ISSN 1364-0380 (on line) 1465-3060 (printed)

Geometry & Topology
Volume 9 (2005) 1295–1336
Published: 26 July 2005

Stabilization for the automorphisms of free groups with boundaries

Allen Hatcher
Nathalie Wahl

Mathematics Department, Cornell University, Ithaca NY 14853, USA
and
Aarhus University, 116 Ny Munkegade, 8000 Aarhus C, Denmark

Email: `hatcher@math.cornell.edu` and `wahl@imf.au.dk`

Abstract

The homology groups of the automorphism group of a free group are known to stabilize as the number of generators of the free group goes to infinity, and this paper relativizes this result to a family of groups that can be defined in terms of homotopy equivalences of a graph fixing a subgraph. This is needed for the second author's recent work on the relationship between the infinite loop structures on the classifying spaces of mapping class groups of surfaces and automorphism groups of free groups, after stabilization and plus-construction. We show more generally that the homology groups of mapping class groups of most compact orientable 3-manifolds, modulo twists along 2-spheres, stabilize under iterated connected sum with the product of a circle and a 2-sphere, and the stable groups are invariant under connected sum with a solid torus or a ball. These results are proved using complexes of disks and spheres in reducible 3-manifolds.

AMS Classification numbers Primary: 20F28

Secondary: 57M07

Keywords: Automorphism groups of free groups, homological stability, mapping class groups of 3-manifolds

Proposed: Martin Bridson — Received: 15 July 2004
Seconded: Ralph Cohen, Thomas Goodwillie — Revised: 20 July 2005

Introduction

There is by now a significant literature on homological stability properties of various families of discrete groups, including matrix groups, mapping class groups of surfaces, and automorphism groups of free groups. The present paper is yet another contribution in this direction, for a family of groups $A^s_{n,k}$ that can be thought of as relative versions of automorphism groups of free groups. The groups $A^1_{n,k}$ arose first in [23] in studying the relation between mapping class groups of surfaces and automorphism groups of free groups, and applications to this work provide our main motivation. The extra parameter s in the family, necessary for the proof of the stability theorem, has also interests of its own. In the present paper we exploit the fact that $A^s_{n,k}$ can be realized as a quotient of the mapping class group of a certain 3–manifold, and our proof of stability applies to give a homological stability result for the analogous quotients of the mapping class groups of general compact orientable 3–manifolds.

The quickest definition of $A_{n,k} = A^1_{n,k}$ is in terms of homotopy equivalences of graphs. Let $G_{n,k}$ denote the graph shown in Figure 1 consisting of a wedge of n circles together with k distinguished circles joined by arcs to the basepoint.

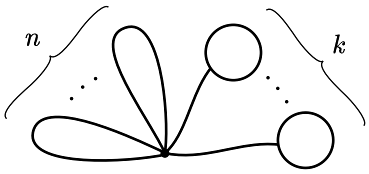

Figure 1: The graph $G_{n,k}$

Then $A_{n,k}$ is the group of path-components of the space of homotopy equivalences of $G_{n,k}$ that preserve the basepoint and restrict to the identity on each of the k distinguished circles. If this last condition is relaxed to allow homotopy equivalences that rotate each of the k distinguished circles, the group $A_{n,k}$ is replaced by the group $\mathrm{Aut}_{n,k}$ of automorphisms of the free group on $n+k$ generators that take each of the last k generators to a conjugate of itself. By restricting the more general homotopy equivalences of $G_{n,k}$ to their rotations of the distinguished circles, we obtain a fibration whose associated long exact sequence of homotopy groups is a short exact sequence

$$1 \longrightarrow \mathbb{Z}^k \longrightarrow A_{n,k} \longrightarrow \mathrm{Aut}_{n,k} \longrightarrow 1$$

expressing $A_{n,k}$ as a central extension of $\mathrm{Aut}_{n,k}$. The groups $\mathrm{Aut}_{n,k}$ can be regarded as interpolations between automorphism groups of free groups, which are the groups $\mathrm{Aut}_{n,0}$, and the groups $\mathrm{Aut}_{0,k}$ which are the 'symmetric' automorphism groups studied for example in [2] and [18]. A purely algebraic description of $A_{n,k}$ is given in [14].

The group $A_{n,k}$ relates to mapping class groups of surfaces in the following way. Let S be a surface of genus g with $k+1$ boundary components and let $\Gamma_{g,k+1}$ be its mapping class group, the group of path components of the space of diffeomorphisms of S fixing the boundary pointwise. There is a natural embedding of the graph $G_{2g,k}$ in S with the basepoint of $G_{2g,k}$ on the first boundary component of S and the k distinguished circles of G identified with the other k boundary components of S, such that the embedding is a homotopy equivalence relative to the basepoint and the k distinguished circles. This defines a map from $\Gamma_{g,k+1}$ to $A_{2g,k}$, which is actually injective. Dehn twists along the last k boundary circles of S are identified under this inclusion with the copy of $\mathbb{Z}^k$ in $A_{2g,k}$ arising in the short exact sequence above.

The natural inclusions $G_{n,k} \hookrightarrow G_{n+1,k}$ and $G_{n,k} \hookrightarrow G_{n,k+1}$ obtained by attaching an extra circle or circle-plus-arc induce stabilization maps $A_{n,k} \to A_{n+1,k}$ and $A_{n,k} \to A_{n,k+1}$ by extending homotopy equivalences via the identity on the added parts. These stabilization maps are easily seen to be injective. The main goal of the paper is to prove the following result:

Theorem A *The stabilization maps $A_{n,k} \to A_{n+1,k}$ and $A_{n,k} \to A_{n,k+1}$ induce isomorphisms $H_i(A_{n,k}) \to H_i(A_{n+1,k})$ and $H_i(A_{n,k}) \to H_i(A_{n,k+1})$ for $n \geq 3i+3$ and $k \geq 0$.*

In particular, the case $k = 0$ in the first stabilization gives a new proof of homology stability for automorphism groups of free groups, although the ranges $n \geq 2i+3$ obtained in [9] and $n \geq 2i+2$ in [11] are better than the one we obtain here. The proof of the theorem also shows that the quotient map $\mathrm{Aut}(F_n) \to \mathrm{Out}(F_n)$ induces an isomorphism on H_i for $n \geq 3i+3$. Again, a better range is given in [11].

The theorem says that the stable $A_{n,k}$ groups have the same homology as the stable automorphism groups of free groups. Not a great deal is known about these stable homology groups. In low dimensions they are trivial rationally [9], [10], and this is conjectured to hold in all dimensions. The only nontrivial torsion known so far is a copy of the homology of the infinite symmetric group, or equivalently $\Omega^\infty S^\infty$, as noted in [7].

The above theorem is one of the main ingredients in [23] in showing that the natural map from the stable mapping class group of surfaces to the stable automorphism group of free groups induces an infinite loop map on the classifying spaces of these groups after plus-construction. The infinite loop space structure of the stable mapping class group is governed by a cobordism category, the surfaces being cobordisms between disjoint unions of circles [21]. Graphs with distinguished circles, thought of as *boundary circles*, can be used to define a generalized cobordism category where graphs as well as surfaces define cobordisms and where gluing along boundary circles is compatible in an appropriate sense with the map $\Gamma_{g,k+1} \to A_{2g,k}$ described above. This category is used to compare the two infinite loop spaces and homological stability for the groups $A_{n,k}$ is needed for identifying the homotopy type of this generalized cobordism category. From [17] and [4] the homology of the stable mapping class group is known (and it is very rich!). Little is known however about the map from the mapping class groups to the automorphism groups. If the map could be shown to be rationally trivial, for example, this would give further evidence for the conjecture that the stable rational homology of the automorphism groups is trivial.

To prove the theorem we actually need to work with the more general groups $A^s_{n,k}$ which can be defined as follows. Consider the graph $G^s_{n,k}$ obtained from $G_{n,k}$ by wedging $s-1$ edges at the basepoint. We think of the free ends of the new edges as extra basepoints. (When $s=0$, we forget the basepoint of $G_{n,k}$.) Define $A^s_{n,k}$ as the group of path components of the space of homotopy equivalences of $G^s_{n,k}$ which fix the k distinguished circles as well as the s basepoints. There is a short exact sequence

$$1 \longrightarrow F_{n+k} \longrightarrow A^s_{n,k} \longrightarrow A^{s-1}_{n,k} \longrightarrow 1$$

where the map $A^s_{n,k} \to A^{s-1}_{n,k}$ forgets the last basepoint. When $s \geq 2$, this map splits and in fact $A^s_{n,k} = (F_{n+k})^{s-1} \ltimes A^1_{n,k}$ where $A^1_{n,k}$ acts diagonally on $(F_{n+k})^{s-1}$ via the map $A^1_{n,k} \to \mathrm{Aut}(F_{n+k})$. Particular cases of these groups are $A^0_{n,0} = \mathrm{Out}(F_n)$, $A^1_{n,0} = \mathrm{Aut}(F_n)$ and $A^2_{n,0} = F_n \ltimes \mathrm{Aut}(F_n)$.

We show in Theorem 4.1 that the stabilization map $A^s_{n,k} \to A^{s+1}_{n,k}$ and the map $A^s_{n,k} \to A^{s-1}_{n,k}$ for $s \geq 1$ also induce isomorphisms in homology when $n \geq 3i+3$. A corollary of the stability for $A^2_{n,0} \to A^1_{n,0}$ is that the twisted homology group $H_i(\mathrm{Aut}(F_n), \mathbf{Z}^n)$ is trivial when $n \geq 3i+9$. This is immediate from the Leray–Hochschild–Serre spectral sequence for the short exact sequence displayed above, with $s=2$ and $k=0$.

Our proof of homological stability for $A_{n,k}$, and in fact $A^s_{n,k}$, is based on an interpretation of these groups in terms of the mapping class group of compact 3-manifolds $M^s_{n,k}$ obtained from the connected sum of n copies of $S^1 \times S^2$ by deleting the interiors of s disjoint balls and k disjoint unknotted, unlinked solid tori $S^1 \times D^2$. Equivalently, $M^s_{n,k}$ is the connected sum of n copies of $S^1 \times S^2$, s balls, and k solid tori. If $\mathrm{Diff}(M^s_{n,k})$ denotes the group of orientation-preserving diffeomorphisms of $M^s_{n,k}$ that restrict to the identity on the boundary, then $A^s_{n,k}$ is the quotient of the mapping class group $\pi_0\mathrm{Diff}(M^s_{n,k})$ by the subgroup generated by Dehn twists along embedded spheres. This subgroup is rather small, the product of finitely many $\mathbb{Z}_2$'s.

It seemed worthwhile to write the proof in its most natural level of generality. Thus we actually consider manifolds $M^s_{n,k}$ obtained from a fixed compact orientable manifold N by taking the connected sum with n copies of $S^1 \times S^2$, s balls, and k solid tori, where we assume that none of the connected summands of N has universal cover a counterexample to the Poincaré conjecture. We let now $A^s_{n,k}$ denote the mapping class group $\pi_0\mathrm{Diff}(M^s_{n,k})$ with twists along 2-spheres factored out. Again these generate just a product of finitely many $\mathbb{Z}_2$'s. When $s \geq 1$ there are natural stabilizations with respect to n, k, and s by enlarging $M^s_{n,k}$ and extending diffeomorphisms by the identity. The more general form of the preceding theorem is then:

Theorem B *The maps $H_i(A^s_{n,k}) \to H_i(A^s_{n+1,k})$, $H_i(A^s_{n,k}) \to H_i(A^s_{n,k+1})$, and $H_i(A^s_{n,k}) \to H_i(A^{s+1}_{n,k})$ induced by stabilization are isomorphisms for $n \geq 3i+3$, $k \geq 0$, and $s \geq 1$.*

One might hope that these stability results could be part of a program to extend the work of Madsen–Weiss on mapping class groups of surfaces to 3-manifolds, but there are some obstacles to doing this. For surfaces the contractibility of the components of Diff was an essential step in the Madsen–Weiss program, allowing π_0Diff to be replaced by Diff. However, the components of $\mathrm{Diff}(M^s_{n,k})$ are far from contractible, according to the results of [12], and the homology of $\mathrm{Diff}(M^s_{n,k})$ itself does not stabilize. This suggests that for 3-manifolds a different sort of stabilization than by connected sum might be better if one wanted to use the Madsen–Weiss techniques, a stabilization that stays within the realm of irreducible 3-manifolds where Diff often has contractible components, as shown in [5] and [3].

Here is how the paper is organized. The first section is devoted to showing the definition of $A^s_{n,k}$ in terms of mapping class groups agrees with the graph-theoretic definition. The second section proves some preliminary properties of systems of spheres and disks in a 3-manifold, generalizing [7]. These properties are used in the third section to prove that the complexes of disks and spheres we are interested in are highly connected. Then in the fourth section the spectral sequence arguments proving homological stability are given.

We have tried to make the paper largely independent of [7], but the reader might still find it helpful to consult this earlier paper at certain points.

The second author was supported by a Marie Curie Fellowship of the European Community under contract number HPMF-CT-2002-01925.

1 Diffeomorphism groups

In the introduction we first defined $A^s_{n,k}$ as the group of path-components of the space of homotopy equivalences of the graph $G^s_{n,k}$ that fix the s basepoints and the k distinguished circles. A homotopy equivalence satisfying this condition is in fact a homotopy equivalence relative to the basepoints and distinguished circles, according to Proposition 0.19 of [8], so $A^s_{n,k}$ is indeed a group, with inverses. When $s \geq 1$, by assigning to each homotopy equivalence its induced automorphism of $\pi_1 G^s_{n,k} = F_{n+k}$ one obtains a short exact sequence

$$1 \longrightarrow \mathbb{Z}^k \times (F_{n+k})^{s-1} \longrightarrow A^s_{n,k} \longrightarrow \mathrm{Aut}_{n,k} \longrightarrow 1$$

The copy of $\mathbb{Z}^k$ in the kernel is generated by the homotopy equivalences produced by rotating one of the distinguished circles through 360 degrees and dragging along the arc connecting this circle to the basepoint so that this arc wraps once around the circle. The ith copy of $F_{n+k} = \pi_1 G^s_{n,k}$ is identified with the homotopy classes of maps from the interval to $G^s_{n,k}$ mapping the endpoints to the endpoints of the ith extra edge.

Our aim in this section is to relate $A^s_{n,k}$ to the mapping class group of $M^s_{n,k}$, where $M^s_{n,k}$ is the connected sum

$$M^s_{n,k} = \big(\#_n(S^1 \times S^2)\big) \,\#\, \big(\#_k(S^1 \times D^2)\big) \,\#\, \big(\#_s D^3\big)$$

For an orientable 3–manifold M we denote by $\mathrm{Diff}(M)$ the group of orientation-preserving diffeomorphisms of M that fix its boundary pointwise. When the boundary of M is nonempty, as will usually be the case in this paper, diffeomorphisms that restrict to the identity on the boundary automatically preserve orientation.

Theorem 1.1 *There is an exact sequence*

$$1 \longrightarrow K^s_{n,k} \longrightarrow \pi_0 \mathit{Diff}(M^s_{n,k}) \longrightarrow A^s_{n,k} \longrightarrow 1$$

where the kernel $K^s_{n,k}$ is the subgroup of $\pi_0 \mathit{Diff}(M^s_{n,k})$ generated by Dehn twists along embedded 2–spheres. This subgroup is a product of at most $n+k+s$ copies of $\mathbb{Z}_2$.

