THE 4-DIMENSIONAL DISC EMBEDDING THEOREM AND DUAL SPHERES

MARK POWELL, ARUNIMA RAY, AND PETER TEICHNER

Abstract. We modify the proof of the disc embedding theorem for 4-manifolds, which appeared as Theorem 5.1A in the book "Topology of 4-manifolds" by Freedman and Quinn, in order to construct geometrically dual spheres. These were claimed in the statement but not constructed in the proof.

Fundamental results in 4-manifold topology such as the existence and exactness of the surgery sequence, the s-cobordism theorem, and thence the classification of closed, simply connected topological 4-manifolds up to homeomorphism, rely on a closely related sphere embedding theorem with geometrically dual spheres, which is proven using the disc embedding theorem. We give an example to show that geometrically dual spheres are essential in applications.

1. Introduction

Manifolds of dimension at least 5 are studied using the Browder-Novikov-Sullivan-Wall surgery theory [Bro72, Sul96, Nov64, Wal99, KS77], Kreck's modified surgery [Kre99], and the s-cobordism theorem of Smale and Barden-Mazur-Stallings [Sma62, KS77, Bar63, Maz63, Sta67]. These strongly depend on the Whitney trick, which does not apply when the ambient manifold is 4-dimensional. Casson [Cas86] gave an infinite construction of certain 4-dimensional objects, now called *Casson handles*, within simply connected 4-manifolds. He showed that exactness of the surgery sequence for simply connected 4-manifolds and the 5-dimensional h-cobordism theorem would follow if every Casson handle were homeomorphic, relative to its attaching region, to the interior of a standard handle $D^2 \times \mathring{D}^2$. This last step is Freedman's celebrated theorem [Fre82].

The *disc embedding theorem* [Fre84] combines work of Casson, Freedman, and Quinn. In a topological 4-manifold with *good* fundamental group, the disc embedding theorem replaces an immersed disc with embedded boundary and an algebraically dual sphere by a locally flat embedded disc with the same boundary as the original disc. The full disc embedding theorem was first stated in the book "Topology of 4-manifolds" by Freedman and Quinn [FQ90, Theorem 5.1A]. This book gave a new proof using a generalisation of Casson handles built out of *capped gropes*, variously called a *skyscraper* [BKKPR], a *cope* [Fre84], or a *generalized infinite tower* [FQ90]. The Freedman-Quinn capped grope approach is key to the proof of the disc embedding theorem for nontrivial fundamental groups.

The goal of this article is to modify part of the Freedman-Quinn proof of the disc embedding theorem, in order to fill a gap in the proof of [FQ90, Theorem 5.1A and Corollary 5.1B] related to geometrically dual spheres. We elucidate further below, but in brief one needs algebraically dual spheres for the input of the disc embedding theorem, while for many applications one needs geometrically dual spheres in the output. Here we say that two surfaces are *geometrically dual* if they intersect transversely in precisely one point.

We will explain how to deduce the *sphere embedding theorem*, which gives conditions under which a collection of immersed spheres with framed algebraically dual spheres can be regularly homotoped, using Whitney discs produced by the disc embedding theorem, to a collection of locally flat pairwise disjoint embedded spheres with geometrically dual spheres.

The geometrically dual spheres in the outcome of the sphere embedding theorem, constructed using the geometrically dual spheres in the disc embedding theorem, are important

2020 *Mathematics Subject Classification.* 57K40, 57N35.

in several applications, especially to surgery theory. Surgery on an embedded, framed sphere without a geometrically dual sphere might change the fundamental group of the ambient 4-manifold, whereas geometrically dual spheres guarantee control over the fundamental group and are hence essential.

At the end of the article (Example 1.5) we shall construct an embedded, framed sphere S in $S^2 \times S^2$ with an algebraically dual sphere, such that surgery on S produces a 4-manifold with nontrivial fundamental group (the first construction of such a sphere is due to [Sat89, Example 4.1]). This implies, unlike in high-dimensional surgery, that geometric duals cannot be found *post hoc* and that the application to surgery theory, and thence to 4-manifold classification results, hinges on providing geometric duals together with the embedded spheres.

