$, so that R_i is essential in M.

If S_i is not a disk, then because it is incompressible and boundary incompressible, we can isotop it into M_i'. It follows that the R_i are all disjoint, and there is at least one for each non-disk S_i.

By Kneser-Haken finiteness (i.e. Theorem 4.7), there is a constant $c(M)$ so that if there are more than $c(M)$ non-disk S_i, then there are at least 4 parallel R_i. It follows that there is some S_i which can be isotoped into $\partial M_i'$. But M_i is obtained from M_i' by attaching finitely many 1-handles, so there is some $S_i' \subset S_i$, obtained by removing finitely many disks from S_i, so that S_i' is isotopic into ∂M_i. If S_i is not a disk, then S_i' contains some essential proper arc which is also essential in S_i; isotoping S_i' into ∂M_i therefore certifies that S_i was boundary compressible (along the arc in question), contrary to hypothesis. □

Proposition 4.10 does not yet give an *a priori* bound on the length of a partial hierarchy, but it is enough to deduce the existence of a hierarchy:

Theorem 4.11 (Hierarchy exists)**.** *Let M be Haken. Then any maximal partial hierarchy in which every surface is boundary incompressible must terminate. In particular, M admits a hierarchy.*

Proof. Consider any partial hierarchy for M, which ends in M_n. Any incompressible and boundary incompressible surface is either a compressing disk for the boundary or not. Each compression of the boundary reduces its complexity, so there is a bound (depending on M_n) on the number of disk surfaces in any extension of the partial hierarchy, before we must either be left with a collection of balls, or we must cut along a non-disk component. By Proposition 4.10 there is a bound (depending only on M) on the number of non-disk components in any partial hierarchy. Thus any maximal partial hierarchy as above must terminate at some M_n. Since M is Haken, every component of M_n has nonempty boundary; but then every boundary component is a sphere (or else we could extend the partial hierarchy) and therefore M_n is a union of balls, so that the partial hierarchy is actually a hierarchy. □

The next lemma lets us rearrange the order of surfaces in a partial hierarchy so that all the "interesting" surfaces appear first. Combining this lemma with Proposition 4.10 and Theorem 4.11 we see that any Haken manifold M admits a partial hierarchy of bounded length (which can be effectively computed from any triangulation of M) which ends at a collection of handlebodies.

Lemma 4.12 (Do disks last). *Let $M_0 \xrightarrow{S_0} \cdots M_n$ be a partial hierarchy. Then M_0 has another partial hierarchy of the same length terminating in M_n, in which all the splittings along disk components are done last.*

Proof. Suppose S_i is a disk, so that when we cut M_i along S_i to get M_{i+1}, we get two new disks $S_i^\pm \subset \partial M_{i+1}$. Whatever ∂S_{i+1} is, we can isotop it in ∂M_{i+1} to be disjoint from $S_i^\pm$. But this means that $\partial S_{i+1} \subset \partial M_i$ and is disjoint from ∂S_i, so we can switch the order of these two surfaces in the hierarchy. □

In fact, if one is prepared for our surfaces S_i to be disconnected, we can find very short hierarchies:

Proposition 4.13. *Let M be Haken. Then there is a hierarchy of length at most 4:*

$$M := M_0 \xrightarrow{S_0} M_1 \xrightarrow{S_1} M_2 \xrightarrow{S_2} M_3$$

Proof. Let S_0 be a maximal collection of pairwise disjoint essential closed surfaces. If a component of M_1 has compressible boundary, then it must be a handlebody, or else it would contain an essential surface (obtained as the boundary after compressing along a maximal collection of disks) disjoint from and not parallel into S_0. For each component of M_1 with incompressible boundary, we can find an essential nonseparating boundary incompressible surface, and call the union of such surfaces (one for each component) S_1. The result M_2 of cutting along S_1 must now be a union of handlebodies, since otherwise as above we could find a new essential surface to add to the collection S_0. Now each handlebody admits a maximal collection of disjoint compressing disks; the union over all components of M_2 is S_2, and the result $M_3 = M_2 - S_2$ is a union of balls. □

4.4. **Homotopy equivalences between Haken manifolds.** Homotopy equivalent 3-manifolds are not necessarily homeomorphic, as the following examples show.

Example 4.14 (Square and Granny knot complements). Let *square knot* is a connect sum of a right- and left-handed trefoil in S^3. The *Granny knot* is a connect sum of two left-handed trefoils. Denote their complements in S^3 by M_1 and M_2. Then $\pi_1(M_1)$ and $\pi_1(M_2)$ are equal, both groups being equal to the amalgam of two trefoil know complements along their meridional subgroups; i.e.

$$\pi_1(M_1) = \pi_1(M_2) = \langle x, y, z \mid xyx = yxy, xzx = zxz \rangle$$

Since the M_i are orientable and irreducible, by the sphere theorem they have π_2 trivial. Since they are non-compact, they have π_3 trivial. Thus both spaces are $K(\pi, 1)$s, and they are homotopy equivalent. However, they are not homeomorphic, although this is harder to see. One way to distinguish them is first to invoke a famous theorem of Gordon–Luecke [10] which says that knot complements in S^3 are homeomorphic if and only if the knots are isotopic. Then one must distinguish the square and Granny knots somehow; the simplest way is to use the *signature*, which is zero for the square knot, and is 4 for the Granny knot (or -4 for its mirror image).

Example 4.15 (Lens spaces). The Lens spaces $L(p,q)$ and $L(p',q')$ are homeomorphic if and only if $|p| = |p'|$ and $q' = \pm q^{\pm 1} \bmod p$. On the other hand, they are homotopy equivalent if and only if $|p| = |p'|$ and $qq' = \pm w^2 \bmod p$ for some w. Thus (for example), $L(7,1)$ and $L(7,2)$ are homotopy equivalent but not homeomorphic. Homotopy equivalent but not homeomorphic Lens spaces can be distinguished by Reidemeister torsion, which was invented by Reideimeister [27] precisely for this purpose.