In the case $k=0$ and $s=0,1$ when $A^1_{n,0}=\mathrm{Aut}(F_n)$ and $A^0_{n,0}=\mathrm{Out}(F_n)$, this is a theorem of Laudenbach [15, III.4.3] who showed that $K^0_{n,0}=K^1_{n,0}$ is the product of exactly n copies of $\mathbb{Z}_2$. (Laudenbach considered diffeomorphisms of $M^0_{n,0}$ fixing a point rather than diffeomorphisms of $M^1_{n,0}$ fixing the boundary, but the group of isotopy classes is the same in both cases.) For our purposes we only need to know that $K^s_{n,k}$ is generated by twists along 2–spheres. As we will see, twists along spheres act trivially on embedded spheres and disks, up to isotopy, so the natural action of the diffeomorphism group of M on the complexes of isotopy classes of spheres and disks that we define in Section 3 will induce an action of $A^s_{n,k}$ on these complexes.

Proof We first consider the case when $s=1$. As before, we drop $s=1$ from the notation so that $A_{n,k}=A^1_{n,k}$ etc. The first step will be to construct a homomorphism $\alpha\colon \pi_0\mathrm{Diff}(M_{n,k})\to A_{n,k}$. To do this we first describe a natural embedding $G_{n,k}\hookrightarrow M_{n,k}$. We can construct $M_{n,k}$ in the following way. Start with the space X obtained from a ball B by removing the interiors of $2n+k$ disjoint subballs B_i, $i=1,\cdots,2n+k$. Then $M_{n,k}$ is the quotient space of X obtained by identifying ∂B_{2i-1} with ∂B_{2i} for $i=1,\cdots,n$ to form nonseparating spheres $S_i\subset M_{n,k}$ and identifying two disjoint disks D'_i and D''_i in ∂B_i for $i=2n+1,\cdots,2n+k$ to form disks $D_i\subset M_{n,k}$. Construct a tree $T\subset X$ by

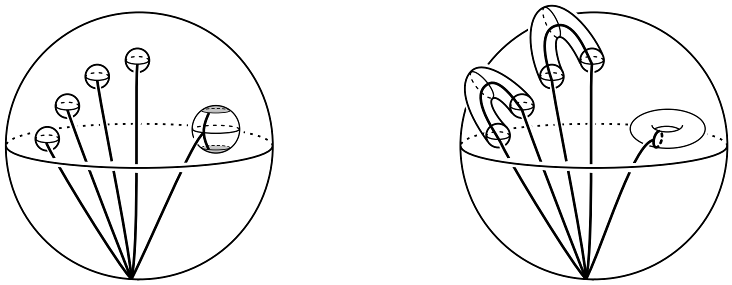

Figure 2: The graph $G_{n,k}$ in $M_{n,k}$

taking arcs a_i joining a basepoint $x_0\in\partial B$ to points $x_i\in\partial B_i$ (with x_i disjoint from the disks D'_i and D''_i for $i=2n+1,\cdots,2n+k$), together with arcs b_i in ∂B_i for $i=2n+1,\cdots,2n+k$ joining points $x'_i\in\partial D'_i$ and $x''_i\in\partial D''_i$ and passing through x_i. We may assume the points x_i, x'_i, and x''_i are chosen so that the pairs (x_{2i-1},x_{2i}) for $i=1,\cdots,n$ and the pairs (x'_i,x''_i) for $i=2n+1,\cdots,2n+k$ match up under the identifications which form $M_{n,k}$, and then the image of T in $M_{n,k}$ is a copy of $G_{n,k}$. Let $i\colon G_{n,k}\to M_{n,k}$ be this embedding. With this embedding the k distinguished circles of $G_{n,k}$ lie in $\partial M_{n,k}$ so we call them *boundary circles*. There is a retraction $X \to T$ obtained by first deformation retracting X onto the union of T with the spheres ∂B_i and then retracting this union onto T by collapsing each sphere ∂B_i to its intersection with T, which is either a point x_i or an arc b_i. These retractions can be taken to respect the identifications which produce $M_{n,k}$, so they induce a retraction $r\colon M_{n,k} \to i(G_{n,k})$. This retraction is not a homotopy equivalence but it does induce an isomorphism on π_1, as does the inclusion map i.

Now we are able to define a map $\alpha\colon \pi_0\mathrm{Diff}(M_{n,k}) \to A_{n,k}$ by sending a diffeomorphism $\varphi\colon M_{n,k} \to M_{n,k}$ to the map $r\varphi i\colon G_{n,k} \to G_{n,k}$. This composition is a homotopy equivalence since it induces an isomorphism on π_1. Since it fixes the basepoint and the boundary circles, it therefore represents an element of $A_{n,k}$.

To show that α is a homomorphism, we need the following fact: For every map $j\colon G_{n,k} \to M_{n,k}$ that agrees with i on the basepoint and boundary circles there is a homotopy $irj \simeq j$ relative to the basepoint and boundary circles. If this is true then by taking $j = \varphi i$ we obtain a homotopy $(r\psi i)(r\varphi i) \simeq r(\psi\varphi)i$. Hence α is a homomorphism.

To verify the fact, it suffices to homotope j to have image in $i(G_{n,k})$ by a homotopy relative to the basepoint and boundary circles. Such a homotopy exists since i induces a surjection on π_1, so the pair $\big(M_{n,k}, i(G_{n,k})\big)$ is 1–connected, which means that any map $(I, \partial I) \to \big(M_{n,k}, i(G_{n,k})\big)$ is homotopic, fixing ∂I, to a map into $i(G_{n,k})$.

The next step of the proof is to show that α is surjective. Following the method of [15] this will be done by showing that generators for $A_{n,k}$ can be realized by diffeomorphisms of $M_{n,k}$. According to [14] the group $A_{n,k}$ has the following generators, described by their effect on a basis $\{x_1, \cdots, x_n, y_1, \cdots, y_k\}$ for $F_{n+k} = \pi_1(G_{n,k})$, where we write the action on generators only when it is nontrivial.

$P_{i,j}$	$x_i \to x_j$ and $x_j \to x_i$	$1 \le i, j \le n,\ i \ne j$
I_i	$x_i \to x_i^{-1}$	$1 \le i \le n$
$(x_i; x_j)$	$x_i \to x_i x_j$	$1 \le i, j \le n,\ i \ne j$
$(x_i; y_j)$	$x_i \to x_i y_j$	$1 \le i \le n,\ 1 \le j \le k$
$(x_i^{-1}; y_j)$	$x_i \to y_j^{-1} x_i$	$1 \le i \le n,\ 1 \le j \le k$
$(y_i^{\pm}; x_j)$	$y_i \to x_j^{-1} y_i x_j$	$1 \le i \le k,\ 1 \le j \le n$
$(y_i^{\pm}; y_j)$	$y_i \to y_j^{-1} y_i y_j$	$1 \le i, j \le k$

When $i = j$ in the last set of generators one has the trivial automorphism of F_{n+k} but a nontrivial element of $A_{n,k}$, a generator of the subgroup $\mathbb{Z}^k$.

The first three types of generators involve only the x_i's and can be realized as in [15]. Thinking of the $S^1 \times S^2$ summands of $M_{n,k}$ as handles, the generator $P_{i,j}$ interchanges the ith and jth handles, I_i switches the two ends of the ith handle, and $(x_i; x_j)$ slides one end of the ith handle over the jth handle. The result of this slide is a Dehn twist along a torus that encloses one end of the ith handle and passes over the jth handle. The next two types of generators, $(x_i; y_j)$ and $(x_i^{-1}; y_j)$, can be realized by similar slides of one end of the ith handle along a loop that represents y_j, going through the jth $S^1 \times D^2$ summand of $M_{n,k}$. For the last two types of generators we think of the ith $S^1 \times D^2$ summand as a punctured solid torus attached to the rest of $M_{n,k}$ along a sphere that gives the connected sum operation, and then we slide this punctured solid torus over the jth handle to realize the conjugation $(y_i^{\pm}; x_j)$, or through the jth $S^1 \times D^2$ summand to realize $(y_i^{\pm}; y_j)$ for $i \neq j$. Like the slide producing $(x_i; x_j)$, these slides also yield Dehn twists along tori. The last remaining type of generator $(y_i^{\pm}; y_i)$ is realized by a Dehn twist along a torus parallel to the ith boundary torus of $M_{n,k}$, with the twisting in the direction of the S^1 factor of the summand $S^1 \times D^2$.

The last step of the proof is to describe the kernel of α, using the following commutative diagram

$$\begin{tikzcd}
1 \arrow[r] & K_{n,k} \arrow[r] & \pi_0 \mathrm{Diff}(M_{n,k}) \arrow[r, "\alpha"] \arrow[dr] & A_{n,k} \arrow[r] \arrow[d] & 1 \\
 & & & \mathrm{Aut}(F_{n+k}) &
\end{tikzcd}$$

where the top row is exact and the diagonal map is the action on π_1. Let φ be a diffeomorphism representing a class in the kernel $K_{n,k}$, so in particular φ induces the identity on $\pi_1(M_{n,k})$. By [15, A-III.3.2] we know that φ also induces the identity on $\pi_2(M_{n,k})$. Let Σ be a system of $n+k$ spheres in $M_{n,k}$ consisting of the n spheres S_i in the $S^1 \times S^2$ summands together with spheres $S_{n+1}, \cdots, S_{n+k}$ separating off the $S^1 \times D^2$ summands. Splitting $M_{n,k}$ along Σ produces a ball with $2n+k$ punctures, along with k copies of a punctured solid torus. Because φ induces the identity on π_2, it can be isotoped so that it is the identity on Σ, using Laudenbach's homotopy-implies-isotopy theorem for 2–spheres, taking care of the spheres S_i one by one. (For the induction step one uses [15, V.4.2] which says that homotopic spheres that are disjoint from a sphere S and not isotopic to S are still isotopic in the complement of S; this is reproved in our Theorem 2.2.) So we can consider φ separately on the components of $M - \Sigma$. On the punctured ball it is isotopic to a product of Dehn twists along a subset of the $2n+k$ boundary spheres of the punctures.

According to Lemma 1.2 which follows below, the mapping class group of a punctured solid torus is $\mathbb{Z}_2 \times \mathbb{Z}$ with the first factor generated by a Dehn twist along the boundary sphere and the second factor generated by a Dehn twist along the boundary torus, in the S^1 direction. The latter twist is the diffeomorphism realizing one of the generators $(y_i; y_i)$ of the $\mathbb{Z}^k$ subgroup of $A_{n,k}$, so only the $\mathbb{Z}_2$ factor lies in the kernel $K_{n,k}$.

The conclusion of all this is that the diffeomorphism φ representing an element of the kernel of α is isotopic to a product of Dehn twists along a subset of the spheres of Σ. The proof of the theorem for $s = 1$ is then completed by observing that all twists along 2–spheres lie in the kernel of α since for a given 2–sphere, the map $i\colon G_{n,k} \to M_{n,k}$ can be homotoped to intersect this sphere only in the fixed points of a twist along the sphere.

An argument for deducing the theorem for arbitrary s from the case $s = 1$ is given in Proposition 1 of [11] when $k = 0$, and the same argument works equally well for arbitrary k. $\square$

Lemma 1.2 *$\pi_0 Diff(M^1_{0,1}) \cong \mathbb{Z}_2 \times \mathbb{Z}$ where the $\mathbb{Z}_2$ is generated by a Dehn twist along the boundary sphere and the $\mathbb{Z}$ is generated by a Dehn twist along the boundary torus in the direction of the S^1 factor of the summand $S^1 \times D^2$ of $M^1_{0,1} = (S^1 \times D^2) \# D^3$.*

Proof Consider the fibration

$$\mathrm{Diff}(M^1_{0,1}) \longrightarrow \mathrm{Diff}(S^1 \times D^2) \longrightarrow E(D^3, S^1 \times D^2)$$

where $E(D^3, S^1 \times D^2)$ is the space of embeddings of a ball in the interior of $S^1 \times D^2$ and the projection to E is obtained by restricting diffeomorphisms to a chosen ball in the interior of $S^1 \times D^2$. Since elements of $\mathrm{Diff}(S^1 \times D^2)$ fix the boundary pointwise, they preserve orientation, so we may as well take $E(D^3, S^1 \times D^2)$ to consist of orientation-preserving embeddings.

The total space $\mathrm{Diff}(S^1 \times D^2)$ of the fibration is contractible. This can be seen by looking at a second fibration

$$\mathrm{Diff}(D^3) \longrightarrow \mathrm{Diff}(S^1 \times D^2) \longrightarrow E(D^2, S^1 \times D^2)$$

whose projection map is restriction to a meridian disk. The base space consists of embeddings of a disk in $S^1 \times D^2$ fixed on the boundary. It is a classical fact that this space is path-connected, and in fact it is contractible by [5]. The fiber consists of diffeomorphisms of $S^1 \times D^2$ that are fixed on the boundary and a meridian disk, so this can be identified with $\mathrm{Diff}(D^3)$ which is contractible by [6]. Hence $\mathrm{Diff}(S^1 \times D^2)$ is also contractible.

The base space of the first fibration is homotopy equivalent to $S^1 \times SO(3)$. The long exact sequence of the fibration then gives an isomorphism

$$\pi_0\mathrm{Diff}(M^1_{0,1}) \cong \pi_1\big(S^1 \times SO(3)\big) \cong \mathbb{Z} \times \mathbb{Z}_2$$

where the Z_2 corresponds to a Dehn twist about the boundary sphere of $M^1_{0,1}$ and the $\mathbb{Z}$ is generated by the diffeomorphism resulting from an isotopy dragging a ball in the interior of $S^1 \times D^2$ around the S^1 factor. This diffeomorphism is equivalent to a Dehn twist about the boundary torus of $M^1_{0,1}$. $\square$

2 Normal form

We are primarily interested in the manifolds $M^s_{n,k}$, but it will be convenient and not much extra work to consider a more general compact manifold M that is the connected sum of n copies of $S^1 \times S^2$, s balls, and k irreducible orientable manifolds Q_i that are neither a ball nor a counterexample to the Poincaré conjecture. The latter condition is needed since we will be using Laudenbach's theorem that homotopy implies isotopy for embedded spheres, which requires this hypothesis. We will in fact assume that no Q_i has universal cover a counterexample to the Poincaré conjecture, although it is possible that this stronger assumption could be avoided by taking extra care at certain points in the proofs.

We will study $\pi_0\mathrm{Diff}(M)$ by studying isotopy classes of embedded spheres and disks in M. By an embedded sphere or disk we mean a submanifold of M that is diffeomorphic to a sphere or disk, but without a specific diffeomorphism being chosen. An isotopy of an embedded sphere or disk is a smooth one-parameter family of such submanifolds. This is the same as choosing a diffeomorphism from the standard sphere or disk to the given submanifold and varying this map to M by isotopy. More generally, if we allow the map to vary by homotopy instead of isotopy we can speak of two embedded spheres or disks being homotopic as well as isotopic.