1.1. **The disc embedding theorem.** This is one of the main tools in 4-dimensional topology, similar in spirit to Dehn's lemma in dimension 3. Under the assumptions given below, the theorem states that a collection of maps of discs to a 4-manifold can be replaced by *pairwise disjoint locally flat embeddings*. Dual spheres are essential both in the input and the output of the precise statement in Theorem A and will be explained next.

The equivariant intersection form of a connected 4-manifold M is a pairing

$$\lambda\colon H_2(M,\partial M;\mathbb{Z}[\pi_1 M]) \times H_2(M;\mathbb{Z}[\pi_1 M]) \longrightarrow \mathbb{Z}[\pi_1 M].$$

As well as having nonempty boundary, the topological 4-manifold M may be nonorientable or noncompact, in which case the definition of λ via cap products involves cohomology with compact support and coefficients twisted by the orientation character $w\colon \pi_1(M) \to \{\pm 1\}$. For topological 4-manifolds, we will explain the *reduced self-intersection number*

$$\widetilde{\mu}\colon \pi_2(M) \longrightarrow \mathbb{Z}[\pi_1 M]/\langle g - w(g)g^{-1}, \mathbb{Z}\cdot 1\rangle$$

in Section 1.2, and in more detail in Section 4. By Corollary 4.5, $\widetilde{\mu}$ is determined by λ if M is orientable. In this case, the assumptions $\widetilde{\mu}(g_i) = 0$ below are implied by $\lambda(g_i, g_i) = 0$.

Given maps $\{f_i\}_{i=1}^k$ of discs or spheres to M, a second collection $\{g_i\colon S^2 \to M\}_{i=1}^k$ is *algebraically dual* to the $\{f_i\}_{i=1}^k$ if the algebraic intersection form satisfies $\lambda(f_i, g_j) = \delta_{ij}$. If this is true geometrically, we say that the collection $\{g_i\}$ is *geometrically dual* to the $\{f_i\}$. More precisely, this means that $f_i \cap g_j$ is a single point for $i = j$ and is empty for $i \neq j$, and that each intersection point has local coordinates $\mathbb{R}^4 \hookrightarrow M$ in which f_i and g_i are linear.

Theorem A (Disc embedding theorem cf. [FQ90, Theorem 5.1A])**.** *Let M be a connected 4-manifold with good fundamental group. Consider a continuous map*

$$F = (f_1, \dots, f_k)\colon (D^2 \sqcup \cdots \sqcup D^2, S^1 \sqcup \cdots \sqcup S^1) \longrightarrow (M, \partial M)$$

that is a locally flat embedding on the boundary and that admits algebraically dual spheres $\{g_i\}_{i=1}^k$ satisfying $\lambda(g_i, g_j) = 0 = \widetilde{\mu}(g_i)$ for all i, j. Then there exists a locally flat embedding

$$\overline{F} = (\overline{f}_1, \dots, \overline{f}_k)\colon (D^2 \sqcup \cdots \sqcup D^2, S^1 \sqcup \cdots \sqcup S^1) \hookrightarrow (M, \partial M)$$

such that $\overline{F}$ has the same boundary as F and admits a generically immersed, geometrically dual collection of framed spheres $\{\overline{g}_i\}_{i=1}^k$, such that $\overline{g}_i$ is homotopic to g_i for each i.

Moreover, if f_i is a generic immersion, then it induces a framing of the normal bundle of its boundary circle. The embedding $\overline{f}_i$ may be assumed to induce the same framing.

Remark 1.1. Our statement of Theorem A differs from [FQ90, Theorem 5.1A] in that we emphasise that the input is purely homotopy theoretic, and we control the homotopy classes of the dual spheres. The interested reader can use the discussion in Remark 3.2 to see that even the original continuous maps f_i induce framings modulo 2 on the boundary circles.

We discuss topological immersions in Section 1.2, where we also explain why a topological generic immersion f admits a linear normal bundle $\nu(f)$ whose total space is embedded in M apart from finitely many plumbings. If the base of $\nu(f)$ is a disc, and so in particular contractible, then this bundle has a unique trivialization, or framing, which is used in the last paragraph of Theorem A.

The notion of a *good* group will be recalled in Definition 5.7. Practically, it suffices to know that the class of good groups is known to contain groups of subexponential growth [FT95a, KQ00], and to be closed under subgroups, quotients, extensions, and colimits [FQ90, p. 44]. In particular, all finite groups and all solvable groups are good. It is not known whether non-abelian free groups are good.