However, it is a remarkable theorem of Waldhausen, that homotopy equivalences between Haken manifolds *which preserve the boundary structure* are (with some well-understood exceptions) homotopic to homeomorphisms. This theorem is proved inductively, using the hierarchical structure. We prove the following theorem:

Theorem 4.16 (Waldhausen). *Let M and N be Haken 3-manifolds, and suppose that $f : (M, \partial M) \to (N, \partial N)$ is injective on π_1, and furthermore the restriction $f : \partial M \to \partial N$ is injective on π_1 on each component. Then f is homotopic through maps of pairs to some map $g : (M, \partial M) \to (N, \partial N)$ for which one of the following holds:*

1. *$g : M \to N$ is a covering map;*
2. *M is an I-bundle over a closed surface, and $g(M) \subset \partial N$; or*
3. *N (hence also M) is a solid torus and $g : M \to N$ is a branched covering with branch set a circle.*

Moreover, if f restricted to a component of ∂M is already a covering map to its image in ∂N, we may assume the homotopy is constant on this component.

The theorem depends on a similar but simpler proposition for surfaces, which we state and prove first.

Proposition 4.17. *Let F and G be compact oriented surfaces with $\pi_1(F) \neq 0$. Let $f : (F, \partial F) \to (G, \partial G)$ be injective on π_1. Then f is homotopic through maps of pairs to $g : (F, \partial F) \to (G, \partial G)$ for which one of the following holds:*

1. *$g : F \to G$ is a covering map; or*
2. *F is an annulus and $g(F) \subset \partial G$.*

Moreover, if f restricted to a component of ∂F is already a covering map to its image in ∂G, we may assume the homotopy is constant on this component.

Proof. Assume ∂G is nonempty. Then since f_* is injective, ∂F is nonempty. Since $\pi_1(F) \neq 0$, boundary components of F are essential, so the restriction of f to each boundary component is homotopic to a covering map, and after replacing f by a homotopic map f', we can assume $f' : \partial F \to \partial G$ is a covering map on each component.

Let G' be the cover of G with fundamental group equal to $f_*(\pi_1(F))$, and let $f' : F \to G'$ be the lift of f. Then f' is an embedding on each component of ∂F. Suppose there are distinct components J_0, J_1 of ∂F with the same image $K \subset \partial G'$. Then we can find an essential arc α from J_0 to J_1 whose image $f'(\alpha)$ is a loop based at some point in K. By surjectivity of f' on π_1, we may modify α as above so that $f'(\alpha)$ is trivial in $\pi_1(G')$. Since f' induces an isomorphism on π_1, we deduce that the elements associated to J_0 and J_1 are conjugate in $\pi_1(F)$; but this immediately implies that F is an annulus, since otherwise killing J_0 (by attaching a disk, say) would produce a non-disk surface in which J_1 is still essential. So either we are in case (2), or we can assume f' restricted to ∂F is an embedding.

Choose a non-separating essential arc β in G'. The preimage $(f')^{-1}(\beta)$ consists of loops, together with exactly one proper arc (because f' is an embedding near the boundary). The loops are inessential in F, because f' is π_1-injective, so after a homotopy, we can assume the preimage is a single non-separating essential arc in F. After cutting along these arcs, we get a new map $f_1 : (F_1, \partial F_1) \to (G_1', \partial G_1')$ between simpler surfaces. Thus the proposition follows by induction (if we are careful about the base cases).

Now, if ∂G is empty, we can find a 2-sided nonseparating loop β in G, homotop the map so that the preimage consists of a collection of essential loops in F, then cut along β and its preimage and apply the argument above. □

We are now ready to give the proof of Theorem 4.16. We follow very closely the proof of the analogous Thm. 13.6 in [12].

Proof. Note that the condition that f should be injective on each component of ∂M is automatic if ∂M is incompressible.

By Proposition 4.17 we homotop f on each component of ∂M so that $f : \partial M \to \partial N$ is a covering map on each component. Let N' be the cover of N with fundamental group equal to $f_*(\pi_1(M))$, and let $f' : M \to N'$ be the lift of f. Suppose first that ∂M is nonempty, and $f' : \partial M \to \partial N'$ is not an embedding. Then we can find a proper arc α in M with distinct endpoints on ∂M whose image is a loop based at some point in $\partial N'$. Since f' is surjective on π_1, we can choose α in such a way that $f'(\alpha)$ is null-homotopic in N'. Let B_0, B_1 be the components of ∂M containing the endpoints of α, and let C be the component of N' they map to.

Now, f' is a covering map when restricted to each B_i, so the images of $\pi_1(B_0)$ and $\alpha_*\pi_1(B_1)$ in $\pi_1(C)$ both have finite index; in particular, they have finite index in each other, so the same is true for the image in $\pi_1(N')$. So if $\tilde{M}$ is the cover of M with fundamental group $\pi_1(B_0)$, the preimage $\tilde{B}_1$ of B_1 in $\tilde{M}$ is compact.

Suppose further that α is not homotopic rel. boundary into ∂M. Then $\tilde{B}_0$ is not equal to $\tilde{B}_1$. It follows that $\tilde{B}_0$ is incompressible in $\tilde{M}$, or else we could compress $\tilde{B}_0$ without changing $\tilde{B}_1$, contrary to the fact that both surfaces have fundamental groups which are finite index in $\pi_1(\tilde{M})$. Hence the inclusion of $\tilde{B}_0$ in $\tilde{M}$ induces an isomorphism in π_1, and is therefore a homotopy equivalence. Thus, $H_2(\tilde{M}; \mathbb{Z}/2\mathbb{Z}) = H_2(\tilde{B}_0; \mathbb{Z}/2\mathbb{Z}) = \mathbb{Z}/2\mathbb{Z}$. On the other hand, we have an exact sequence

$$0 \to H_3(\tilde{M}, \tilde{B}_0 \cup \tilde{B}_1; \mathbb{Z}/2\mathbb{Z}) \to H_2(\tilde{B}_0 \cup \tilde{B}_1; \mathbb{Z}/2\mathbb{Z}) \to H_2(\tilde{M}; \mathbb{Z}/2\mathbb{Z}) = \mathbb{Z}/2\mathbb{Z}$$

and $H_2(\tilde{B}_0 \cup \tilde{B}_1; \mathbb{Z}/2\mathbb{Z}) = \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$, so that $\tilde{M}$ is compact, and $\pi_1(B_0)$ has finite index in $\pi_1(M)$. From this, it follows on general grounds that M is an I-bundle over a closed surface, and we obtain conclusion (2).