Besides individual spheres and disks, we also consider systems of finitely many embedded spheres and disks in M. The spheres and disks in a system are assumed to be disjoint, except that the boundary circles of different disks are allowed to coincide. We will assume the disks have their boundaries contained in a fixed finite collection C of disjoint circles in ∂M. In the end we will only need the case that C is a single circle, either in a boundary sphere of M or in the boundary torus of a solid torus summand $Q_i = S^1 \times D^2$, with the circle being a meridian $\{x\} \times \partial D^2$ in the latter case. However, to prove the results in Section 3 we will have to allow larger collections C. It will suffice to assume:

- C consists of at most two circles in each sphere component of ∂M and at most one circle in the boundary torus of each summand $Q_i = S^1 \times D^2$, this circle being a meridian.

The reason for allowing two circles in a boundary sphere is that we will be cutting M open along systems of disks with boundaries in C. Cutting M along a disk bounded by a circle of C that is a meridian in a boundary torus replaces the torus by a boundary sphere containing two copies of the original circle of C (bounding two copies of the disk). On the other hand, cutting M along a disk bounded by a circle of C in a boundary sphere replaces the sphere by two spheres, each containing a copy of the original circle of C, so the number of circles in a boundary sphere does not increase.

The spheres and disks in a system will be assumed to be nontrivial. For spheres this means that they do not bound balls in M and are not isotopic to boundary spheres of M disjoint from C. Nontriviality for disks means they are not isotopic, fixing their boundary, to disks in ∂M containing no circles of C in their interior. Another assumption we make is that no two spheres in a system are isotopic, and no two disks are isotopic fixing their boundaries. This is equivalent to saying that no two spheres bound a product $S^2 \times I$ in M and no two disks bound a ball.

To save words in the rest of the paper we will always assume that isotopies and homotopies of disks in a system fix the boundaries of these disks.

Our purpose in this section is to develop a notion of normal form for systems of spheres and disks in the manifold M, generalizing [7] which dealt with systems of spheres in the manifolds $M^s_{n,0}$. The simplest definition of normal form is to say that a system S of spheres and disks in M is in *normal form* with respect to a fixed maximal sphere system Σ if S is transverse to Σ and the number of circles of $S \cap \Sigma$ is minimal among all systems isotopic to S and transverse to Σ. As we will see, in the case of the manifolds $M^s_{n,0}$ this boils down to almost the same thing as the more complicated definition given in [7], with only a minor difference in how spheres of S that are isotopic to spheres of Σ are treated.

The advantage of the present definition is that the existence of a system in normal form within each isotopy class of systems is obvious. This allows us to focus on what is the real point of normal form, which is to obtain a reasonable uniqueness statement. The strongest uniqueness statement would be that if two normal form systems are isotopic then they are isotopic through normal form systems. However, this could hold only if one first eliminated certain knotting and linking phenomena. Fortunately knotting and linking are not really an issue in view of the fact that Laudenbach's homotopy-implies-isotopy theorem for spheres and sphere systems holds also for systems of spheres and disks, as we show in Theorem 2.2 below. We can then avoid knotting and linking problems by making the following definition. Two systems S and S' transverse to Σ are called *equivalent* if there is a homotopy S_t from S to S' through immersions such that S_t is transverse to Σ for all t and the self-intersections of S_t are disjoint from Σ for all t. We also allow a sphere of S that is disjoint from Σ and hence isotopic to a sphere of Σ to be moved from one side of this sphere to the other. The main result is then:

Theorem 2.1 *Isotopic systems in normal form are equivalent.*

Before beginning the proof of this we will need to establish a more basic result:

Theorem 2.2 *Homotopic systems are isotopic. In particular, equivalent systems are isotopic.*

Proof For single spheres this is Theorem III.1.3 of [15], and the generalization to systems of spheres follows from the Lemma on page 124 of [15]. We need to extend the proof to allow disks as well. First consider systems consisting of a single disk, so let D_0 and D_1 be two disks that are homotopic. If the common boundary of these two disks lies in a torus component T of ∂M, we can associate to D_0 and D_1 spheres S_0 and S_1 that are the boundaries of ε-neighborhoods of $T \cup D_0$ and $T \cup D_1$. Alternatively, we can view S_i as the result of surgering T along D_i. A homotopy from D_0 to D_1 induces a homotopy from S_0 to S_1. By Laudenbach's theorem, S_0 and S_1 are then isotopic, so there is an ambient isotopy of M, fixing ∂M, taking S_0 to S_1 and taking D_0 to a new disk D_0' that lies in the ε-neighborhood of $T \cup D_1$. This neighborhood is a once-punctured solid torus, the manifold $M^1_{0,1}$. Any two disks in $M^1_{0,1}$ are equivalent under a diffeomorphism of $M^1_{0,1}$ fixing $\partial M^1_{0,1}$ since after filling in the puncture they are isotopic fixing their boundary. We saw in the previous section that the mapping class group of $M^1_{0,1}$ is $\mathbb{Z}_2 \times \mathbb{Z}$ generated by Dehn twists along the boundary sphere and boundary torus. Twists along the boundary sphere act trivially on disks with boundary on the torus. So, up to isotopy, D_0' and D_1 are equivalent under twists along the boundary torus. Looking in the universal cover of M, lifts of D_0' and D_1 having the same boundary circle are homotopic and therefore homologous. But the homology classes represented by the lifts of the disks obtained by the $\mathbb{Z}$'s worth of Dehn twists along T applied to a single disk are all distinct, so it follows that D_0' and D_1 are isotopic.

The other possibility is that the common boundary circle of D_0 and D_1 lies in a sphere S of ∂M. The argument in this case is similar. An ε-neighborhood of $S \cup D_i$ is a 3-punctured sphere bounded by S and two other spheres S_i' and S_i''. A homotopy from D_0 to D_1 induces a homotopy from $S_0' \cup S_0''$ to $S_1' \cup S_1''$, so from the known case of sphere systems we can conclude that these two sphere systems are isotopic. Arguing as before, this gives a disk D_0' isotopic to D_0 lying in the ε-neighborhood of $S \cup D_1$. Since D_0' is not a trivial disk, it is unique up to isotopy in this neighborhood, which finishes the argument in this case.

Now we turn to the general case. Let $X = X_1 \cup \cdots \cup X_n$ and $Y = Y_1 \cup \cdots \cup Y_n$ be two homotopic systems of spheres and disks, with X_i corresponding to Y_i under the homotopy. By induction on n we may assume the subsystems $X_1 \cup \cdots \cup X_{n-1}$ and $Y_1 \cup \cdots \cup Y_{n-1}$ are isotopic, so after an ambient isotopy we may assume they coincide. The surfaces X_n and Y_n are homotopic in M so it will suffice to show they are homotopic in the complement of $X_1 \cup \cdots \cup X_{n-1}$. By another induction the problem reduces further to the case $n = 2$, that is, to show that for systems $X_1 \cup X_2$ and $Y_1 \cup Y_2$ with $X_1 = Y_1$, if X_2 and Y_2 are homotopic in M then they are homotopic in $M - X_1$.

To simplify the notation, let the two systems be $X \cup Z$ and $Y \cup Z$. In the universal cover $\widetilde{M}$ choose a lift $\widetilde{X}$ of X and lift the homotopy from X to Y to a homotopy from $\widetilde{X}$ to a lift $\widetilde{Y}$ of Y. If X and Y are disks then $\widetilde{X}$ and $\widetilde{Y}$ have the same boundary since the homotopy from X to Y fixes the boundary. After we perturb X and Y to meet transversely we can triangulate M so that X, Y, and Z are all subcomplexes, and we can lift this triangulation to a triangulation of $\widetilde{M}$. Since $\widetilde{X}$ and $\widetilde{Y}$ are homotopic in $\widetilde{M}$ they are homologous, so there is a simplicial 3-chain bounded by $\widetilde{X}$ and $\widetilde{Y}$, suitably oriented. Geometrically, this chain is a subcomplex W of $\widetilde{M}$ bounded by $\widetilde{X} \cup \widetilde{Y}$. Note that W is unique since there are no 3-cycles in $\widetilde{M}$. Also W is connected since the component of W containing $\widetilde{X}$ must also contain $\widetilde{Y}$. This is clear if $\widetilde{X} \cap \widetilde{Y} \neq \varnothing$, in particular if $\widetilde{X}$ and $\widetilde{Y}$ are disks. If $\widetilde{X}$ and $\widetilde{Y}$ are disjoint spheres they cannot be trivial in $H_2(\widetilde{M}) \cong \pi_2(\widetilde{M}) \cong \pi_2(M)$, so the component of W containing $\widetilde{X}$ must contain $\widetilde{Y}$ in this case as well.

Let $\widetilde{Z}$ be any lift of Z. We claim that $\widetilde{Z}$ is disjoint from W, except perhaps for points in $\partial\widetilde{Z}$ when $\widetilde{Z}$ is a disk. For suppose there are other points of intersection of $\widetilde{Z}$ with W. Then $\widetilde{Z}$ is entirely contained in W since $\partial W = \widetilde{X} \cup \widetilde{Y}$ and $\widetilde{Z}$ is disjoint from $\widetilde{X}$ and $\widetilde{Y}$, apart from $\partial\widetilde{Z}$. Since $\widetilde{Z}$ separates $\widetilde{M}$, it splits W into two parts W_1 and W_2. There are only two possibilities:

- W_1 or W_2 has $\widetilde{Z}$ as its complete boundary. This is obviously impossible if $\widetilde{Z}$ is a disk, and if it is a sphere this would make $\widetilde{Z}$ homologous to zero in $\widetilde{M}$, which is not the case since Z is a nontrivial sphere in M.

- One of W_1, W_2 is bounded by $\widetilde{X} \cup \widetilde{Z}$ and the other is bounded by $\widetilde{Y} \cup \widetilde{Z}$. In particular this would say $\widetilde{X}$ and $\widetilde{Z}$ are homologous. If one is a disk, the other would also have to be a disk with the same boundary and the sphere $\widetilde{X} \cup \widetilde{Z}$ would be homologous to zero, hence homotopic to zero, making $\widetilde{X}$ homotopic to $\widetilde{Z}$, forcing the same to be true for X and Z. This would make X and Z isotopic, contrary to the assumption that $X \cup Z$ is a system. The other alternative is that both $\widetilde{X}$ and $\widetilde{Z}$ are spheres, but then they would be homotopic since they are homologous, again forcing X and Z to be homotopic and therefore isotopic, a contradiction.

Now if we split $\widetilde{M}$ along all the lifts of Z, W will lie in one of the resulting components, a simply-connected manifold N. In the case that $\widetilde{X}$ and $\widetilde{Y}$ are spheres, this says they are homologous in N and hence homotopic in N. Projecting to M, we deduce that X and Y are homotopic in the complement of Z, the conclusion we wanted. In the opposite case that $\widetilde{X}$ and $\widetilde{Y}$ are disks, they together give a map $S^2 \to \widetilde{M}$ that is homologous to zero in N, hence homotopic to zero in N, making $\widetilde{X}$ and $\widetilde{Y}$ homotopic in N. Again this makes X and Y homotopic in the complement of Z. □

Before proving Theorem 2.1 we need to understand what normal form systems look like. So suppose the system S is in normal form with respect to the maximal sphere system Σ. Splitting M along Σ produces connected manifolds P_i, and splitting S along Σ produces connected surfaces which we call *pieces* of S. Components of S disjoint from Σ are either spheres parallel to components of Σ or disks that are of one of two types:

(a) In a $P_i = S^2 \times I$ that has a component of Σ as one boundary sphere and a component of ∂M containing two circles of C as the other boundary sphere, one can have a disk with boundary on one circle of C, the disk cutting off a ball from P_i that contains the other circle of C.

(b) In a P_i that is a punctured $S^1 \times D^2$ having a circle of C in its boundary torus, one can have a disk with boundary this circle of C.

For pieces that actually meet Σ there are the following possibilities:

(1) In a P_i that is a 3-punctured sphere, a piece can be:

   - a disk with boundary on one boundary sphere and separating the other two boundary spheres, or

- a cylinder with its two boundary circles on two different boundary spheres of P_i, or
- a pair of pants with each boundary circle on a different boundary sphere of P_i.

(2) In a $P_i = S^2 \times I$ with a component of Σ at one end and a component of ∂M at the other, a piece can be a cylinder with one boundary circle on each boundary sphere, the boundary circle in ∂M being a circle of C.

(3) In a P_i that is a once-punctured summand Q_j, a piece can be the boundary surface of a tubular neighborhood of a tree in P_i obtained by joining $p \geq 2$ points in the boundary sphere of P_i to an interior point of P_i by disjoint arcs such that a lift of this tree to the universal cover $\widetilde{P}_i$ of P_i has its endpoints all in distinct boundary spheres of $\widetilde{P}_i$. This is equivalent to saying that the boundary circles of a lift of the piece all lie in different components of $\partial\widetilde{P}_i$.

(4) In a P_i that is a once-punctured $S^1 \times D^2$ with a circle of C in its boundary torus a piece can be of the type in (3) or it can be:

- a cylinder joining the circle of C in the boundary torus to the boundary sphere, or
- a pair of pants with one boundary circle on C and two boundary circles on the boundary sphere of P_i, such that after lifting to the universal cover the three boundary circles of a lift of the pair of pants all lie in different components of $\partial\widetilde{P}_i$.

To show that if the number of circles of intersection of S with Σ is minimal then the pieces have only the types listed we argue as follows. By a sequence of surgeries on S the circles of $S \cap \Sigma$ can be eliminated one by one. This converts S into a collections of disks and spheres disjoint from Σ, and S is obtained from this collection by the inverse sequence of tubing operations, where the tubes might be nested one inside another. So each piece of S is obtained from a sphere or disk in a P_i by inserting tubes that run from this sphere or disk to one or more of the boundary spheres of P_i that are in Σ. See Figure 3 for an example.

Consider first the case of a sphere S_0 connected to boundary spheres of P_i by tubes. Since Σ is maximal, S_0 must be trivial in P_i, either bounding a ball or parallel to a sphere of ∂P_i. Suppose first that S_0 bounds a ball. If there were nesting among the tubes from S_0, then there would be an outermost tube connecting S_0 on the outside to a sphere of Σ and a next-outermost tube inside this going from the inside of S_0 to the same sphere of Σ.

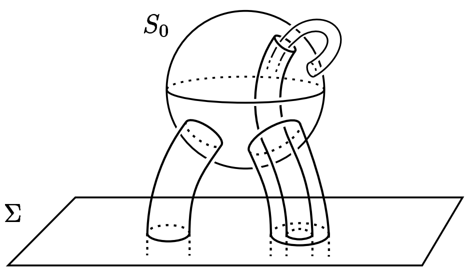

Figure 3

There is an isotopy of S that takes the end of this inner tube attached to S_0 and slides it along S_0 over to the outer tube and then along the outer tube until it lies on the other side of Σ. If there are other tubes inside this inner tube, they are to be carried along during this sliding process. After this isotopy the number of circles of $S \cap \Sigma$ has temporarily increased, but now there is a regular homotopy of the new tube that pushes it across Σ so as to eliminate both its new and old intersections with Σ. Again any tubes inside this tube can be carried along during the regular homotopy. The net result is a new embedding of S which is homotopic to the original one and therefore isotopic to it by Theorem 2.2, and which has fewer circles of intersection with Σ than the original S had, contrary to the minimality assumption. Thus nesting of tubes is ruled out. Similarly, if two tubes joining S_0 to the same sphere of Σ were homotopic in P_i, then we could slide one tube across the other and again decrease the number of circles of $S \cap \Sigma$. Thus homotopic tubes are ruled out, or in terms of the universal cover, tubes joining a lift of S_0 to the same lift of a sphere of Σ. Since there have to be at least two tubes from S_0, we see in this case that the piece is as in (3) or is a cylinder or pair of pants as in (1).