The proof of the disc embedding theorem from [FQ90] begins with immersed discs and has three distinct steps. We describe the $k = 1$ case for ease of exposition. First upgrade the immersed disc f to a 1-*storey capped tower* $\mathcal{T}$ with at least four surface stages, whose attaching region coincides with the framed boundary of f (see Section 4 for a precise definition of capped towers). Then show that every 1-storey capped tower with at least four surface stages contains a skyscraper with the same attaching region. Then, using *decomposition space theory*, prove that every skyscraper is homeomorphic to a handle $D^2 \times D^2$ relative to its attaching region. The proof given in [FQ90, p. 86-7] does not mention geometrically dual spheres, although they appear in the theorem statement.

In this paper, we give a modified proof of the disc embedding theorem, including the construction of the claimed geometrically dual spheres. More precisely, we only modify the first step of the proof, producing 1-storey capped towers additionally equipped with geometrically dual spheres. Our modification utilises dual *capped surfaces* obtained from *Clifford tori*. These dual capped surfaces act as a sort of dual sphere factory, allowing us to create arbitrarily many collections of pairwise disjoint geometrically dual spheres. The remainder of our proof appeals to the original proof of Freedman and Quinn that 1-storey capped towers contain skyscrapers and that skyscrapers are standard handles.

1.2. **Generic immersions and intersection numbers.** In the smooth category, the first step in the proof of Theorem A is to bring the original maps into the most "generic" position. Specifically, as we discuss in Section 3, homotopy classes of smooth maps of a compact surface to a 4-manifold are represented by *generic immersions*, which are immersions whose only singularities are transverse double points in the interior.

In the topological category, we use this local description as the definition of a generic immersion. In particular, a generic immersion is locally a flat embedding and hence restricts to a locally flat embedding of the boundary. Proposition 3.1 states that every continuous map of a compact surface is homotopic to a generic immersion $f\colon \Sigma \looparrowright M$, and in fact for M connected and $p \in M \setminus f(\Sigma)$ there is a smooth structure on Σ and on $M \setminus \{p\}$ with respect to which f is a smooth generic immersion. In particular it follows that a topological generic immersion $f\colon \Sigma \looparrowright M$ has a linear normal bundle $\nu(f)$.

A *regular homotopy* in the smooth category is a homotopy through immersions. A smooth regular homotopy of generically immersed surfaces in a 4-manifold is generically a concatenation of (smooth) isotopies, finger moves, and Whitney moves [GG73, Section III.3]. A *topological regular homotopy* of generically immersed surfaces in a 4-manifold is by definition a concatenation of (topological) isotopies, finger moves, and Whitney moves. We show the following useful fact, which combines [FQ90, Lemma 1.2 and Proposition 1.6]. The latter proposition was asserted there without proof.

Theorem 1.2. *Let Σ be a disjoint union of discs or spheres, and let M be a 4-manifold. The subspace of generic immersions in the space of all continuous maps leads to a bijection*

$$\frac{\{(f_1, \ldots, f_m)\colon \Sigma = \Sigma_1 \sqcup \cdots \sqcup \Sigma_m \looparrowright M \mid \mu(f_i)_1 = 0,\ i = 1, \ldots, m\}}{\{\textit{isotopies, finger moves, Whitney moves}\}} \longleftrightarrow [\Sigma, M]_\partial,$$

where $\mu(f_i)_1 \in \mathbb{Z}$ denotes the signed sum of double points of f_i whose double point loops are trivial in $\pi_1(M)$, and $[\Sigma, M]_\partial$ denotes the set of homotopy classes of continuous maps that restrict on $\partial\Sigma$ to locally flat embeddings disjoint from the image of the interior of Σ. Moreover, any such homotopy between generic immersions is homotopic rel. $\Sigma \times \{0,1\}$ to a sequence of isotopies, finger moves and Whitney moves.