So suppose that α as above is homotopic rel. boundary into ∂M. Since the image of α is a loop in C, and $B_0 \to C$ is a covering, this image is nontrivial in $\pi_1(C)$. But by hypothesis, the image of α *is* trivial in $\pi_1(N')$. Thus C is compressible in N'. We would like to argue that C is a torus. Notice that we may assume that f takes distinct components of ∂M to distinct components of $\partial N'$, or else we could have found α as in the previous case. Then $f'_* : H_2(\partial M) \to H_2(\partial N')$ is nonzero, since it is a covering map on each component, and different components have different images. But then $f'_* : H_3(M, \partial M) \to H_3(N, \partial N')$ is nonzero, so N' is compact, and every component of $\partial N'$ is in the image of some component of ∂M. Since f' is a homotopy equivalence,

$$\chi(\partial M) = 2\chi(M) = 2\chi(N') = \chi(\partial N')$$

On the other hand, labeling components of ∂M as B_i, and components of $\partial N'$ as C_i, if $f : B_i \to C_i$ has degree n_i, then $n_i\chi(C_i) = \chi(B_i)$. No component is a sphere, so it follows that $n_i = 1$ for each i unless B_i (and C_i) is a torus. Thus C as above is a torus, as claimed. Hence N' is a solid torus, and so is M, and we are in case (3).

So finally we can assume that f' is an embedding when restricted to ∂M. Choose a proper essential surface S in N' (which is the first surface of a hierarchy), and homotop f' rel. boundary so that $R := (f')^{-1}(S)$ is an essential surface in M. Note that f' induces an injection on π_1 on each component of R; since S has boundary, and since f' restricted to ∂R is an embedding, it follows that $f' : R \to S$ is homotopic to a homeomorphism, by Proposition 4.17. So we can cut along R and S, and the proof follows by induction. □

4.5. **Examples of Haken manifolds.** Many naturally occurring classes of 3-manifolds are easily seen to be Haken, and many natural operations on 3-manifolds stay in the world of Haken ones. We describe some examples

Example 4.18 (Handlebodies). Every handlebody of genus at least 1 is Haken. They contain many interesting incompressible surfaces, but the only incompressible and boundary incompressible surfaces they contain are meridian disks. To see this, intersect any such surface S with a maximal family $\mathcal{D}$ of meridian disks, and minimize the complexity of the intersection. By incompressibility of S, there can be no loops of intersection. By boundary incompressibility of S, innermost arcs of intersection can be eliminated by sliding S over a disk in $\mathcal{D}$; this produces a new surface S', still incompressible and boundary incompressible if S is, with the same topology as S but with fewer intersections with $\mathcal{D}$. So we ultimately obtain S'' disjoint from $\mathcal{D}$, and therefore deduce that S'' is a disk.

Example 4.19 (Knot complements). Let K be a knot in S^3. Then the complement $S^3 - N(K)$ of an open tubular neighborhood of K is a Haken manifold with torus boundary. Irreducibility follows from Alexander's Theorem, plus the fact that K is connected. By Mayer–Vietoris, $H^1(S^3 - N(K)) = \mathbb{Z}$; a properly embedded 2-sided surface dual to the generator is called a *Seifert surface*. Any minimal genus Seifert surface is incompressible, thus showing that $S^3 - N(K)$ is Haken.

Note that by the sphere theorem, $\pi_2(S^3 - N(K)) = 0$, and by noncompactness, $\pi_3(S^3 - N(K)) = 0$. Thus, knot complements are $K(\pi, 1)$'s.

Example 4.20 (Branched covers). Let $M \to N$ be a branched cover, and suppose N is Haken. For the sake of argument, let's suppose M and N are oriented. A branched cover between oriented manifolds has nonzero degree, and the image of $\pi_1(M)$ therefore has finite index in $\pi_1(N)$ (for, otherwise, the map would factor through an infinite — and hence noncompact — cover of N, and would be trivial on H_3). Since N is Haken, $\pi_1(N)$ admits a nontrivial action on a tree, and therefore so does $\pi_1(M)$.

This does not yet imply that M is Haken. There are two ways that M could contain essential spheres:

(1) there is a sphere S in M disjoint from L, and such that for each ball component B of $M - S$ the intersection $B \cap L$ is not an unlink; or
(2) there is a sphere S in M intersecting L in exactly one component K, so that $K \cap S$ consists of exactly 2 points, and for each ball component B of $M - S$ the intersection $B \cap L$ is not an unlink or an unlink together with an unknotted arc.

In every other case, M will be irreducible. To see this, first pass to a further cover which is regular — i.e. obtained by quotient by a finite group action. Then look for a *least area* essential embedded sphere Σ in M, with respect to a group-invariant metric. Then Σ is either disjoint from or equal to its translates under the group action, and therefore covers a spherical orbifold S in N. Note that because we are assuming the manifolds are orientable, S is topologically a sphere with finitely many orbifold points.

Because N is Haken, the underlying sphere S is inessential in N, so each component of L intersects S an even number of times. Giving N its orbifold structure, we see that S has at most 3 orbifold points (because it has spherical geometry); so in fact, it must have at most 2 orbifold points, and with the same order. So at most one component K of L can intersect S in at most 2 points (after an isotopy for which the number of points of intersection is minimal).

4.6. **Examples of non-Haken manifolds.** Haken manifolds by definition are irreducible, and have infinite π_1, so reducible 3-manifolds, and manifolds with finite fundamental group, are non-Haken. It is more interesting to give examples with infinite π_1.

Example 4.21 (Small Seifert fibered spaces). A compact 3-manifold is a *Seifert fibered space* if it is foliated by circles. Epstein showed that every circle γ in such a manifold has a neighborhood that looks like a solid torus constructed as $D^2 \times I/(x,1) \sim (\varphi(x),0)$ where $\varphi : D^2 \to D^2$ is a twist through angle $2\pi p/q$ for some coprime (p,q), where γ is the core circle $0 \times I/(0,1) \sim (0,0)$, and where the nearby circles are finite unions of the vertical intervals $x \times I$. Each nearby circle winds q times around γ, and sits on the boundary of a tubular neighborhood of γ like a (p,q) torus knot. A circle γ for which $q > 1$ is called a *singular fiber*; there are finitely many of these. Other circles are called *ordinary*

If M is a Seifert fibered space, the quotient space of M by its circle fibers is a 2-dimensional orbifold O. Singular fibers become orbifold points in O; a singular fiber of type (p,q) becomes an orbifold point of order q.