Suppose now that S_0 is a sphere parallel to a sphere of ∂P_i. If there is a tube from S_0 to this sphere of ∂P_i, then the piece can also be obtained by joining a sphere bounding a ball to ∂P_i by tubes, reducing to the previous case. If there is no tube joining S_0 to the sphere of ∂P_i it is parallel to, then P_i must be a 3-punctured sphere as in (1). By the reasoning in the preceding paragraph there can be no nesting of tubes, and there can be at most one tube joining S_0 to each of the other two spheres of ∂P_i since P_i is simply-connected. Thus the piece is a disk or a cylinder, and it must be of the type in (1).

The other possibility is that the piece is a disk D_0 with tubes connecting it to Σ. If P_i is as in (2), then arguing as before we see that there can be only one tube and the piece is a cylinder of the type described. If P_i is as in (4) then D_0

is a nontrivial disk in P_i joined by one or more tubes to the boundary sphere of P_i. The disk D_0 is unique up to isotopy if we allow its boundary to move in the boundary torus of P_i, so we may assume D_0 is obtained by applying Dehn twists along this torus to a meridian disk of P_i. Minimality rules out nesting of tubes, and it rules out two tubes being connected to the same side of D_0 since if there were, the two tubes would be homotopic. Thus either there is one tube or there are two tubes attached on opposite sides of D_0. These are the configurations in (4).

Proof of Theorem 2.1 What we will show is that if two isotopic systems S and S' are in normal form in the sense that their pieces have the forms described above, then they are equivalent.

Let $\widetilde{\Sigma}$ be the preimage of Σ in the universal cover $\widetilde{M}$ of M. We first consider the case that S has a single component, a sphere or disk. Let $\widetilde{S}$ be a lift of S to $\widetilde{M}$. An isotopy of S to S' lifts to an isotopy of $\widetilde{S}$ to a lift $\widetilde{S}'$ of S'. Our goal in the first part of the proof will be to show:

- **(I)** $\widetilde{S}$ and $\widetilde{S}'$ are isotopic staying transverse to $\widetilde{\Sigma}$, except in the special case that they are spheres isotopic to the same sphere of $\widetilde{\Sigma}$ but lying on opposite sides of this sphere.

The proof will involve several steps. To begin, let us choose an orientation for $\widetilde{S}$. This is equivalent to choosing a transverse orientation once we fix an orientation for $\widetilde{M}$. The orientation for $\widetilde{S}$ carries over to an orientation for $\widetilde{S}'$ by means of the isotopy between them, and likewise for the transverse orientation.

Dual to $\widetilde{\Sigma}$ is a tree T, with a vertex for each component of $\widetilde{M}-\widetilde{\Sigma}$ and an edge for each component of $\widetilde{\Sigma}$. Similarly, dual to the collection of circles $\widetilde{S}\cap\widetilde{\Sigma}$ in $\widetilde{S}$ is a tree $T(\widetilde{S})$ with a vertex for each component of $\widetilde{S}-\widetilde{\Sigma}$ and an edge for each circle of $\widetilde{S}\cap\widetilde{\Sigma}$. The natural map $T(\widetilde{S})\to T$ is injective since it is locally injective, the pieces of S having the types listed in (1)–(4) when $S\cap\Sigma\neq\varnothing$. Thus we can view $T(\widetilde{S})$ as a subtree of T. In the same way we have a subtree $T(\widetilde{S}')$ of T.

- **(a)** $T(\widetilde{S})=T(\widetilde{S}')$, except in the case that S is a sphere isotopic to a sphere of Σ, when $T(\widetilde{S})$ and $T(\widetilde{S}')$ can be adjacent vertices of T.

To prove this it will suffice to show that $T(\widetilde{S})$ is determined by the homology class of the oriented surface $\widetilde{S}$ in $H_2(\widetilde{M},\partial\widetilde{S})$. Let e be an oriented edge of T, by which we mean an edge together with a choice of one of its two possible orientations. We can split $\widetilde{M}$ into two components along the sphere of $\widetilde{\Sigma}$

corresponding to e, and we let $\widetilde{M}_e$ be the component at the tail of e. If $T(\widetilde{S})$ is not a single vertex, then an end vertex v of $T(\widetilde{S})$ corresponds to a piece of $\widetilde{S}$ that is either a disk in a 3-punctured sphere or an annulus in a once-punctured D^3 or $S^1 \times D^2$. This disk or annulus is nonzero in $H_2(\widetilde{M}, \widetilde{M}_e \cup \partial\widetilde{S})$ where e is the edge of $T(\widetilde{S})$ with v as its head vertex. Thus if we partially order the submanifolds $\widetilde{M}_e$ by inclusion as e ranges over the oriented edges of T, then the maximal $\widetilde{M}_e$'s for which $\widetilde{S}$ is nonzero in $H_2(\widetilde{M}, \widetilde{M}_e \cup \partial\widetilde{S})$ are those corresponding to oriented edges e whose head vertex is an end vertex of $T(\widetilde{S})$, at least if $T(\widetilde{S})$ is not a single vertex. This remains true also when $T(\widetilde{S})$ is a single vertex and S is a disk. In the case that $T(\widetilde{S})$ is a single vertex and S is a sphere the maximal $\widetilde{M}_e$'s have e with its head vertex at either end of the edge of T corresponding to the component of $\widetilde{\Sigma}$ isotopic to $\widetilde{S}$. This gives a homological characterization of $T(\widetilde{S})$, proving (a).

Assuming that we are not in the exceptional case in (a) from now on, then since $T(\widetilde{S}) = T(\widetilde{S}')$, the pieces of $\widetilde{S}$ are diffeomorphic to those of $\widetilde{S}'$. In fact a stronger statement is true:

- **(b)** The pieces of $\widetilde{S}$ are isotopic within their $\widetilde{P}_i$'s to the corresponding pieces of $\widetilde{S}'$. Here we mean proper isotopies of the pieces, so their boundary circles in $\widetilde{\Sigma}$ stay in $\widetilde{\Sigma}$ during the isotopies.

This is proved by examining the various types of pieces. If a piece is the boundary of a neighborhood of a tree in $\widetilde{P}_i$ then two different embeddings of the tree having the same endpoints are homotopic fixing their endpoints since $\widetilde{P}_i$ is simply-connected, and a homotopy can be improved to an isotopy by sliding crossings of edges over the boundary spheres at their outer ends — the so-called lightbulb trick of unknotting the cord to a hanging light bulb by letting the knot drop over the bulb. This works in the present situation since the various edges in the tree all go to different boundary spheres of $\widetilde{P}_i$. This argument covers pieces of types (2) and (3) and cylinders and pairs of pants in (1). A disk in (1) is unique up to isotopy since the isotopy class of a disk in a ball with punctures is determined by how it separates the punctures. For (4) we have a piece that is the lift of a meridian disk with one or two tubes attached, and possibly with Dehn twists along the boundary torus applied to the meridian disk. Again the lightbulb trick applies in $\widetilde{P}_i$ since if there are two tubes, they connect to different boundary spheres of $\widetilde{P}_i$. Thus we have (b) in all cases.

The different ways of gluing together two pieces that lie on opposite sides of a sphere of $\widetilde{\Sigma}$ and have a boundary circle in this sphere depend only on choosing an isotopy taking one boundary circle to the other in this sphere. The space of circles in a 2-sphere has the homotopy type of the projective plane, with fundamental group $\mathbb{Z}_2$, so there are two essentially distinct isotopies taking one circle to another. This two-fold ambiguity is eliminated if we require the gluing to preserve transverse orientations for the two pieces since the space of oriented circles in a 2-sphere is simply-connected, being homotopy equivalent to S^2. Thus in order to finish the proof of assertion (I) that $\widetilde{S}$ and $\widetilde{S}'$ are isotopic staying transverse to $\widetilde{\Sigma}$, it will suffice to refine (b) to:

(c) Corresponding pieces of $\widetilde{S}$ and $\widetilde{S}'$ are isotopic preserving transverse orientations.

This is automatically true for a piece that is a pair of pants in a 3-punctured sphere since there is an isotopy of this piece that starts and ends with the same position and reverses the transverse orientation. Namely, take the three punctures to lie on a great circle in S^3 and then the family of great 2-spheres containing this great circle gives such an isotopy. (Alternatively, think of the 3-punctured sphere as a 2-punctured ball, with the two punctures aligned along a diameter of the ball, and intersect the punctured ball with the family of planes containing this diameter.) The same thing happens for a piece that is an n-punctured sphere in a $\widetilde{P}_i$ that is an n-punctured S^3. Such a $\widetilde{P}_i$ is the universal cover of a once-punctured Q_j with $\pi_1 Q_j$ of order n.

We claim that cutting $\widetilde{S}$ along $\widetilde{\Sigma}$ and regluing according to different transverse orientations in any subset of the remaining pieces changes the class of $\widetilde{S}$ in $H_2(\widetilde{M}, \partial\widetilde{S})$. If $\partial S \neq \varnothing$ it is clear that reorienting the piece containing ∂S changes the homology class since this changes its image under the boundary map to $H_1(\partial\widetilde{S})$. So we may assume the orientation of this piece is fixed. To treat the remaining cases it is convenient to pass to a quotient manifold N of $\widetilde{M}$ constructed in the following way. Consider a $\widetilde{P}_i$ containing a piece of type (3) that does not touch all the boundary spheres of $\widetilde{P}_i$. This piece is the boundary of a neighborhood X of a tree in $\widetilde{P}_i$. Let Y be the union of X with an arc joining it to a boundary sphere of $\widetilde{P}_i$ disjoint from the piece. Let Z be an ϵ-neighborhood in $\widetilde{P}_i$ of the union of Y with the boundary spheres of $\widetilde{P}_i$ that meet Y. Then ∂Z consists of some boundary spheres of $\widetilde{P}_i$ together with one sphere in the interior of $\widetilde{P}_i$. The latter sphere splits $\widetilde{M}$ into two components. Collapsing the component not containing Z to a point produces a quotient manifold of $\widetilde{M}$. Doing this collapsing operation on all such $\widetilde{P}_i$'s gives the manifold N. This still contains $\widetilde{S}$, and it contains a subcollection $\widetilde{\Sigma}_N$ of the spheres of $\widetilde{\Sigma}$, with dual tree T_N containing $T(\widetilde{S})$. Edges of T_N that touch $T(\widetilde{S})$ but are not contained in $T(\widetilde{S})$ we call *abutting* edges.

Consider the effect of changing the transverse orientation on a type (3) piece that meets all but one of the boundary spheres of the n-punctured sphere of $N - \widetilde{\Sigma}_N$ that contains it. If we fill in with a ball the one boundary sphere that it does not meet, then we have seen that there is an isotopy of the piece that reverses its transverse orientation. This isotopy sweeps across the filled-in ball exactly once, so if we take out the filled-in ball we see that reorienting the piece changes the homology class of $\widetilde{S}$ by adding the boundary sphere of the filled-in ball, with one of its two possible orientations. The same argument applies also to a type (1) piece that is an annulus. For a type (1) piece that is a disk, if we surger $\widetilde{S}$ along the boundary of this disk using one of the two disks it bounds in $\widetilde{\Sigma}$, we obtain a surface of two components, one of which is a copy of one of the other two boundary spheres of the $\widetilde{P}_i$ containing the given piece, and we can recover S from this surgered surface by a connected sum operation reversing the surgery. The connected sum operation preserves the homology class. As Figure 4 shows, changing the boundary sphere we use for the connected sum corresponds to changing the orientation of the piece.

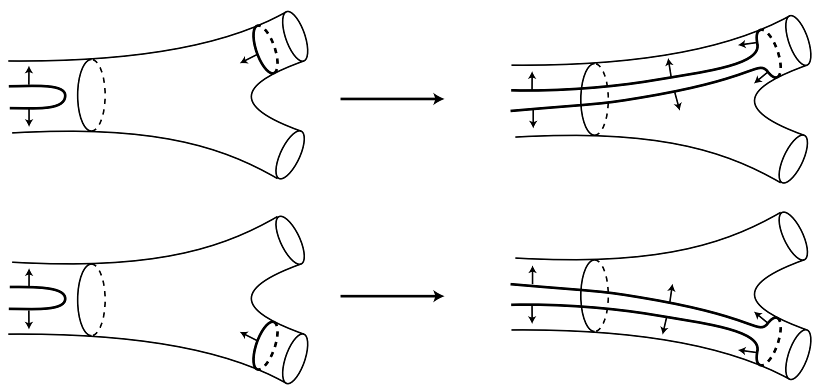

Figure 4

Hence the homology class of $\widetilde{S}$ changes by subtracting one boundary sphere of $\widetilde{P}_i$ and adding another. In terms of transverse orientations the homology class changes by adding two boundary spheres of $\widetilde{P}_i$, one oriented into $\widetilde{P}_i$ and the other out of $\widetilde{P}_i$.

Thus we see that reorienting some collection of pieces disjoint from $\partial\widetilde{S}$ changes the homology class of $\widetilde{S}$ in $H_2(N, \partial\widetilde{S})$ by adding the spheres corresponding to some of the abutting edges of T_N, with certain signs. If this signed sum of spheres were homologically a boundary it would have to consist of spheres for all the abutting edges, oriented all toward $\widetilde{S}$ or all away from $\widetilde{S}$, but this is not the case as there is at least one abutting edge corresponding to a disk. This finishes the proof of (c) and hence of (I).

We now consider the general case that S and S' have an arbitrary number of components. Choose a transverse orientation for S. This gives a transverse orientation to the circles of $S\cap\Sigma$ in Σ. In each sphere Σ_0 of Σ, each circle of $S\cap\Sigma_0$ then partitions the remaining circles of $S\cap\Sigma_0$ into those on the positive side and those on the negative side of the given circle. Let us call this partition data the *side relation* for $S\cap\Sigma$. The transverse orientation for S induces a transverse orientation for S' via the isotopy between them, so the circles of $S'\cap\Sigma$ also have a side relation. From what we have shown so far we have a bijective correspondence between the pieces of S and S' and hence also between the circles of $S\cap\Sigma$ and those of $S'\cap\Sigma$.

(II) The side relation on $S\cap\Sigma$ agrees with that on $S'\cap\Sigma$ under this bijection.