More generally, the self-intersection number $\mu(f)$ of a generic immersion f of a disc or sphere into M is obtained by summing signed group elements $g \in \pi_1(M)$ corresponding to double points in the interior of f (see Section 4.2 for details). The ambiguity of the choice of sheets at each double point leads to the relations $g - w(g)g^{-1}$ in $\mathbb{Z}[\pi_1 M]$. Local cusp homotopies allow us to change the coefficient $\mu(f)_1$ at the trivial fundamental group element 1 at will, and correspondingly we factor out by $\mathbb{Z} \cdot 1$. We obtain the *reduced* self-intersection number, already used in the assumptions of Theorem A:

$$\widetilde{\mu} \colon \pi_2(M) \longrightarrow \mathbb{Z}[\pi_1 M]/\langle g - w(g)g^{-1}, \mathbb{Z} \cdot 1\rangle.$$

While $\mu(f)$ depends on the choice of generically immersed representative $f \colon S^2 \looparrowright M$, the reduced version $\widetilde{\mu}$ only depends on $[f] \in \pi_2(M)$. Any two generic immersions homotopic to f are of course themselves homotopic. Use the injectivity in Theorem 1.2 to obtain a regular homotopy, and deduce that $\widetilde{\mu}$ is well-defined on $\pi_2(M)$, since the quantity is evidently unchanged by isotopies, finger moves, and Whitney moves.

For any $f \colon S^2 \looparrowright M$, the invariant μ satisfies

$$\lambda(f, f) = \mu(f) + \overline{\mu(f)} + e(f) \cdot 1 \in \mathbb{Z}[\pi_1 M] \tag{1}$$

where $e(f) \in \mathbb{Z}$ is the *Euler number* of the normal bundle $\nu(f)$ (see Lemma 4.4) and the involution $\overline{g} := w(g)g^{-1}$ is extended linearly to the group ring. Apply the augmentation map $\varepsilon \colon \mathbb{Z}[\pi_1 M] \to \mathbb{Z}$ to this equation, to see that modulo 2, $\varepsilon(\lambda(f, f)) \equiv e(f)$ only depends on the homotopy class of f. This is also the *Stiefel-Whitney number* $w_2(f) \in \mathbb{Z}/2$ of $\nu(f)$ which will be used in the assumption of the next theorem. As in the discussion above one sees that $w_2(a) = 0$ if and only if $a \in \pi_2(M)$ is represented by a generic immersion $f \colon S^2 \looparrowright M$ that can be framed, i.e. whose normal bundle is trivial.

1.3. **The sphere embedding theorem.** Most applications of the disc embedding theorem, including the existence and exactness of the topological surgery sequence and the s-cobordism theorem, involve changing a collection of generically immersed spheres with vanishing intersection and self-intersection numbers, by a regular homotopy, to pairwise disjoint embedded spheres. The *sphere embedding theorem*, which we shall deduce from the disc embedding theorem in Section 8, describes precisely when such embedded spheres may be found, with the main assumption being the existence of algebraically dual spheres and the output giving geometrically dual spheres.

Theorem B (Sphere embedding theorem with framed duals)**.** *Let M be a connected 4-manifold with good fundamental group and consider a continuous map*

$$F = (f_1, \dots, f_k) \colon (S^2 \sqcup \cdots \sqcup S^2) \longrightarrow M$$

satisfying $\widetilde{\mu}(f_i) = 0$ for every i and $\lambda(f_i, f_j) = 0$ for $i \neq j$, with a collection of algebraically dual spheres $\{g_i\}_{i=1}^k$ with $w_2(g_i) = 0$ for each i. Then there is a locally flat embedding

$$\overline{F} = (\overline{f}_1, \dots, \overline{f}_k) \colon (S^2 \sqcup \cdots \sqcup S^2) \hookrightarrow M$$

with $\overline{F}$ homotopic to F and with a generically immersed, geometrically dual collection of framed spheres $\{\overline{g}_i\}_{i=1}^k$, such that $\overline{g}_i$ is homotopic to g_i for each i.

Moreover, if f_i is a generic immersion and $e(f_i) \in \mathbb{Z}$ is the Euler number of the normal bundle $\nu(f_i)$, then $\overline{f}_i$ is regularly homotopic to f_i if and only if $e(f_i) = \lambda(f_i, f_i)$.