If α is an essential embedded loop in O, then the preimage of α in M is a *vertical torus*, and is essential in M; thus every such M is Haken unless O contains no essential loop. This can happen only if O is a sphere with at most three orbifold points. Such Seifert fibered spaces are called *small*. If O is a Euclidean or hyperbolic orbifold (for example, S^2 with orbifold points of orders $(2,3,6)$ is Euclidean, while S^2 with orbifold points of orders (p,q,r) with $1/p + 1/q + 1/r < 1$ is hyperbolic) then M is irreducible, and has infinite π_1.

Example 4.22 (Dehn fillings on the Figure 8 knot complement). Thurston [39] § 4.10 classified incompressible surfaces in the Figure 8 knot complement. There are essentially only 6 such surfaces: the boundary torus, the once-punctured torus Seifert surfaces, two one-sided once-punctured Klein bottles (each of which is obtained as a colored region by a two-coloring of one of the obvious projections), and two twice-punctured tori which are the boundaries of tubular neighborhoods of the Klein bottles.

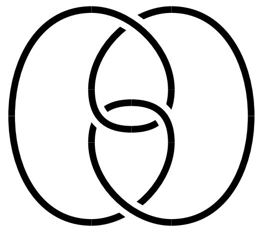


FIGURE 7. A symmetrical projection of the figure 8 knot.

Now, suppose $M(p,q)$ is obtained by Dehn filling the Figure 8 knot complement M by attaching a solid torus whose meridian is the (p,q) curve on ∂M. An essential surface S in $M(p,q)$ can be made transverse to the core α of the filling torus, and gives rise to a surface $S \cap M$ properly embedded in M. If $S \cap M$ is compressible, a compressing disk $D \subset M$ must have $\partial D = \partial D'$ for some $D' \subset S$ which intersects α. If we cut out D' from S and replace it with D, we obtain a new surface S' which intersects α in fewer curves than S does; moreover, S' is incompressible if S is. Similarly, if $S \cap M$ is boundary incompressible, we can perform a boundary compression by an isotopy of S which eliminates intersections with α. Thus, if $S \cap \alpha$ is minimal, $S \cap M$ is incompressible in M, and must be one of the 6 surfaces above. It can't be ∂M, which does not stay incompressible in any filling; so it has to be one of the other 5 surfaces. But each component of $\partial(S \cap M)$ must bound a disk in $M(p,q)$, so (p,q) must be one of the slopes bounded by these surfaces.

Thus: $M(p,q)$ contains an essential surface if and only if (p,q) is one of $(0,\pm 1)$ or $(\pm 4,\pm 1)$. On the other hand, for all but finitely many other surgeries, $M(p,q)$ is hyperbolic, and therefore is certainly irreducible and has infinite fundamental group.

Example 4.23 (Seifert–Weber dodecahedral manifold). Jaco–Oertel [16], building on key ideas of Haken, gave an algorithm to determine whether or not a 3-manifold is Haken. This algorithm is extremely impractical, but has recently been improved to the point where it can actually be used on small examples. We give an outline of this algorithm in § 4.7.

The *Seifert–Weber* dodecahedral manifold is obtained from a regular dodecahedron by gluing opposite faces by a twist of $3\pi/5$. Edges are identified in six groups of five.

For any angle α between 60° and (approximately) 116.565° there is a regular hyperbolic dodecahedron with every dihedral angle equal to α. Choosing $\alpha = 72^\circ$ gives a dodecahedron for which the gluing above gives rise to a complete hyperbolic structure on the Seifert–Weber space.

Thurston conjectured that the Seifert–Weber manifold is non-Haken. This was proved by Burton–Rubinstein–Tillmann [3] in 2012 by a computer implementation of their improvements to the Jaco–Oertel algorithm.

4.7. **Recognizing Haken manifolds.**

5. THE TORUS THEOREM AND THE JSJ DECOMPOSITION

5.1. **The Torus Theorem.** After the sphere, the next simplest kind of oriented closed surface is the torus. The Sphere Theorem lets one replace an essential immersed sphere by an embedded one, and it is natural to want a similar theorem for tori. The first such

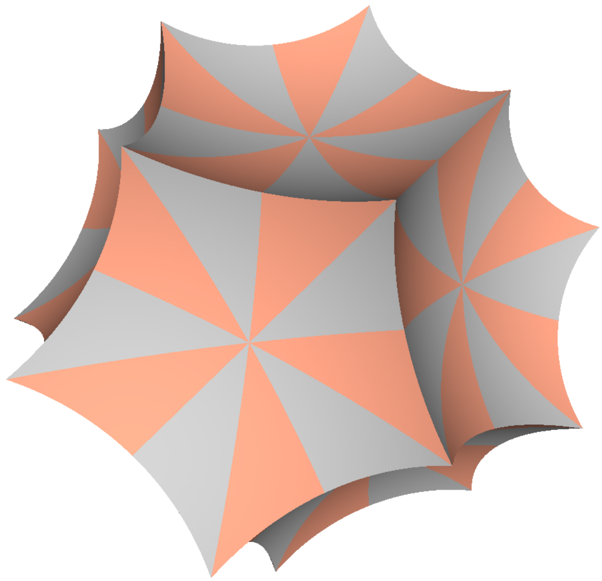


FIGURE 8. A regular hyperbolic dodecahedron with dihedral angles 72°

torus theorem was proved by Waldhausen under the (rather strong) assumption that the manifold is Haken. Scott [32] formulated and proved a stronger version, which we now state:

Theorem 5.1 (Torus Theorem)**.** *Let M be a closed orientable irreducible 3-manifold, and suppose that there is a π_1-injective map $f : T \to M$ where T is a torus. Then*

(1) either M contains an embedded incompressible torus, which can be taken to be contained in any neighborhood of the image of T; or
(2) $\pi_1(M)$ has an infinite cyclic normal subgroup.

The latter possibility can certainly occur: a small Seifert Fibered Space with Euclidean or hyperbolic base orbifold is irreducible, and contains many immersed π_1-injective tori. However it is not even Haken; see Example 4.21.

It is natural to wonder whether these are the only examples; this turns out to be the case, and its proof combines work of Mess, Tukai, Gabai, Casson–Jungreis:

Theorem 5.2 (Seifert fibered theorem)**.** *Let M be irreducible, and suppose that $\pi_1(M)$ contains an infinite cyclic normal subgroup. Then M is Seifert fibered.*

Combining Theorem 5.2 with Theorem 5.1 we obtain the satisfying conclusion that a closed orientable irreducible 3-manifold admits a π_1-injective map of a torus if and only if it admits an essential embedded torus, or is a small Seifert Fibered Space.