For suppose this is false. Then in some sphere Σ_0 of Σ we have circles c_p and c_q of $S\cap\Sigma_0$ with corresponding circles c'_p and c'_q of $S'\cap\Sigma_0$ such that c_q and c'_q lie on opposite sides of c_p and c'_p, respectively. Choose a lift $\widetilde{\Sigma}_0$ of Σ_0 to $\widetilde{M}$, and let $\widetilde{S}_j$ and $\widetilde{S}_k$ be the lifts of components of S containing the lifts to $\widetilde{\Sigma}_0$ of c_p and c_q, respectively. Similarly we have $\widetilde{S}'_j$ and $\widetilde{S}'_k$ containing the lifts of c'_p and c'_q to $\widetilde{\Sigma}_0$. We know that $\widetilde{S}'_j$ is isotopic to $\widetilde{S}_j$ staying transverse to $\widetilde{\Sigma}$. This isotopy can be realized by an ambient isotopy of $(\widetilde{M},\widetilde{\Sigma})$. Restricting this ambient isotopy to $\widetilde{S}'_k$, we obtain an isotopy of $\widetilde{S}'_k$ to a surface $\widetilde{S}^*_k$. Then $\widetilde{S}_k$ and $\widetilde{S}^*_k$ lie on opposite sides of $\widetilde{S}_j$, hence are disjoint except for their common boundary circle when they are disks. Since they are isotopic, both being isotopic to $\widetilde{S}'_k$, there is a product region between them. This product contains $\widetilde{S}_j$ since $\widetilde{S}_k$ and $\widetilde{S}^*_k$ lie on opposite sides of $\widetilde{S}_j$. Thus if $\widetilde{S}_k$ and $\widetilde{S}^*_k$ are spheres, so is S_j, and if they are disks, so is S_j. There is a unique isotopy class of disks or nontrivial spheres in the product between $\widetilde{S}_k$ and $\widetilde{S}^*_k$, so there is also a product region between $\widetilde{S}_j$ and $\widetilde{S}_k$. If $\widetilde{S}_j$ and $\widetilde{S}_k$ project to different components of S, these two components would be homotopic and hence isotopic, contrary to the definition of a system of disks and spheres. So $\widetilde{S}_j$ and $\widetilde{S}_k$ must be lifts of the same component of S. This component must be a sphere, otherwise the isotopy between $\widetilde{S}_j$ and $\widetilde{S}_k$ would give $\partial\widetilde{S}_j=\partial\widetilde{S}_k$ and hence $\widetilde{S}_j=\widetilde{S}_k$, which is impossible since $c_p\neq c_q$.

Lemma 2.3 *If a 3–manifold M contains a nontrivial sphere S having two lifts to the universal cover $\widetilde{M}$ that bound a product $S^2\times I$, then M is either an S^2–bundle over S^1 or M has a connected summand that is a real projective 3–space split off by S.*

Proof Let S_1 and S_2 be two lifts of S bounding a product $S^2 \times I$ in $\widetilde{M}$. By rechoosing S_2 if necessary, we may assume no other lifts of S lie in this product. If the deck transformation taking S_1 to S_2 takes the product to itself, it is a free involution on this product, yielding a projective space summand of M bounded by S by a classical theorem of Livesay [16]. If the deck transformation takes the product $S^2 \times I$ outside itself, then M is obtained from this product by identifying its two ends, so M is an S^2-bundle. $\square$

Now we finish the proof of (II). If M is an S^2-bundle, it contains a unique isotopy class of nontrivial spheres, and if S and S' were in normal form with respect to Σ they would have to be disjoint from Σ, making (II) true vacuously. So we may assume M is not an S^2-bundle. In the case of a projective space summand bounded by S_j, this summand lifts to the region between $\widetilde{S}_j$ and $\widetilde{S}_k$. Applying the same argument with the roles of S and S' reversed, we find a projective space summand of M bounded by S'_j lifting to the region between $\widetilde{S}'_j$ and $\widetilde{S}'_k$. After extending the isotopy taking S'_j to S_j to an isotopy of M, the projective space summand bounded by S'_j becomes a second projective space summand bounded by S_j, and this lies on the opposite side of S_j from the first summand since $\widetilde{S}_k$ and $\widetilde{S}'_k$ lie on opposite sides of $\widetilde{S}_j$ and $\widetilde{S}'_j$, where we are using the transverse orientations to distinguish sides. Thus S_j has a projective space summand on each side, so M is the connected sum of two projective spaces. It is an elementary fact that in a manifold such as this which is the sum of two irreducible manifolds there is only one isotopy class of nontrivial spheres, just as for S^2-bundles, and (II) would again be true vacuously. (If one wants to avoid quoting Livesay's theorem, one can say that S_j splits M as the sum of two closed 3-manifolds with fundamental group of order 2, and these manifolds are irreducible since M has no fake 3-sphere summands.)

Conclusion of the proof of Theorem 2.1 It is easy to see that a system of transversely oriented circles in a sphere is determined up to isotopy by its associated side relation. In view of (II) this means that by an isotopy of S staying transverse to Σ, the transversely oriented circles of $S \cap \Sigma$ can be made to agree with the corresponding transversely oriented circles of $S' \cap \Sigma$. By the earlier arguments there are then isotopies of the pieces of $\widetilde{S}$ to the corresponding pieces of $\widetilde{S}'$, preserving transverse orientations, and we can take these isotopies to fix the boundaries of these pieces since the space of transversely oriented circles in a 2-sphere is simply-connected, as noted earlier. Thus these isotopies fit together to give an isotopy of $\widetilde{S}$ to $\widetilde{S}'$ fixing $\widetilde{S} \cap \widetilde{\Sigma}$. This isotopy projects to an equivalence of S and S' in M. $\square$

3 Complexes of disks and spheres

We continue with the notation of the previous section, so M is a connected sum of n copies of $S^1 \times S^2$ with s copies of the ball D^3 and k additional irreducible manifolds, and C is a collection of circles in the sphere and torus boundary components of M subject to the same restrictions as before. Let $\mathcal{DS}(M,C)$ be the simplicial complex whose p–simplices are the isotopy classes of systems of $p+1$ disks and spheres in M. The faces of such a simplex are obtained by passing to subsystems. If we restrict attention to systems formed just of disks or just of spheres we obtain subcomplexes $\mathcal{D}(M,C)$ and $\mathcal{S}(M,C)$, respectively. We write $\mathcal{S}(M)$ for $\mathcal{S}(M,\varnothing)$. Eventually we will be most interested in the subcomplex $\mathcal{D}_c(M,C)$ of $\mathcal{D}(M,C)$ formed by disk systems with connected complement. The main result we are heading for in this section is

Theorem 3.1 *$\mathcal{D}_c(M,C)$ is $(n-2)$–connected if $C \neq \varnothing$, and $(n-1)$–connected if C is contained in torus components of ∂M and $C \neq \varnothing$.*

The plan of the proof is to show:

(1) $\mathcal{DS}(M,C)$ is contractible except when $n=0$ and $k \leq 1$, in which cases it is $(2k+s-5)$–connected if $C=\varnothing$ and $(k+s-4)$–connected if $C \neq \varnothing$.

(2) $\mathcal{D}(M,C)$ is contractible if C is contained in torus components of ∂M and $C \neq \varnothing$.

(3) $\mathcal{D}(M,C)$ is $(2n+k+s-4)$–connected in general, if $C \neq \varnothing$.

(4) $\mathcal{D}_c(M,C)$ is $(n-2)$–connected if $C \neq \varnothing$.

(5) $\mathcal{D}_c(M,C)$ is $(n-1)$–connected if C is contained in torus components of ∂M, $C \neq \varnothing$.

Our convention is that (-1)–connected means nonempty and (-2)–connected is a vacuous condition. The empty set is (-2)–connected, for example.

The proof of (3) is the most delicate so we will postpone this until last.

Proof of (1) There are three main steps. The first is the following, where t denotes the number of boundary spheres of M that contain circles of C:

(a) $\mathcal{DS}(M,C)$ is contractible if $\mathcal{S}(M) \neq \varnothing$ and either $s \leq 1$ or $s = t$.

This is proved using the surgery technique of the proof of Theorem 2.1 of [7], with a few minor simplifications. Choose a maximal sphere system Σ and put an arbitrary disk-sphere system $S = S_0 \cup \cdots \cup S_p$ in normal form with respect to Σ. A point in the p-simplex of $\mathcal{DS}(M, C)$ determined by S can be thought of as a weighted sum $\sum_i t_i S_i$ of the components of S, with weights t_i the barycentric coordinates of the point in the simplex. Barycentric coordinates are normalized to have $\sum_i t_i = 1$. To make the surgery process clearer it is helpful to replace each S_i by a family $S \times [0, t_i]$ of parallel copies of S_i of 'thickness' t_i. When a weight t_i goes to zero at a face of the simplex, the family $S_i \times [0, t_i]$ shrinks to thickness zero and is deleted. We will allow the family $S_i \times [0, t_i]$ to be split into several parallel families of total thickness t_i, as well as the inverse operation of combining parallel families into a single family, adding their weights.

Our aim is to construct a sequence of surgeries on S to eliminate the circles of $S \cap \Sigma_0$ for Σ_0 one of the spheres of Σ. Let T_0 be the dual tree of $S \cap \Sigma_0$ in Σ_0, with a vertex for each components of $\Sigma_0 - S$ and an edge for each circle of $S \cap \Sigma_0$. The weights on the S_i's define lengths for these edges, so T_0 is a metric tree. There is a canonical way to shrink T_0 to a point by shortening all its extremal edges simultaneously at unit speed. Once an extremal edge has disappeared one continues shrinking all remaining extremal edges. The surgery process will realize this shrinking of T_0 by a path in $\mathcal{DS}(M, C)$ starting at S and ending with a system disjoint from Σ_0. If T_0 is not already a point, its extremal vertices v correspond to disjoint disks D_v in Σ_0 with $D_v \cap S = \partial D_v$. These disks can be used to surger S to a new system in which the circles ∂D_v have been eliminated from $S \cap \Sigma_0$. Taking weights into account, one gradually surgers through the appropriate families $S_i \times [0, t_i]$ at unit speed, decreasing the thicknesses of these families while increasing the thicknesses of the new families created by the surgery. The old and new families can be taken to be disjoint, so the surgery can be viewed as simply transferring weights from the old family to the new. When the thickness of a family has shrunk to zero, this family is deleted and one continues the surgery process on the remaining families. Near the end of the process of shrinking T_0 to a point it can happen that all that remains of T_0 is a single edge, and then both ends of this edge are shrinking at the same time, so the corresponding family of surfaces is being surgered on both sides simultaneously.

Thus the surgery process produces a family $S(t_1, \cdots, t_p, u)$ of weighted collections of surfaces, for $(t, u) \in \Delta^p \times [0, \infty)$, with $S(t_1, \cdots, t_p, 0)$ the given family. After the process is completed we combine parallel families and discard any trivial spheres or disks that are produced by the surgeries. We need to make sure that some nontrivial disks or spheres always remain. Surgery on a nontrivial sphere S_0 produces a pair of spheres, and the only way these can both be trivial is if S_0 splits off a pair of boundary spheres from M, neither of which meets C. This cannot happen if $s \le 1$ or $s = t$. Surgery on a nontrivial disk D_0 produces a sphere and a disk, and the only way these can both be trivial is if ∂D_0 lies in a sphere of ∂M and D_0 splits off another boundary sphere of M that is disjoint from C. Again we are safe if $s \le 1$ or $s = t$.

Thus after discarding trivial spheres and disks we are left with systems of spheres and disks. Then renormalizing the weights of the remaining nontrivial spheres and disks to have sum 1, the surgery process defines a path in $\mathcal{DS}(M, C)$. This is independent of the choice of normal form system S within its isotopy class because the surgeries always take equivalent systems to equivalent systems, and equivalent systems are isotopic. The surgeries may destroy normal form, but that does not matter.

It is evident that $S(t_1, \cdots, t_p, u)$, regarded as a point in $\mathcal{DS}(M, C)$ by discarding trivial components and renormalizing weights, depends continuously on the parameters t_i and u. When we pass to a face of the parameter simplex Δ^p by letting some t_i go to zero, we get the corresponding family $S(t_1, \cdots, \widehat{t_i}, \cdots, t_p, u)$ for the system obtained by deleting S_i from S. So we have constructed a deformation retraction of $\mathcal{DS}(M, C)$ into the subspace of systems disjoint from Σ_0. This subspace is contractible since it is the star of the vertex corresponding to Σ_0. Hence $\mathcal{DS}(M, C)$ is contractible.

Note that instead of surgering along just the sphere Σ_0 we could just as well surger along any subsystem of Σ, and this would produce a deformation retraction onto the star of this subsystem in $\mathcal{DS}(M, C)$.

This takes care of (a). The other two steps in the proof of (1) will provide a reduction to (a). Suppose M has boundary spheres $\partial_1, \cdots, \partial_s$, one of which, say ∂_s, is disjoint from C. Let M' be the manifold obtained from M by filling in ∂_s with a ball.

(b) If $\mathcal{S}(M) \ne \varnothing$ and $s \ge t + 2$, the connectivity of $\mathcal{DS}(M, C)$ is one greater than the connectivity of $\mathcal{DS}(M', C)$.

The proof of this is similar to the proof of Lemma 2.2 in [7]. Call a vertex of $\mathcal{DS}(M, C)$ *special* if it is either

- a sphere splitting off a 3-punctured sphere from M whose other two boundary spheres are ∂_s and some ∂_i which is disjoint from C, or

- a disk splitting off a 2–punctured sphere from M having ∂_s as one boundary sphere and whose other boundary sphere intersects C only in the boundary circle of the disk.

Special vertex spheres exist if $s \geq t+2$, so let Σ be such a sphere, splitting off a submanifold P of M that is a 3–punctured sphere bounded by Σ, ∂_s, and some other ∂_i. Let $\mathcal{DS}'(M,C)$ be the subcomplex of $\mathcal{DS}(M,C)$ consisting of simplices with no special vertices. A deformation retraction of $\mathcal{DS}'(M,C)$ onto the link of Σ can be obtained in the following way. Take a system S defining a simplex in $\mathcal{DS}'(M,C)$ and put it in normal form with respect to some maximal sphere system containing Σ. Then if $S \cap P$ is not empty, it consists of parallel disks separating ∂_s from ∂_i. These disks can be eliminated by pushing them across ∂_s and then outside P. Since no components of S are special disks or spheres, this process of modifying S by pushing across ∂_s produces no trivial disks or spheres. The process could also be described in terms of surgering S along Σ and discarding the resulting spheres that are isotopic to ∂_s, so the process determines a deformation retraction of $\mathcal{DS}'(M,C)$ onto the link of Σ, as desired.

We obtain $\mathcal{DS}(M,C)$ from $\mathcal{DS}'(M,C)$ by attaching the star of each special vertex along the link of this vertex. These stars are the cones on the links, and have disjoint interiors since no simplex can contain two distinct special vertices. The star of Σ is contractible, so by the preceding paragraph the union of this star with $\mathcal{DS}'(M,C)$ is also contractible. We obtain $\mathcal{DS}(M,C)$ from this contractible space by attaching the stars of the other special vertices, so $\mathcal{DS}(M,C)$ is homotopy equivalent to the wedge of the suspensions of the other links. Each link is a copy of $\mathcal{DS}(M',C)$, so the connectivity of $\mathcal{DS}(M,C)$ is one greater than the connectivity of $\mathcal{DS}(M',C)$.

(c) The connectivity of $\mathcal{DS}(M,C)$ is one greater than the connectivity of $\mathcal{DS}(M',C)$ if $s > t$ and $t > 0$. The same is true for $\mathcal{D}(M,C)$ and $\mathcal{D}(M',C)$.

The proof parallels the previous one. First we have a preliminary observation:

- If C is obtained from C' by adding a second circle in some sphere component of ∂M then $\mathcal{DS}(M,C)$ is homotopy equivalent to the suspension of $\mathcal{DS}(M,C')$, and this is also true for $\mathcal{D}(M,C)$ and $\mathcal{D}(M,C')$.