Theorem 5.2 has a much easier argument if one adds the further hypothesis that M is *Haken*. Under this hypothesis, the conclusion (that M is Seifert fibered) was proved by Waldhausen, and his proof was simplified by Scott.

But at the time Scott proved Theorem 5.1, there were already well-known examples of non-Haken 3-manifolds whose fundamental group contains an infinite cyclic normal subgroup, namely the *small Seifert fibered spaces* discussed in Example 4.21. Theorem 5.2 says that these are the *only* non-Haken examples.

We will sketch proofs of these theorems in the next few sections.

5.2. **Minimal surfaces and Casson's proof of the Torus Theorem.** We will now sketch the proof of the Torus Theorem 5.1 following Casson. Casson's proof follows Scott's outline, but with a major simplification that depends on using results from the theory of minimal surfaces.

Definition 5.3. A map of a surface $f : S \to M$ is *least area* if it globally minimizes the area amongst all maps in its free homotopy class.

We summarize some of the most important properties of least area surfaces:

1. Suppose $\pi_2(M) = 1$. Then if no simple loop in S is conjugate into the kernel of $f_* : \pi_1(S) \to \pi_1(M)$, a least area surface exists. [34], [30]
2. Least area surfaces are immersed; i.e. they have no branch points. [22]
3. If M is orientable and irreducible, and S is orientable and not S^2, then double curves are essential. [9]
4. If $f : S \to M$ is least area and π_1-injective, if M is orientable and irreducible, and if S is not a sphere, the preimage of $f(S)$ in the universal cover consists of a union of properly embedded planes, and the intersection between any two such planes is a union of lines. [9]
5. If $f : T \to M$ is a least area incompressible torus, then the preimages have the *1-line property:* i.e. they intersect pairwise in at most one line. [9]

Casson's main technical contribution beyond these results was the following:

Lemma 5.4 (Casson)**.** *If $f : T \to M$ is a least area torus, the double curves of f represent primitive elements of $\pi_1(T)$.*

Proof. We give a proof of this lemma following Max Neumann–Coto [21], using the results of Freedman–Hass–Scott [9] described above. Let P and Q be transverse planes in the universal cover $\tilde{M}$ covering T. By the 1-line property, their intersection $P \cap Q$ is a single line ℓ. Suppose α stabilizes Q but does not stabilize P. We further let ℓ_i denote the line $P \cap \alpha^i P$ whenever this is nonempty and the intersection is transverse (again, we are using the 1-line property). Note that ℓ, the ℓ_i, and all their translates by powers of α are *parallel* in $\tilde{M}$ (i.e. they are stabilized by some common nontrivial covering translation). For, the generator λ of the common stabilizers of P and Q commutes with α, and therefore stabilizes the set of intersections between Q and all the $\alpha^i P$. This fact lets us reduce the analysis of the configuration of Q and the various $\alpha^i P$ to configurations of lines in the plane, where separation arguments become more powerful.

The line ℓ divides P into two parts; we call the ends of these P^+ and P^-. Clearly, if P intersects αP, then αP^+ and $\alpha^{-1} P^+$ lie on different sides of P in $\tilde{M}$. Similarly, if P intersects αP then P intersects $\alpha^2 P$; for, both P and $\alpha^2 P$ intersect Q, and both P^+ and $\alpha^2 P^+$ are on the same side of Q.

Finally, we claim that for all positive n, the ends αP^+ and $\alpha^n P^+$ lie on the same side of P. Suppose by induction $\alpha^i P^+$ lies to the right of P for all $1 \le i \le n$. We can't have $\alpha^{n+1} P = P$, or else $\alpha^n P^+ = \alpha^{-1} P^+$ contrary to the fact that αP^+ and $\alpha^{-1} P^+$ are on different sides of P. The subset X on the positive side of Q to the right of $\alpha^n P$ and to the left of P is invariant under λ, and its quotient is compact, so it can't contain $\alpha^{n+1}P^+$; thus this end must be to the right of P, as claimed. But then, no power of α stabilizes P, and we deduce that stabilizers of lines are primitive. □

It follows from this lemma that the double curves of f are homotopic to simple curves in T. Given this, we now explain the proof of the Torus Theorem.

Let $f : T \to M$ be a least area π_1-injective torus. By the 1-line property, each double curve α on T is covered by a line ℓ in $\tilde{M}$ which must necessarily meet two distinct planes P, Q covering T, stabilized by subgroups G_P, G_Q, both isomorphic to $\mathbb{Z}^2$, and intersecting in a primitive $\mathbb{Z}$ subgroup which is the stabilizer of ℓ. Suppose there are double curves α, β in distinct homotopy classes in T. Then there are planes P, Q, R stabilized by G_P, G_Q, G_R all distinct, for which $P \cap Q = l_\alpha$, $P \cap R = l_\beta$ and (since $\alpha \cap \beta$ is nonempty) $Q \cap R$ is nonempty and equal to some l_γ. Now, the stabilizer γ of l_γ has no power conjugate into $\langle \alpha, \beta \rangle = \mathbb{Z}^2$; on the other hand, it commutes with both of them. Thus $\langle \alpha, \beta, \gamma \rangle = \mathbb{Z}^3$, and it follows for homological reasons that M is finitely covered by T^3. It is known that M is Seifert fibered in this case.

Otherwise we may assume that all double curves are parallel, and therefore (because primitive) freely homotopic. Look at the universal cover, and consider a connected component X of the preimage of T, so that X is a union of planes which are pairwise disjoint or intersect in lines, and any two lines of intersection contained in the same plane are parallel in that plane. We claim that the stabilizer of X normalizes α. This is proved by induction: let P be some plane of X, and let gP be another plane. Because X is connected, we can join P to gP by a sequence of translates

$$P := P_0, P_1, \cdots, P_n := gP$$

where $P_i = g_i P$ for some g_i, and where $P_i \cap P_{i+1} = \ell_i$ is stabilized by α_i. We have $\alpha = \alpha_0$ by hypothesis. Moreover, $g_i g_{i-1}^{-1} : P_{i-1} \to P_i$ takes ℓ_{i-1} to a line in P_i which is parallel both to ℓ_{i-1} and to ℓ_i; since by Lemma 5.4 these lines are stabilized by the same primitive element, we have that α_{i-1} is equal to $\alpha_i^{\pm}$, and is normalized by $g_i g_{i-1}^{-1}$. Thus, by induction, α is normalized by g.

If X is not stabilized by all of $\pi_1(M)$, the boundary of a regular neighborhood covers a closed surface whose fundamental group contains an infinite cyclic normal subgroup, and is therefore a torus or Klein bottle; since M is orientable, there is an embedded essential torus in this case. Otherwise, X is stabilized by all of $\pi_1(M)$, which means that $\langle \alpha \rangle$ is normal in $\pi_1(M)$, as claimed.

5.3. **The Seifert Fibered Theorem: groups quasi-isometric to planes.**

Recall that the Seifert Fibered Theorem 5.2 says that if M is irreducible, and $\pi_1(M)$ contains an infinite cyclic normal subgroup, then M is Seifert fibered. We abbreviate $G := \pi_1(M)$ and let C be the infinite cyclic normal subgroup.

The first step in the proof of the Seifert Fibered Theorem is due to Mess [18], and is unfortunately unpublished. In this section we sketch the contents of Mess's preprint. This takes several steps, each of which is highly original and technical, and therefore our survey necessarily omits many details.

Let's begin. We can assume by passing to a cover if necessary that M is a closed, oriented 3-manifold whose fundamental group contains a central $\mathbb{Z}$ subgroup. For simplicity, let's in fact assume that the center is actually equal to $\mathbb{Z}$; it is easy to reduce to the case that the center has rank 1, but it is subtle to deal with the possibility that the center might be infinitely generated. In any case, the first main theorem Mess proves ([18] Thm. 1, page 2), by a "bare hands" topological argument, is:

Proposition 5.5 (Cover is solid torus)**.** *Let M be closed, irreducible, orientable. Suppose that center of $\pi_1(M)$ is $\mathbb{Z}$. Then the covering space $\hat{M}$ with fundamental group equal to this center is homeomorphic to a solid torus.*

Proof. Let's let a denote the generator of the center. Because it is central, the element a is well-defined as an element of $\pi_1(M, p)$ for any point p, so we can build (e.g. inductively on the skeleta of a triangulation) a homotopy $H : M \times S^1 \to M$ such that the track of every point in M under the homotopy is in the class of a. We can lift this homotopy to $H : \hat{M} \times S^1 \to \hat{M}$; because M was compact, the length of the tracks of the homotopy have uniformly bounded length. For homological reasons, $\hat{M}$ is one-ended, and the first observation is that every compact set K in $\hat{M}$ can be separated from this end by an embedded torus T in such a way that a is still central in $\pi_1(E)$, where E is the noncompact region bounded by T. To see this, first observe that K can be included in a big compact set K'' such that the track of $\partial K''$ under the homotopy H stays disjoint from K (this uses the fact that the tracks themselves have uniformly bounded length). The surface $\partial K''$ is essential in $H_2(\hat{M} - K)$, and its image under H sweeps out an immersed 3-manifold whose image G in $\pi_1(\hat{M} - K)$ contains a central $\mathbb{Z}$ subgroup (the image of the tracks of the homotopy). Pass to the cover $\hat{M}_G$ of $\hat{M} - K$; this manifold has nontrivial H_2, and is therefore Haken, so (because it has a central $\mathbb{Z}$ subgroup) is already known to be a Seifert fibered space. Thus the surface $\partial K''$ can be replaced by a homologically equivalent embedded torus, which necessarily bounds a solid torus in $\hat{M}$. So $\hat{M}$ is an increasing union of solid tori; a further standard argument shows that these tori nest nicely in each other, and the union is a solid torus. □

Now, at this stage, $\hat{M}$ has two useful structures: topologically it is homeomorphic to a solid torus $\mathbb{R}^2 \times S^1$, while geometrically it admits a homotopy $H : \hat{M} \times S^1 \to \hat{M}$ whose tracks have bounded length. The next step is to find a relationship between these two structures:

Proposition 5.6 (Circles of bounded length)**.** *With M, $\hat{M}$ and H as above, there is a homotopy $J : \hat{M} \times S^1 \times [0,1] \to \hat{M}$ whose $S^1 \times [0,1]$ tracks have uniformly bounded diameter, which starts at H and ends at a free circle action on $\hat{M}$ witnessing its topological product structure.*

Proof. In words, J is a bounded homotopy from H to the Seifert structure. In particular, because J has fibers of bounded diameter, $\hat{M}$ admits a product structure for which the circle fibers have uniformly bounded length. The homotopy J is constructed inductively out of "round handles" — i.e. products of circles with ordinary (2-dimensional) handles. First, we can pick any unknotted core γ of the solid torus, and take this to be the image of some track of H under the homotopy J. The deck group $G := \pi_1(M)/\pi_1(\hat{M})$ (which is a group because $\pi_1(\hat{M})$ is central and therefore normal) acts on $\hat{M}$ by isometries, and therefore by homeomorphisms; and thus permutes the set of positively oriented unknotted cores, since these are the only unknotted circles which represent a homotopically in a solid torus. Choose a separated net in G — a collection of elements g_i such that no two are very close, and such that every element is not too far away from something in the net. Evidently we can choose such a net so that the translates of γ by elements of the net are all mutually unlinked, and collectively represent an unknotted collection of circles in $\hat{M}$. Thicken each such circle to a round 0-handle; these will be the round 0-handles in our decomposition.

Building the round 1-handles is tricky, and requires quite an ingenious argument. Because we chose a separated net, every round 0-handle is close to some, but not too many, other round 0-handles. Any two round 0-handles which are close enough can be connected by some annulus (because their cores are isotopic), and we can least area representatives. Two such least area annuli cannot intersect on their boundaries (unless they agree), by the roundoff trick. Thus, any two of them will intersect transversely in finitely many essential circles. So we pick a starting 0-handle B_0 and inductively attach least area annuli one at a time, choosing the absolute smallest area one among the finitely many (up to isotopy) which join an unattached 0-handle (which we will call B_n) to one of the $B_0, \cdots, B_{n-1}$ constructed so far, and by a roundoff argument, we see that the result is embedded. By transfinite induction, all the round 0-handles can be connected up in this way after some countable ordinal stage. The annuli we attach can be thickened to become round 1-handles, and the result is a tree of round 0-handles, connected up by round 1-handles, all with uniformly bounded diameter (this is because at every stage some B_n yet to be connected is bounded distance from the union of the handles connected so far, so the annuli which are attached have uniformly bounded diameter).