To see this, let C_0 and C_1 be two circles of C in the same boundary sphere of M and let these bound disks D_0 and D_1 that lie in a neighborhood of the boundary sphere and represent vertices of $\mathcal{D}(M,C)$. Let $\mathcal{DS}'(M,C)$ be the subcomplex of $\mathcal{DS}(M,C)$ consisting of simplices containing neither of the vertices represented by D_0 and D_1. Then $\mathcal{DS}'(M,C)$ deformation retracts onto the link of D_1 in $\mathcal{DS}(M,C)$ by shifting disks with boundary on C_0 over to disks with boundary on C_1, using the annulus between C_0 and C_1 in ∂M. In the same way $\mathcal{DS}'(M,C)$ deformation retracts onto the link of D_0. Since $\mathcal{DS}(M,C)$ is the union of $\mathcal{DS}'(M,C)$ with the stars of the vertices D_0 and D_1, it follows that $\mathcal{DS}(M,C)$ is homotopy equivalent to the suspension of the link of either of these vertices. These links can be identified with $\mathcal{DS}(M,C')$. The same argument applies with $\mathcal{D}$ in place of $\mathcal{DS}$.

Now to prove (c) we may assume C has at most one circle in each sphere of ∂M. Let D_0 be a disk representing a special vertex of $\mathcal{D}(M,C)$, with ∂D_0 in ∂_i, and let Σ be a sphere splitting off a 3-punctured sphere P containing D_0 and having ∂_s and ∂_i as its other two boundary spheres. As in (b), let $\mathcal{DS}'(M,C)$ be the subcomplex of $\mathcal{DS}(M,C)$ consisting of simplices having no special vertices. We may deformation retract $\mathcal{DS}'(M,C)$ onto the link of D_0 by putting systems S representing simplices of $\mathcal{DS}'(M,C)$ into normal form with respect to a maximal sphere system containing Σ and then eliminating the intersections of S with an arc in P joining ∂_i to ∂_s by pushing across ∂_s. The intersections of S with P can be disks with boundary on Σ, cylinders joining Σ to ∂_i, or spheres parallel to ∂_i. Figure 5 shows which combinations of these surfaces are possible (along with parallel copies of these surfaces).

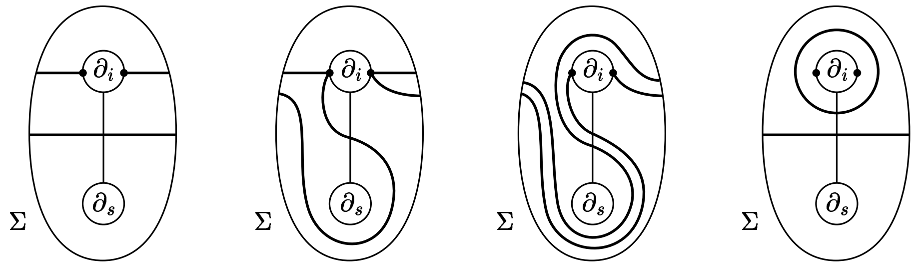

Figure 5

The rest of the argument then follows the reasoning in (b). The proof works equally well with $\mathcal{D}$ in place of $\mathcal{DS}$.

To deduce (1) from (a), (b), and (c) there are three cases:

(i) If $n > 0$ or $k > 1$ then $\mathcal{S}(M) \neq \varnothing$, so (b) gives the reduction to (a) when $t = 0$ and (c) gives the reduction to (a) when $t > 0$. The conclusion in both cases is that $\mathcal{DS}(M,C)$ is contractible.

(ii) Suppose $n = k = 0$. If $t = 0$ then $\mathcal{DS}(M,C) = \mathcal{S}(M) \neq \varnothing$ for $s \geq 4$ so (b) gives the reduction to the case $s = 4$ and we conclude that $\mathcal{DS}(M,C)$ is $(s-5)$–connected. If $t = 1$ then $\mathcal{S}(M) \neq \varnothing$ for $s \geq 3$ so (b) gives a reduction to the case $s = 3$ and $\mathcal{DS}(M,C)$ is $(s-4)$–connected. If $t \geq 2$ then $\mathcal{S}(M) \neq \varnothing$ so (c) gives a reduction to the case $s = t$ and we see that $\mathcal{DS}(M,C)$ is contractible.

(iii) Suppose $n = 0$ and $k = 1$. Then $\mathcal{S}(M) \neq \varnothing$ for $s \geq 2$. Reasoning as in (ii) we see that $\mathcal{DS}(M,C)$ is contractible if $t \geq 2$ and $(s-3)$–connected if $t \leq 1$. Note that $s - 3 = 2k + s - 5 = k + s - 4$ when $k = 1$.

This finishes the proof of (1). ◻

Proof of (2) Let Σ be a system of spheres that splits off all the $S^1 \times D^2$ summands of M whose boundary tori contain circles of C. Doing surgery along this Σ as in the first step of the proof of (1) and discarding all spheres produced by the surgeries gives a deformation retraction of $\mathcal{D}(M,C)$ into the subcomplex consisting of disk systems disjoint from Σ. This subcomplex is the join of copies of the disk complex for $M^1_{0,1}$, which is homeomorphic to the real line. Namely, up to diffeomorphism of $M^1_{0,1}$ fixing the boundary torus there is a unique nontrivial disk, and applying Dehn twists on the boundary torus gives an infinite sequence of disks, the vertices of the disk complex. There is an edge joining each vertex to each of its two neighbors in the infinite sequence, and there are no simplices of higher dimension, so the disk complex for $M^1_{0,1}$ is a line. In particular it is contractible, so $\mathcal{D}(M,C)$ is contractible as well.

This argument does not cover the special case $M = M^0_{0,1}$, but in this case $\mathcal{D}(M,C)$ is just a point. ◻

Proof that (3) implies (4) For $n = 0$ the result holds vacuously, and for $n = 1$ it is true since $\mathcal{D}_c(M,C)$ is nonempty if $n > 0$ and $C \neq \varnothing$. So we may assume $n \geq 2$. To show that $\mathcal{D}_c(M,C)$ is $(n-2)$–connected, start with a map $f\colon S^m \to \mathcal{D}_c(M,C)$, $m \leq n-2$. By (3) this extends to a map $F\colon D^{m+1} \to \mathcal{D}(M,C)$ since $2n + k + s - 4 \geq n - 2$ if $n \geq 2$. We may assume F is simplicial with respect to some triangulation of D^{m+1}. Each p–simplex σ of D^{m+1} then maps to a disk system $D = D_0 \cup \cdots \cup D_q$ with $q \leq p$. Choose σ of maximal dimension p such that the system D is purely separating, meaning that each D_i separates the complement of the other disks. This is equivalent to saying that all edges in the dual graph of D have distinct endpoints. This implies that σ is not contained in ∂D^{m+1} since F maps this to systems with connected complement, having dual graph a wedge of circles. Thus the link

L_σ of σ in the triangulation of D^{m+1} is a sphere S^{m-p}. The maximality of σ implies that any simplex properly containing σ is sent by F to a disk system whose dual graph is obtained from the dual graph of D by attaching loops at some vertices. This implies that F maps L_σ to the join

$$J_\sigma = \mathcal{D}_c(M_1, C_1) * \cdots * \mathcal{D}_c(M_d, C_d)$$

where the M_i's are the components of the manifold obtained by splitting M along D and the C_i's are the preimages of C in these components. Each (M_i, C_i) is smaller than (M, C) with respect to the lexicographic ordering on the 4-tuples (n, k, s, u) where u denotes the number of circles in C. So by induction $\mathcal{D}_c(M_i, C_i)$ is $(n_i - 2)$-connected. Since the quantity 'connectivity-plus-two' is additive for joins by [19], it follows that J_σ is $\big(\sum_i n_i - 2\big)$-connected. We have $\sum_i n_i \geq n - q \geq n - p$ since the system D of $q+1$ disks separates M, so splitting M along D can decrease the rank of the dual graph by at most q. Thus J_σ is at least $(n - p - 2)$-connected. Since L_σ is a sphere of dimension $m - p$ we deduce that if $m \leq n - 2$, the map $F\colon L_\sigma \to J_\sigma$ extends to a map $G\colon D^{m-p+1} \to J_\sigma$. The star of σ in D^{m+1} is the join $\sigma * L_\sigma$ but it can also be regarded as $\partial\sigma * D^{m-p+1}$ for a disk $D^{m-p+1} \subset D^{m+1}$ with $\partial D^{m-p+1} = L_\sigma$. Thus we can redefine F on the interior of the star of σ by taking the join of the restriction of F to $\partial\sigma$ with G. For the new F there are no purely separating simplices of dimension p or greater in the star. Hence after a finite number of modifications like this we eliminate all purely separating simplices, thereby making F into a map to $D_c(M, C)$, finishing the proof that $D_c(M, C)$ is $(n - 2)$-connected. □

Proof that (4) implies (5) This is very similar to the preceding proof. The case $n = 0$ is easy since $\mathcal{D}_c(M, C)$ is nonempty, so we may assume $n \geq 1$. We start with a map $S^m \to \mathcal{D}_c(M, C)$ and extend it to a map $D^{m+1} \to \mathcal{D}(M, C)$ using (2). As before, we look at a maximal simplex σ mapping to a purely separating disk system D. The inequality $\sum_i n_i \geq n - p$ now becomes $\sum_i n_i \geq n - p + 1$ since splitting along the first disk in a disk system just changes a boundary torus into a sphere. The join J_σ is $\big(\sum_i n_i - 2\big)$-connected using (4) instead of induction. Then if $m \leq n - 1$ the connectivity of J_σ is at least $\sum_i n_i - 2 \geq n - p + 1 - 2 \geq m - p$ so we can finish the proof as before. □

Proof of (3) It is natural to try the same approach as in the two previous proofs. Starting with a map $f\colon S^m \to \mathcal{D}(M, C)$ we can extend it to a map $F\colon D^{m+1} \to \mathcal{DS}(M, C)$ using (1), and we may assume this is a simplicial map. We wish to modify F so as to eliminate all sphere systems in its image, so one's first guess would be to take σ to be a maximal dimension simplex in

D^{m+1} such that $F(\sigma)$ is a sphere system. This does not work, however. Here is an example of what can go wrong. Suppose we take $m=0$ and we let C consist of two circles in one sphere of ∂M. Then there are nontrivial disks D_0 and D_1 lying in a neighborhood of this boundary sphere and having boundaries on the two circles of C. These two disks correspond to two distinct vertices of $\mathcal{D}(M,C)$, and the easiest way to join these two vertices by a path in $\mathcal{DS}(M,C)$ is to interpolate the vertex formed by a sphere S parallel to this boundary component of M, so there are edges from D_0 to S and from S to D_1. Then σ would be this vertex S, but there is no way to modify F on the star of σ by connecting D_0 and D_1 by a path of disk systems disjoint from S since the only such disk systems are D_0 and D_1.

So a more subtle approach is needed. Fortunately a model for what to do in an analogous situation in one lower dimension is available in Theorem 2.7 of [13], and we will follow this model closely.

Recall that t is the number of boundary spheres of M intersecting C. Note that we can assume that $t>0$ by (2), and then that $t=s$ by (1)(c).

For a p–simplex σ of D^{m+1} mapped by F to a sphere system S let $X=X_\sigma$ be the closure of the union of the components of $M-S$ that meet C, and let $Y=Y_\sigma$ be the closure of $M-X$. Call σ *regular* if no spheres of S lie in the interior of Y. Every simplex σ of D^{m+1} mapping to a sphere system contains a subsimplex which is regular, obtained by discarding vertices mapping to spheres in the interior of Y. Such vertices are called *superfluous*. We will assume from now on that σ is regular. Let $Y_1,\cdots,Y_c$ be the components of Y and let $X_1,\cdots,X_d$ be the components of the manifold obtained from X by splitting along the spheres of S.

Among the various regular simplices of D^{m+1} suppose that σ is chosen so that the lexicographically ordered pair $(c(Y_\sigma),p)$ is maximal, where the complexity $c(Y_\sigma)$ is defined as the number of components in the complement of a maximal sphere system in Y_σ. (This does not depend on the maximal sphere system chosen, by the unique prime decomposition theorem for 3–manifolds.) With this maximality assumption on σ we claim that F takes the link L_σ of σ in D^{m+1} to the join

$$J_\sigma=\mathcal{D}(X_1,C_1)*\cdots*\mathcal{D}(X_d,C_d)*\mathcal{S}(Y_1)*\cdots*\mathcal{S}(Y_c)$$

where C_i is the part of C in X_i. Indeed, let v be a vertex in L_σ, so $F(v)$ is a disk or sphere in the complement of the system $S=F(\sigma)$. If $F(\sigma)$ lies in some Y_j it must be a sphere defining a vertex of $\mathcal{S}(Y_j)$ and hence in J_σ. The other possibility is that $F(\sigma)$ lies in some X_i, where it could be a disk or a sphere, and we need to show it must be a disk. So suppose it is a sphere. If it does not split off a piece of X_σ disjoint from C then we can enlarge σ to include v without changing X_σ or Y_σ, contradicting the maximality of $(c(Y_\sigma), p)$. On the other hand, if $F(\sigma)$ does split off a piece of X_σ disjoint from C then by enlarging σ to include v and then discarding any resulting superfluous vertices we obtain a new σ' with $Y_{\sigma'}$ the union of Y_σ and the piece split off from X_σ, so $c(Y_{\sigma'}) > c(Y_\sigma)$, contradicting maximality again.

So we have $F\colon L_\sigma \to J_\sigma$. We calculate now the connectivity J_σ. If any Y_j has $n(Y_j) > 0$ or $k(Y_j) > 1$ then $\mathcal{S}(Y_j)$ is contractible by (1), hence J_σ is also contractible. So we may assume $n(Y_j) = 0$ and $k(Y_j) \leq 1$ for each j. For the connectivity of the join J_σ we have

$$\operatorname{conn}(J_\sigma) = \textstyle\sum_i \operatorname{conn}\big(\mathcal{D}(X_i, C_i)\big) + \sum_j \operatorname{conn}\big(\mathcal{S}(Y_j)\big) + 2c + 2d - 2$$

By induction on the lexicographically ordered triple (n, k, s) we know the connectivities in the first summation, and (1) gives the connectivities in the second summation, so if the number of Y_j's with $k(Y_j) = 1$ is a, we get

$$\begin{aligned}\operatorname{conn}(J_\sigma) = 2\textstyle\sum_i n(X_i) + \sum_i k(X_i) + \sum_i s(X_i) - 4d \\ + \textstyle\sum_j s(Y_j) - 5c + 2a + 2c + 2d - 2\end{aligned}$$

We can start the induction with $(n, k, s) = (0, 0, 3)$ when $\mathcal{D}(M, C)$ is nonempty if $C \neq \varnothing$. If $F(\sigma)$ is a q–simplex, consisting of $q+1$ spheres, then by splitting M along these spheres and counting the number of boundary spheres in the resulting manifold we see that $s(M) + 2q + 2 = \sum_i s(X_i) + \sum_j s(Y_j)$. Looking at the dual graph to the system $F(\sigma)$ we see that $n(M) = \sum_i n(X_i) + q + 1 - (c + d - 1)$ since it takes $c + d - 1$ edges to connect together the $c + d$ subgraphs corresponding to the X_i's and Y_j's and we assumed $n(Y_j) = 0$. Also we have $k(M) = \sum_i k(X_i) + a$. Substituting into the preceding displayed formula and letting $n = n(M)$, $k = k(M)$, $s = s(M)$, we see that the connectivity of J_σ is

$$\begin{aligned}(2n - 2q - 4 + 2c + 2d) + (k - a) - 4d + (s + 2q + 2) - 5c + 2a + 2c + 2d - 2 \\ = 2n + k + s - 4 + a - c\end{aligned}$$

Note that $q+1 \geq 3(c-a)$ since each Y_j with $k(Y_j) = 0$ must have at least three boundary spheres and we assumed that no boundary sphere of M was disjoint from C. This implies that $q \geq c - a$. If we assume that $m \leq 2n + k + s - 4$, then

$$\begin{aligned}\dim L_\sigma = m - p \leq m - q \leq 2n + k + s - 4 - q \\ \leq 2n + k + s - 4 + a - c = \operatorname{conn}(J_\sigma)\end{aligned}$$

Hence if $m \leq 2n + k + s - 4$ we can modify F in the star of σ as before, taking the join of its values on $\partial\sigma$ with an extension of $F\colon L_\sigma \to J_\sigma$ to a map $D^{m-p+1} \to J_\sigma$.

It remains to check that this modification improves the situation, namely that the maximum value of $(c(Y_\sigma), \dim\sigma)$ has not increased, while the number of simplices σ realizing this maximum value has actually decreased. We only need to check simplices τ in $D^{m-p+1} * \partial\sigma$ that are sent by the new F' to sphere systems, and we need only consider such τ that are regular. The simplex τ is the join of a simplex α in D^{m-p+1} and a simplex β in $\partial\sigma$, where α or β is taken to be empty when τ lies in $\partial\sigma$ or D^{m-p+1}. Since $F'(\tau)$ is a sphere system, so are $F'(\alpha)$ and $F'(\beta)$. By construction $F'(\alpha)$ is a simplex in J_σ, so since it is a sphere system it must lie in $\mathcal{S}(Y_1) * \cdots * \mathcal{S}(Y_c)$ and hence it must consist of spheres in Y_σ. On the other hand, $F'(\beta)$ is a subsimplex of $F(\sigma)$. These two facts imply that $Y_\tau \subset Y_\sigma$. Hence $c(Y_\tau) \leq c(Y_\sigma)$. Equality can occur only if $Y_\tau = Y_\sigma$, in which case $\alpha = \varnothing$ since τ is regular. Then the simplex $\tau = \beta$ is a proper face of σ and hence has smaller dimension. So in all cases $(c(Y_\tau), \dim\tau) < (c(Y_\sigma), \dim\sigma)$. □

This finishes the proof of Theorem 3.1. We will also need a similar result with spheres in place of disks. Let $\mathcal{S}_c(M)$ be the subcomplex of $\mathcal{S}(M)$ consisting of sphere systems with connected complement and let $\mathcal{S}_c^{\pm}(M)$ be the complex of oriented sphere systems with connected complement. The simplices of $\mathcal{S}_c^{\pm}(M)$ thus correspond to systems of spheres each of which has a chosen orientation, or equivalently, a chosen normal orientation. (There is no need to consider oriented disk complexes since disks can be canonically oriented by choosing orientations for the circles of C.)

Proposition 3.2 *$\mathcal{S}_c(M)$ and $\mathcal{S}_c^{\pm}(M)$ are $(n-2)$-connected.*

This was shown in Proposition 3.1 of [7] in the case $k = 0$. The generalization to $k > 0$ is straightforward, and we give the proof here mainly to correct a small error at the beginning of the proof in [7].

Proof For the case of $\mathcal{S}_c(M)$ this will follow from (1) by the same argument used to show that (3) implies (4). The result is trivially true for $n = 0$, so we may assume that $n > 0$. Consider a map $f\colon S^m \to \mathcal{S}_c(M)$ with $m \leq n-2$. Since $\mathcal{S}(M)$ is contractible, this extends to a map $F\colon D^{m+1} \to \mathcal{S}(M)$. Let σ be a purely separating simplex in D^{m+1} of maximal dimension p with image the sphere system $S = S_0 \cup \cdots \cup S_q$ where $q \leq p$. Then F maps the link $L_\sigma = S^{m-p}$

to $J_\sigma = \mathcal{S}_c(M_1) * \cdots * \mathcal{S}_c(M_d)$ where the M_i's are the manifolds obtained by cutting M along S. By induction on the triple (n,k,s), the connectivity of $\mathcal{S}_c(M_i)$ is $n_i - 2$ and thus J_σ is $(\sum_i n_i - 2)$-connected. Now $\sum_i n_i \geq n - q$ as the system S separates M. So $\sum_i n_i - 2 \geq n - q - 2 \geq m - p$ as $m \leq n-2$ and $q \leq p$. Hence $F\colon L_\sigma = S^{m-p} \to J_\sigma$ can be extended to the disc D^{m-p+1}. We use this to modify F on the interior of the star of σ. After a finite number of such modifications, F becomes a map with image in $\mathcal{S}_c(M)$, so $\mathcal{S}_c(M)$ is $(n-2)$-connected.

We want to deduce that $\mathcal{S}_c^{\pm}(M)$ is also $(n-2)$-connected. For this, choose a positive orientation on the spheres of $\mathcal{S}_c(M)$ so that $\mathcal{S}_c(M)$ is identified with the subcomplex $\mathcal{S}_c^{+}(M)$ of $\mathcal{S}_c^{\pm}(M)$ of spheres with positive orientation. It will suffice to take a map $f\colon S^m \to \mathcal{S}_c^{\pm}(M)$ with $m \leq n-2$ and homotope this to have image in $\mathcal{S}_c^{+}(M)$. Let σ be a simplex of S^m of maximal dimension p whose image under f is a system $S = S_0^- \cup \cdots \cup S_q^-$ consisting entirely of negatively oriented spheres. The link $L_\sigma = S^{m-p-1}$ then maps to $\mathcal{S}_c^{+}(M')$, where M' is M cut along S. The manifold M' has $n-q-1$ copies of $S^1 \times S^2$ in its decomposition as $M - S$ is connected. So $\mathcal{S}_c^{+}(M')$ is $(n-q-3)$-connected. Now $m-p-1 \leq n-q-3$ as $m \leq n-2$ and $q \leq p$. Hence we can modify f on the interior of the star of σ as before. Note that the new f is homotopic to the old one by coning. After a finite number of such modifications we obtain an f having image in $\mathcal{S}_c^{+}(M)$. □

We conclude this section with two results that will be needed for the spectral sequence argument. We use the notation $A(M)$ for the quotient of $\pi_0 \mathrm{Diff}(M)$ by twists along 2-spheres. This quotient acts on all the disk and sphere complexes for M since twists along 2-spheres act trivially, as they obviously preserve the homotopy classes of embedded spheres and disks.

Proposition 3.3 *If C consists of a single circle, then $A(M)$ acts transitively on the p-simplices of $\mathcal{D}_c(M,C)$ for each p. The same is true also for $\mathcal{S}_c(M)$ and $\mathcal{S}_c^{\pm}(M)$.*

Proof First consider the case that C is contained in a torus boundary component. Let $D = D_0 \cup \cdots \cup D_p$ be a disk system representing a p-simplex of $\mathcal{D}_c(M,C)$. Thinking of C as a longitude of the torus T of ∂M that it lies in, take a meridian circle of T and push this slightly into the interior of M. Orienting this meridian circle then determines an ordering of the disks D_i, which we may assume agrees with the ordering by increasing subscripts. If $D' = D_0' \cup \cdots \cup D_p'$ is another such disk system, we want to find a diffeomorphism $M \to M$ fixed on ∂M and taking each D_i to D_i'.

Let N be the manifold obtained from M by splitting along D. Splitting along D_0 replaces the $S^1 \times D^2$ summand of M containing T by a D^3 summand, and then, since $M - D$ is connected, splitting along each subsequent D_i replaces an $S^1 \times S^2$ summand by another D^3 summand. The same is true when we split along D' to produce another manifold N'. By the uniqueness of prime decompositions for compact orientable 3–manifolds, N and N' are diffeomorphic, and the diffeomorphism can be chosen to be the identity on the components of ∂M other than T. The diffeomorphism can also be chosen to permute the new boundary spheres in any way we like, so it can be chosen in such a way that it passes down to a quotient diffeomorphism $M \to M$ taking each D_i to D_i' and fixed on ∂M.

The case that C is contained in a sphere boundary component is covered by the same argument since in this case splitting along disks again replaces $S^1 \times S^2$ summands by D^3 summands. For $\mathcal{S}_c(M)$ and $\mathcal{S}_c^{\pm}(M)$ splitting along spheres replaces $S^1 \times S^2$ summands by pairs of D^3 summands, so a similar argument applies. □

For a system S of disks and spheres in M let $A(M, S)$ be the subgroup of $A(M)$ represented by diffeomorphisms restricting to the identity on S. If M' denotes M split along S, there is thus a surjection $A(M') \to A(M, S)$.

Proposition 3.4 *The surjection $A(M') \to A(M, S)$ is an isomorphism.*

Proof Choose a system Σ in M' consisting of spheres that split off all prime summands together with spheres in $S^1 \times S^2$ summands and disks in $S^1 \times D^2$ summands. Represent an element of the kernel of $A(M') \to A(M, S)$ by a diffeomorphism $f \in \mathrm{Diff}(M)$ that is isotopic to the identity and restricts to the identity on S. The system $f(\Sigma)$ is then isotopic to Σ in M, hence in M' as shown in the proof of Theorem 2.2. So we may assume f takes each component of Σ to itself. We may also assume f is the identity on disk components of Σ since it is the identity on their boundaries. On the spheres in Σ that are separating, f must preserve orientation since it cannot switch the sides of these spheres, and f must also preserve orientations of the nonseparating spheres in Σ since it induces the identity on homology. Thus we may assume f is the identity on Σ as well as on S. Modulo twists along 2–spheres we can then deform f, staying fixed on Σ, to be the identity on all of M except the prime summands that are neither $S^1 \times S^2$ nor $S^1 \times D^2$, since an orientation-preserving diffeomorphism of a punctured S^3 that takes each puncture to itself is isotopic to the identity. For the remaining prime summands we appeal to the theorem in Section 5 of [12] which has as a corollary the fact that the product of the mapping class groups of these prime summands injects into the mapping class group of M. □

4 Proof of stability

We make a small change of notation now and let $M^s_{n,k}$ be the connected sum of n copies of $S^1 \times S^2$, k copies of $S^1 \times D^2$, s balls, and perhaps also another manifold N that is a sum of prime factors satisfying the conditions imposed at the beginning of Section 2. Then we have the following stabilization result for the groups $A^s_{n,k}$, which are the mapping class groups of the manifolds $M^s_{n,k}$ with twists along 2–spheres factored out:

Theorem 4.1 *For all $k \geq 0$ and $s \geq 1$ the natural stabilization maps*

(1) $H_i(A^s_{n,k}) \to H_i(A^{s+1}_{n,k})$

(2) $H_i(A^s_{n,k}) \to H_i(A^s_{n,k+1})$

(3) $H_i(A^s_{n,k}) \to H_i(A^s_{n+1,k})$

are isomorphisms whenever $n \geq 3i + 3$.

These isomorphisms will follow from three other similar isomorphisms. Consider the following three maps, for $s \geq 0$:

- $\alpha\colon A^{s+2}_{n,k} \to A^{s+1}_{n+1,k}$, induced by the map $M^{s+2}_{n,k} \to M^{s+1}_{n+1,k}$ identifying disks in each of the last two boundary spheres of M, or equivalently, attaching a 1–handle $D^1 \times D^2$ joining these two boundary spheres.
- $\beta\colon A^{s+2}_{n,k} \to A^{s}_{n+1,k}$, induced by the map $M^{s+2}_{n,k} \to M^{s}_{n+1,k}$ identifying the two last boundary spheres of M, or equivalently, joining them by a product $S^2 \times I$.
- $\gamma\colon A^{s+1}_{n,k} \to A^{s}_{n,k+1}$, induced by the map $M^{s+1}_{n,k} \to M^{s}_{n,k+1}$ identifying two disks in the last boundary sphere of $M^{s+1}_{n,k}$, or equivalently, attaching a 1–handle $D^1 \times D^2$ from this boundary sphere to itself.

What we will actually prove is that for all $k \geq 0$ and $s \geq 0$ the induced maps

(A) $\alpha_i\colon H_i(A^{s+2}_{n,k}) \to H_i(A^{s+1}_{n+1,k})$

(B) $\beta_i\colon H_i(A^{s+2}_{n,k}) \to H_i(A^{s}_{n+1,k})$

(C) $\gamma_i\colon H_i(A^{s+1}_{n,k}) \to H_i(A^{s}_{n,k+1})$

are surjective when $n \geq 3i$ and isomorphisms when $n \geq 3i+2$. To see how this implies the theorem, we use two other maps:

- $\delta\colon A^{s+1}_{n,k} \to A^s_{n,k}$, induced by filling in the last boundary sphere with a ball.
- $\varepsilon\colon A^s_{n,k} \to A^{s+1}_{n,k}$, induced by attaching a 2–handle $D^2 \times D^1$ to the last boundary sphere, if $s \geq 1$.

Since β is the composition $\delta\alpha$, we deduce that $\delta_i\colon H_i(A^{s+1}_{n,k}) \to H_i(A^s_{n,k})$ is an isomorphism for $n \geq 3i+3$. The composition $\delta\varepsilon$ is the identity, so we conclude that ε induces an isomorphism on homology:

(1) The stabilization $H_i(A^s_{n,k}) \to H_i(A^{s+1}_{n,k})$ is an isomorphism for $n \geq 3i+3$ and $s \geq 1$.

The composition $\gamma\varepsilon$ comes from attaching disjoint 1– and 2–handles to the last boundary sphere, so this is the standard stabilization $A^s_{n,k} \to A^s_{n,k+1}$. Thus we have:

(2) The stabilization $H_i(A^s_{n,k}) \to H_i(A^s_{n,k+1})$ is an isomorphism for $n \geq 3i+3$ and $s \geq 1$.

Finally, the standard stabilization $A^s_{n,k} \to A^s_{n+1,k}$ is $\beta\varepsilon^2$, so:

(3) The stabilization $H_i(A^s_{n,k}) \to H_i(A^s_{n+1,k})$ is an isomorphism for $n \geq 3i+3$ and $s \geq 1$.

Corollary 4.2 *The quotient map* $\mathrm{Aut}(F_n) \to \mathrm{Out}(F_n)$ *induces an isomorphism on* H_i *for* $n \geq 3i+3$.

Proof The map $\mathrm{Aut}(F_n) \to \mathrm{Out}(F_n)$ is a special case of δ as $A^1_{n,0} = \mathrm{Aut}(F_n)$ and $A^0_{n,0} = \mathrm{Out}(F_n)$, and we have just seen that δ induces an isomorphism on H_i in the stated range. □

Case (A) We do an induction on i, using the relative spectral sequence argument of [22]. The statement is clearly true for $i=0$. The induction hypothesis will be that the following two statements are true for all $q < i$:

(a_q) $\alpha_q\colon H_q(A^{s+2}_{n,k}) \to H_q(A^{s+1}_{n+1,k})$ is surjective for all $n \geq 3q$.

(b_q) $\alpha_q\colon H_q(A^{s+2}_{n,k}) \to H_q(A^{s+1}_{n+1,k})$ is an isomorphism for all $n \geq 3q+2$.

We want to show that this implies (a_i) and (b_i).