Now consider a component X of the boundary of the union of round 0- and 1-handles constructed so far. Note that X is partitioned into annuli T_i of bounded diameter which are on the boundaries of the 0-handles, and A_j which are on the boundaries of the 1-handles. They appear in a particular order $\cdots T_{-1}, T_0, T_1 \cdots$. Adding further round 1-handles splits X into components, some of which might be bounded. We would like to add new annuli, to split X up into components of uniformly bounded (combinatorial) size; to do this, we need to find pairs of T_i, T_j which are a uniformly big combinatorial distance apart, but which can be joined by and embedded annuli of uniformly bounded diameter. It is intuitively clear that this can be done: if X is noncompact, the two "ends" of X can't get too far away from each other, or else there would be an arbitrarily big embedded ball contained in the complement, which is incompatible with the fact that we chose a separated net's worth of translates of our original 0-handle. A similar argument works when X is compact but sufficiently big (alternately one can suppose not and take pointed limits, since this is a purely geometric argument). Thus we can attach round 1-handles of uniformly bounded diameter so that at the end, every component X itself has bounded diameter, and can be filled in with a round 2-handle. The construction of J with this handle decomposition as the end result is routine. □

This brings us to section 3 of Mess' paper (page 11), entitled, *On groups which are coarse quasi-isometric to planes*. The group in question is G, i.e. $\pi_1(M)/\pi_1(\hat{M})$. This is the group that we hope will turn out to be the orbifold fundamental group of some

2-orbifold O (the base of a Seifert fibration of M), if the Seifert Conjecture is true. Since it is infinite, we want to show that G is a lattice in the group of isometries of the Euclidean or hyperbolic plane; in fact, a cocompact lattice, since M is closed. In particular, this should imply at least that G is quasi-isometric either to the Euclidean or the hyperbolic plane. By the Schwarz lemma, we know that G is quasi-isometric to $\hat{M}$, and we have constructed a product structure on $\hat{M}$ whose fibers have uniformly bounded length. It is therefore straightforward (e.g. by averaging over fibers) to construct a complete Riemannian metric on the plane (which we denote P) so that G is quasi-isometric to P. The next main result is Thm. 7 (page 13) which says:

Theorem 5.7. *Suppose a finitely generated group G is quasi-isometric to a plane P with a complete Riemannian metric. If P is conformally equivalent to the hyperbolic plane, then G is quasi-isometric to the hyperbolic plane.*

Mess's proof of this theorem is interesting, but can be shortened by appealing to a theorem of Candel [4].

Proof. Note that P has bounded geometry (i.e. 2-sided curvature bounds, and injectivity radius bounded below). One subtlety, observed by Mess, is that the plane admits complete Riemannian metrics with bounded geometry, and in the conformal class of the hyperbolic plane, but for which 0 is the bottom of the spectrum of the Laplacian; a group quasi-isometric to such a space would be amenable, by a famous theorem of Brooks, whereas no group quasi-isometric to the hyperbolic plane can be amenable.

Nevertheless, Candel proves that if L is a compact Riemann surface lamination all of whose leaves are conformally hyperbolic, then the leafwise uniformization map is continuous; in particular, since L is compact, the uniformization map is bilipschitz (and in particular is a quasi-isometry). Now, a Riemannian manifold with bounded geometry can be realized as a dense leaf in a lamination by taking its closure in pointed Gromov-Hausdorff space; if we do this to P, we obtain a lamination L. *A priori* a lamination can have leaves of different conformal type; but in this case P is uniformly quasi-isometric to G, and therefore (since G acts cocompactly on itself) the same must be true for every leaf of L. Now apply Candel's theorem. □

Finally we must deal with the case that P is quasi-isometric to the Euclidean plane. In this case, Thm. 10 (page 20) says (paraphrasing):

Theorem 5.8 (Conformally Euclidean)**.** *Suppose $G = \pi_1(M)/\pi_1(\hat{M})$ is quasi-isometric to a plane P with a complete Riemannian metric, which is conformally equivalent to the Euclidean plane. Then G is virtually rank 2 abelian, and M is Seifert fibered; thus, the Seifert fiber Conjecture holds in this case.*

Proof. The argument is a beautiful application of ideas from the theory of random walks, combined with a theorem of Varopoulos. It is a well-known fact that a simple random walk is recurrent (i.e. returns to a bounded region infinitely often) in Euclidean space of dimension 1 and 2, and transient otherwise. This is not hard to show: under random walk on Euclidean space, after n steps each coordinate function is distributed like a Gaussian with variance of order n; thus the probability that a given coordinate function will be bounded by a constant C after n steps is of order $O(\sqrt{n})$. By independence, in m-dimensional space, the probability that all coordinate functions will be bounded by the same constant C at the same time after n steps is $O(n^{-m/n})$; thus, when m is at least 3, the total number of times this should happen in an infinite walk is bounded, by the Borel-Cantelli Lemma.

Now, in the continuum limit, a simple random walk rescales to Brownian motion, and Brownian motion is conformally invariant in dimension 2; this means that if you have a complete Riemannian metric on a plane P, you can tell whether it is conformally hyperbolic or conformally Euclidean by whether Brownian motion is transient or recurrent. Using the quasi-isometry between P and G, one concludes that if P is conformally Euclidean, random walk on G is recurrent. But this is an extremely confining possibility for finitely presented groups; Varopoulos [40] showed (when combined with Gromov's famous theorem that groups of polynomial growth are virtually nilpotent) that it implies that G is virtually abelian of rank at most 2; this is enough to complete the proof, using the (known) classification of nilpotent 3-manifold groups. ☐

Mess' paper thus reduces the Seifert Fibered Conjecture to the question of whether groups quasi-isometric to the hyperbolic plane are virtually isomorphic to Fuchsian groups — i.e. to (cocompact) lattices in the group of isometries of the hyperbolic plane. Much progress on this question had already been made by Tukia, and while Mess' paper was still under consideration at JAMS this question was solved in the affirmative independently (and in quite different ways) by Casson-Jungreis, and Gabai.