To ease the notation we write the stabilization $A^{s+2}_{n,k} \to A^{s+1}_{n+1,k}$ as $G_n \to G_{n+1}$. Consider the action of G_{n+1} on the complex $X_{n+1} = \mathcal{D}_c(M^{s+1}_{n+1,k}, C)$, where C is a single circle in the last sphere component of ∂M. The stabilizer of a vertex is G_n, and more generally the stabilizer of a p–simplex is $G_{n-p} = A^{s+p+2}_{n-p,k}$. If we take X_n to be $\mathcal{D}_c(M^{s+2}_{n,k}, C)$ with C again a single circle in the last boundary sphere, then the vertex stabilizer G_n acts on X_n, and the map $(M^{s+2}_{n,k}, C) \to (M^{s+1}_{n+1,k}, C)$ that induces α also induces an embedding of X_n in X_{n+1} preserving the actions of G_n and G_{n+1}. Thus if (C_*, ∂_*) and (C'_*, ∂'_*) are the augmented chain complexes of X_n and X_{n+1} respectively, then we have a map of double complexes

$$i\colon E_qG_n \otimes_{G_n} C_p \longrightarrow E_qG_{n+1} \otimes_{G_{n+1}} C'_p$$

where (E_*G_n, d_*) is a free G_n–resolution of $\mathbb{Z}$ and similarly for (E_*G_{n+1}, d_*).

With mapping cones in mind, consider now the double complex

$$(E_{q-1}G_n \otimes_{G_n} C_p) \oplus (E_qG_{n+1} \otimes_{G_{n+1}} C'_p)$$

with vertical boundary maps $(a \otimes b, a' \otimes b') \mapsto \big(-da \otimes b, d'a' \otimes b' + i(a \otimes b)\big)$ and horizontal boundary maps $(a \otimes b, a' \otimes b') \mapsto (-1)^q(a \otimes \partial b, a' \otimes \partial' b')$. There are two spectral sequences associated to this double complex. The first spectral sequence, arising from the horizontal filtration, converges to 0 for $p \le n-2$ since $H_p(C_*) = 0$ for $p \le n-2$ and $H_p(C'_*) = 0$ for $p \le n-1$ by Theorem 3.1. So in the second spectral sequence, associated to the vertical filtration, we have $E^\infty_{p,q} = 0$ for $p+q \le n-2$. The columns in the double complex are the mapping cones of the chain maps $i\colon E_*G_n \otimes_{G_n} C_p \to E_*G_{n+1} \otimes_{G_{n+1}} C'_p$. So the E^1 term of the second spectral sequence is

$$E^1_{p,q} = H_q(E_*G_{n+1} \otimes_{G_{n+1}} C'_p, E_*G_n \otimes_{G_n} C_p)$$

By Proposition 3.3 the actions of G_n and G_{n+1} on X_n and X_{n+1} are transitive on the sets of p–simplices. For the map $X_n \to X_{n+1}$, each orbit of a p–simplex in X_{n+1} comes from the orbit of a p–simplex in X_n if $p \le \dim X_n = n-1$. The stabilizers of p–simplices in X_n and X_{n+1} are G_{n-p-1} and G_{n-p} respectively, so Shapiro's Lemma (or its proof) gives an isomorphism

$$E^1_{p,q} \cong H_q(G_{n-p}, G_{n-p-1}) \quad \text{for } p \le n-1$$

where the coefficients are untwisted $\mathbb{Z}$'s since the stabilizers do not permute the vertices of a simplex.

To prove (a_i), we will show that $H_i(G_{n+1}, G_n) = 0$ when $n \ge 3i$. Consider the differential

$$d^1\colon E^1_{0,i} = H_i(G_n, G_{n-1}) \longrightarrow E^1_{-1,i} = H_i(G_{n+1}, G_n)$$

By assumption, $H_q(G_{n-p}, G_{n-p-1}) = 0$ if $q < i$ and $p+q = i$ as $n-p-1 \geq 3i-p-1 \geq 3i-3p = 3q$ since $p \geq 1$. As $n \geq 3i$ also implies $i-1 \leq n-2$ when $i > 0$, the term $E^1_{-1,i}$ must be killed by differentials and hence the differential d^1 above is surjective as the other differentials originate in trivial groups. Now a diagram chase in

$$\begin{array}{ccccccc} H_i(G_n) & \longrightarrow & H_i(G_n, G_{n-1}) & \longrightarrow & H_{i-1}(G_{n-1}) & \xrightarrow{\cong} & H_{i-1}(G_n) \\ \downarrow & & \downarrow{\scriptstyle d^1} & & & & \\ H_i(G_{n+1}) & \longrightarrow & H_i(G_{n+1}, G_n) & & & & \end{array}$$

shows that $H_i(G_{n+1}, G_n) = 0$. Indeed, given an element $x \in H_i(G_{n+1}, G_n)$ there is an $x' \in H_i(G_n, G_{n-1})$ mapping to x. The top right map is an isomorphism by (b_{i-1}) as $n-1 \geq 3i-1 = 3(i-1)+2$. So x' maps to 0 in $H_{i-1}(G_{n-1})$, and thus there is an $x'' \in H_i(G_n)$ mapping to x'. Following x'' around the other two sides of the square gives 0 since these groups are part of an exact sequence, so $x = 0$. This proves (a_i).

The proof of (b_i) is similar. From the preceding paragraph and the assumption $n \geq 3i+2$ in (b_i) we have $E^1_{p,q} = H_q(G_{n-p}, G_{n-p-1}) = 0$ for all $q \leq i$ and $p+q = i+1$ as $n-p-1 \geq 3i+2-p-1 \geq 3i-3p+3 = 3q$ since $p \geq 1$. From the spectral sequence we conclude by the same reasoning as before that the differential

$$d^1 : E^1_{0,i+1} = H_{i+1}(G_n, G_{n-1}) \longrightarrow E^1_{-1,i+1} = H_{i+1}(G_{n+1}, G_n)$$

is surjective as $i \leq n-2$ when $n \geq 3i+2$. A diagram chase in

$$\begin{array}{ccccc} H_{i+1}(G_n, G_{n-1}) & \longrightarrow & H_i(G_{n-1}) & & \\ \downarrow{\scriptstyle d^1} & & \downarrow & & \\ H_{i+1}(G_{n+1}, G_n) & \longrightarrow & H_i(G_n) & \xrightarrow{\alpha_i} & H_i(G_{n+1}) \end{array}$$

now shows that $\alpha_i \colon H_i(G_n) \to H_i(G_{n+1})$ is injective. Indeed, let $x \in H_i(G_n)$ be in the kernel of α_i. Then there exists $x' \in H_{i+1}(G_{n+1}, G_n)$ mapping to x. As the left vertical map is surjective, there is a lift $x'' \in H_{i+1}(G_n, G_{n-1})$ of x'. Following x'' around the other two sides of the square gives 0, so $x = 0$. This proves (b_i) since (a_i) gives the surjectivity of α_i. □

Case (B) The result in this case is shown in [11] by a slightly different spectral sequence argument, and we include a proof here mainly for the reader's convenience.

As in the previous case, we do an induction on i. The statement is clearly true for $i = 0$. Suppose inductively that the following are valid for all $q < i$:

(a_q) $\beta_q\colon H_q(A^{s+2}_{n,k}) \to H_q(A^s_{n+1,k})$ is surjective for all $n \geq 3q$.

(b_q) $\beta_q\colon H_q(A^{s+2}_{n,k}) \to H_q(A^s_{n+1,k})$ is an isomorphism for all $n \geq 3q+2$.

We write the stabilization $A^{s+2}_{n,k} \to A^s_{n+1,k}$ as $G_n \to G_{n+1}$ and consider the action of G_{n+1} on $X_{n+1} = \mathcal{S}^{\pm}_c(M^s_{n+1,k})$. A vertex stabilizer is G_n, but unfortunately the stabilizer of a p–simplex for $p > 0$ is larger than $G_{n-p} = A^{s+2p+2}_{n-p,k}$ since the stabilizer includes diffeomorphisms that permute the vertices of the p–simplex. To fix this problem one could attempt to refine the definition of $\mathcal{S}^{\pm}_c(M^s_{n+1,k})$ so as to include the data of an ordering of the spheres in a sphere system. However, it is not evident that this complex would be highly connected. So instead we do something slightly different, letting X_{n+1} be the complex whose p–simplices are all the simplicial maps $\Delta^p \to \mathcal{S}^{\pm}_c(M^s_{n+1,k})$. These maps take vertices to vertices, but different vertices of Δ^p can have the same image, so we are allowing 'degenerate' simplices. (This means that X_{n+1} is no longer a simplicial complex, but it is a CW complex with the added structure of a Δ–complex, in the terminology of [8].) It is a standard fact in algebraic topology that enlarging a simplicial complex by adjoining degenerate ordered simplices in this way does not change its homology groups; see for example Theorem 4.6.8 of [20]. The group G_{n+1} still acts on the new X_{n+1}, and the action is transitive on vertices with G_n as vertex stabilizer, but the action is no longer transitive on higher-dimensional simplices. The stabilizers of p–simplices are the groups G_{n-r} for $r \leq p$, and this will be good enough for the proof in case (A) to go through without significant change.

The action of G_{n+1} on X_{n+1} restricts to an action of G_n on X_n, so as in case (A) we look at the double complex $(E_{q-1}G_n \otimes_{G_n} C_p) \oplus (E_q G_{n+1} \otimes_{G_{n+1}} C'_p)$ with boundary maps as before. The spectral sequence coming from the horizontal filtration converges to 0 when $p \leq n-2$, so the second spectral sequence has $E^\infty_{p,q} = 0$ for all $p+q \leq n-2$.

The inclusion $X_n \hookrightarrow X_{n+1}$ induces a bijection on orbits of p–simplices for $p \leq n-1$, so Shapiro's Lemma says that the term $E^1_{p,q}$ of the spectral sequence obtained from the vertical filtration is a direct sum of terms $H_q(G_{n-r}, G_{n-r-1})$ for $r \leq p$ if $p \leq n-1$. Again the coefficients are untwisted $\mathbb{Z}$'s since stabilizers do not permute vertices of a simplex.

The rest of the argument now proceeds as in case (A) as $n-r-1 \geq 3q$ for all $r \leq p$ if and only if this holds when $r = p$. ◻

Case (C) In order to keep the same notation as in the previous cases we take the liberty of writing $A^{s+1}_{n,k} \to A^s_{n,k+1}$ as $G_n \to G_{n+1}$. This is the map induced by the inclusion $M^{s+1}_{n,k} \hookrightarrow M^s_{n,k+1}$. We consider the action of G_{n+1} on $X_{n+1} = \mathcal{D}_c(M^s_{n,k+1}, C)$ where C is a single circle in the last boundary torus. Then G_n is a vertex stabilizer, and this acts on $X_n = \mathcal{D}_c(M^{s+1}_{n,k}, C)$ where C is a single circle in the last boundary sphere, as in case (A). The connectivities of X_n and X_{n+1} are $n-2$ and $n-1$ respectively and the inclusion $M^{s+1}_{n,k} \hookrightarrow M^s_{n,k+1}$ induces an equivariant map $X_n \to X_{n+1}$. The actions are transitive on simplices in each dimension. The stabilizer of a p–simplex for the action of G_{n+1} on X_{n+1} is $G_{n-p} = A^{s+p+1}_{n-p,k}$ and for the action of G_n on X_n it is $A^{s+p+2}_{n-p-1,k} = G_{n-p-1}$. Thus the stabilizers are the same as in (A). More precisely, taking $\widetilde{s} = s+1$ in (A), the E^1–pages are the same when $0 \le p \le n-1$, so we already know that the appropriate terms are 0. The result thus follows by the same argument as for (A), except that we do not need a full induction argument, but can instead appeal to the results proved in (A). $\square$

It is worth noting that γ automatically induces injections on all homology groups since if we follow it by the map coming from attaching a 2–handle to the last boundary torus to surger it to a sphere, the composition is the identity. One can also see from the proof of case (C) that an improvement in the stable range for stabilization with respect to n and s would yield an improvement for stabilization with respect to k.

References

[1] **K S Brown**, *Cohomology of Groups*, Graduate Texts in Mathematics 87, Springer-Verlag (1982)

[2] **D J Collins**, *Cohomological dimension and symmetric automorphisms of a free group*, Comment. Math. Helv. 64 (1989) 44–61

[3] **D Gabai**, *The Smale conjecture for hyperbolic 3–manifolds:* $\mathrm{Isom}(M^3) \simeq \mathrm{Diff}(M^3)$, J. Differential Geom. 58 (2001) 113–149

[4] **S Galatius**, *Mod p homology of the stable mapping class group*, Topology 43 (2004) 1105–1132

[5] **A Hatcher**, *Homeomorphisms of sufficiently large P^2–irreducible 3–manifolds*, Topology 15 (1976) 343–347
Revised version: *Spaces of incompressible surfaces*, available at:
`http://www.math.cornell.edu/~hatcher`

[6] **A Hatcher**, *A proof of the Smale conjecture*, Ann. of Math. 117 (1983) 553–607

[7] **A Hatcher**,*Homological stability for automorphism groups of free groups*, Comment. Math. Helv. 70 (1995) 39–62

[8] **A Hatcher**, *Algebraic topology*, Cambridge University Press, Cambridge (2002)

[9] **A Hatcher**, **K Vogtmann**, *Cerf theory for graphs*, J. London Math. Soc. 58 (1998) 633–655

[10] **A Hatcher**, **K Vogtmann**, *Rational homology of* $\mathrm{Aut}(F_n)$, Math. Res. Lett. 5 (1998) 759–780

[11] **A Hatcher**, **K Vogtmann**, *Homology stability for outer automorphism groups of free groups*, Algebr. Geom. Topol. 4 (2004) 1253–1272

[12] **H Hendriks**, **F Laudenbach**, *Difféomorphismes des sommes connexes en dimension trois*, Topology 23 (1984) 423–443

[13] **N V Ivanov**, *On the stabilization of the homology of the Teichmüller modular groups*, Algebra i Analiz 1 (1989) 110–126; translation in Leningrad Math. J. Vol. 1 (1990) 675–691

[14] **C A Jensen**, **N Wahl**, *Automorphisms of free groups with boundaries*, Algebr. Geom. Topol. 4 (2004) 543–569

[15] **F Laudenbach**, *Topologie de la dimension trois: homotopie et isotopie*, Astérisque 12, Société Mathématique de France, Paris (1974)

[16] **G R Livesay**, *Fixed point free involutions on the* 3*-sphere*, Ann. of Math. 72 (1960) 603–611

[17] **I Madsen**, **M Weiss**, *The stable moduli space of Riemann surfaces: Mumford's conjecture*, `arXiv:math.AT/0212321`

[18] **J McCool**, *On basis-conjugating automorphisms of free groups*, Canad. J. Math. 38 (1986) 1525–1529

[19] **J Milnor**, *Construction of universal bundles II*, Ann. of Math. 63 (1956) 430–436

[20] **E H Spanier**, *Algebraic topology*, McGraw-Hill Book Co., New York (1966)

[21] **U Tillmann**, *On the homotopy of the stable mapping class group*, Invent. Math. 130 (1997) 257–275

[22] **K Vogtmann**, *Homology stability for* $\mathrm{O}_{n,n}$, Comm. Algebra 7 (1979) 9–38

[23] **N Wahl**,*From mapping class groups to automorphisms of free groups*, to appear in J. London Math. Soc. `arXiv:math.AT/0406278`