5.4. **The Seifert Fibered Theorem: the convergence group theorem.** To complete the proof of the Seifert Fibered Theorem 5.2 it suffices to show that any subgroup of $\mathrm{Homeo}^+(S^1)$ acting as a convergence group is conjugate into $\mathrm{PSL}(2,\mathbb{R})$. This will imply that G/C (as above) is isomorphic to the fundamental group of a hyperbolic orbifold, and therefore either that M contains an essential torus, or that this orbifold is a triangle orbifold, and M is a small Seifert Fibered Space.

6. Acknowledgments

Danny Calegari was supported by NSF grant DMS 1005246.

References

[1] I. Agol, *Criteria for virtual fibering*, J. Topol. **1** (2008), no. 2, 269–284
[2] I. Agol (with an appendix by D. Groves and J. Manning), *The virtual Haken conjecture*, preprint; arXiv:1204.2810
[3] B. Burton, H. Rubinstein and S. Tillmann, *The Weber–Seifert dodecahedral space is non-Haken*, Trans. AMS **364** (2012), 911–932
[4] A. Candel, *Uniformization of surface laminations*, Ann. Sci. École Norm. Sup. (4) **26** (1993), no. 4, 489–516
[5] A. Casson and C. Gordon, *Reducing Heegaard splittings*, Top. Appl. **27** (1987), 275–283
[6] J. Cerf, *La stratification naturelle des espaces de fonctions différentiables réelles et le théorème de la pseudo-isotopie*, IHES Pub. Math. **39** (1970), 5–173
[7] M. Dehn, *Über die Topologie des dreidimensionalen Raumes*, Math. Ann. **69** (1910), 137–168
[8] D. Epstein, *Ends*, Topology of 3-manifolds and related topics (Proc. The Univ. of Georgia Institute, 1961), 110–117; Prentice-Hall, Englewood Cliffs, 1962
[9] M. Freedman, J. Hass and P. Scott, *Least area incompressible surfaces in 3-manifolds*, Invent. Math. **71** (1983), 609–642

[10] C. Gordon and J. Luecke, *Knots are determined by their complements*, J. Amer. Math. Soc. **2** (1989), no. 2, 371–415
[11] W. Haken, *Some results on surfaces in 3-manifolds*, Studies in Modern Topology (MAA Studies in Mathematics ed. P. Hilton), 39–98; Prentice–Hall, Englewood Cliffs, 1968
[12] J. Hempel, *3-Manifolds*, Ann. of Math. Studies **86**, Princeton University Press, Princeton, 1976
[13] W. Jaco, *Heegaard splitting and splitting homomorphisms*, Trans. AMS **144** (1969), 365–379
[14] W. Jaco, *Finitely presented subgroups of three-manifold groups*, Invent. Math. **13** (1971), 335–346
[15] W. Jaco, *Lectures on three-manifold topology*, CBMS Regional Conference Series in Math. **43**, AMS, Providence, 1980
[16] W. Jaco and U. Oertel, *An algorithm to decide if a 3-manifold is a Haken manifold*, Topology **23** (1984), no. 2, 195–209
[17] F. Laudenbach, *A proof of Reidemeister–Singer's theorem by Cerf's methods*, preprint, arXiv:1201.1130
[18] G. Mess, *The Seifert conjecture and groups which are coarse quasiisometric to planes*, preprint, available from `https://lamington.files.wordpress.com/2014/08/mess_seifert_conjecture.pdf`
[19] J. Milnor, *A unique factorization theorem for 3-manifolds*, Amer. J. Math. **84** (1962), 1–7
[20] E. Moise, *Affine structures in 3-manifolds V. the triangulation theorem and the Hauptvermutung*, Ann. of Math. **55** (1952), 96–114
[21] M. Neumann-Coto, *On the double curves of least area tori*, Proc. AMS. **125** (1997), no. 8, 2463–2469
[22] R. Osserman, *A proof of the regularity everywhere of the classical solution to Plateau's problem*, Ann. Math. (2) **91** (1970), 550–569
[23] C. Papakyriakopoulos, *On Dehn's Lemma and the asphericity of knots*, Ann. Math. **66** (1957), no. 1, 1–26
[24] G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, preprint, arXiv:math/0211159
[25] G. Perelman, *Ricci flow with surgery on three-manifolds*, preprint, arXiv:math/0303109
[26] G. Perelman, *Finite extinction time for the solutions to the Ricci flow on certain three-manifolds*, preprint, arXiv:math/0307245
[27] K. Reidemeister, *Homotopieringe und Linsenräume*, Hamburger Abhandl, **11** (1935), 102–109
[28] Y. Rieck, *A proof of Waldhausen's uniqueness of splittings of S^3 (after Rubinstein and Scharlemann)*, Geom. Top. Monog. **12** (2007), 277–284
[29] H. Rubinstein and M. Scharlemann, *Comparing Heegaard splittings of non-Haken 3-manifolds*, Topology **35** (1996), 1005–1026
[30] J. Sacks and K. Uhlenbeck, *Minimal immersions of closed Riemann surfaces*, Trans. AMS **271** (1982), 639–652
[31] P. Scott, *Compact submanifolds of 3-manifolds*, J. London Math. Soc. (2) **7** (1973), 246–250
[32] P. Scott, *A new proof of the annulus and torus theorems*, Amer. J. Math. **102** (1980), no. 2, 241–277
[33] J.-P. Serre, *Trees*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2003
[34] R. Schoen and S.-T. Yau, *Existence of incompressible minimal surfaces and the topology of three-dimensional manifolds with non-negative scalar curvature*, Ann. Math. (2) **110** (1979), no. 1, 127–142
[35] J. Stallings, *Some topological proofs and extensions of Grushko's theorem*, Ph.D thesis, Princeton University, 1959
[36] J. Stallings, *How not to prove the Poincaré conjecture*, Ann. of Math. Studies **60** (1966), 83–88
[37] J. Stallings, *On torsion-free groups with infinitely many ends*, Ann. of Math. (2) **88** (1968), 312–334
[38] J. Stallings, *Group theory and three-dimensional manifolds*, Yale Mathematical Monographs **4**, Yale University Press, New Haven, 1971
[39] W. Thurston, *The Geometry and Topology of 3-Manifolds*, a.k.a "Thurston's Notes"; available from the MSRI at `http://www.msri.org/publications/books/gt3m`
[40] N. Varopoulos, *Isoperimetic inequalities and Markov chains*, J. Funct. Anal. **63** (1985), no. 2, 215–239
[41] F. Waldhausen, *Heegaard-Zerlegungen der 3-Sphäre*, Topology **7** (1968), 195–203

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