Remark 1.3. The assumption $\widetilde{\mu}(f_i) = 0$ implies that $\lambda(f_i, f_i) \in \mathbb{Z} \cdot 1 \subseteq \mathbb{Z}[\pi_1 M]$, by (1). In this case, (1) gives an equation of integers $\lambda(f_i, f_i) = 2 \cdot \mu(f_i)_1 + e(f_i)$ and hence the last condition $e(f_i) = \lambda(f_i, f_i)$ in Theorem B is equivalent to the vanishing of $\mu(f_i)$.

The geometrically dual spheres in the conclusion imply that inclusion induces an isomorphism $\pi_1(M) \cong \pi_1(M \setminus \bigcup_i \overline{f}_i)$. We refer to a collection of immersed surfaces whose removal does not change the fundamental group as π_1*-negligible*.

If $\lambda(f_i, f_i) = 0$ for all i, the sphere embedding theorem is equivalent to the decomposition of M as a connected sum with copies of $S^2 \times S^2$, as explained in Remark 8.3. We formulate a special case of this decomposition when the $\{g_i\}$ also form a Lagrangian.

Corollary 1.4. *Let M be a connected 4-manifold with good fundamental group and let H be a hyperbolic form in $(\pi_2(M), \lambda_M, \widetilde{\mu}_M)$, meaning that H is a $\mathbb{Z}[\pi_1 M]$-submodule of $\pi_2(M)$, generated by a hyperbolic basis consisting of classes $a_1, \dots, a_k, b_1, \dots, b_k \in H$ with*

$$\lambda(a_i, b_j) = \delta_{ij},\ \lambda(a_i, a_j) = 0 = \lambda(b_i, b_j) \text{ and } \widetilde{\mu}(a_i) = 0 = \widetilde{\mu}(b_i) \text{ for all } i, j.$$

Then there is a homeomorphism $M \approx (\#_{i=1}^k S^2 \times S^2) \# M'$ with a connected sum that on π_2 sends a_i to $[S^2_i \times \{\mathrm{pt}_i\}]$ and b_i to $[\{\mathrm{pt}_i\} \times S^2_i]$. In particular, H is an orthogonal summand freely generated by $\{a_i, b_i\}$ and $\pi_2(M) \cong H \perp \pi_2(M')$.

This corollary was stated on the first page of Freedman's ICM talk [Fre84]. Together with Donaldson's theorem on definite intersection forms for smooth 4-manifolds [Don83], it implies the existence of infinitely many non-smoothable 4-manifolds. For example, the $E_8 \# E_8$-manifold can be obtained from the $K3$ surface by removing a hyperbolic form of rank 6. Example 1.5 below shows that Corollary 1.4 does not follow from the disc embedding theorem stated by Freedman later in his ICM paper. In fact, we could not find a reference where Corollary 1.4 was proven or even stated.

Theorem B is the key input in the proof of the existence and exactness of the topological surgery sequence in dimension 4 and the 5-dimensional s-cobordism theorem for good fundamental groups. Indeed, the sphere embedding theorem is integral to any known classification result for topological 4-manifolds, including those that use Kreck's modified surgery theory [Kre99], for example [HKT09]. Another example is the main step in [HK88, Lemma 4.1], which uses Corollary 1.4 to move from the easier stable classification up to connected sums with copies of $S^2 \times S^2$ to unstable classification results.

As mentioned earlier, an important observation is that the geometrically dual spheres $\{\overline{g}_i\}$ in the output of the sphere embedding theorem ensure that the fundamental group of the ambient 4-manifold remains unaffected by surgery on the embedded spheres $\{\overline{f}_i\}$. This subtlety only occurs in dimension 4, because in higher dimensions the fundamental group does not change when removing middle dimensional spheres, and geometric duals can be constructed using the Whitney trick after constructing the embedded spheres.

Example 1.5. There is a framed, embedded 2-sphere in $S^2 \times S^2$ that admits an algebraically dual sphere but no geometrically dual sphere. See Section 9.

This explicitly exhibits the phenomenon that surgery on an embedded sphere with an algebraically dual sphere can change the fundamental group. As a consequence, in dimension 4 geometric duals have to be constructed concurrently with the embedded spheres, as we shall do in this paper.

Theorem B gives criteria under which a map of a union of spheres into M with framed duals can be improved to an embedding. In work with Kasprowski, we plan to extend this from unions of spheres to arbitrary compact surfaces, and to allow unframed dual spheres. When dual spheres cannot be framed, a secondary obstruction to embedding called the *Kervaire-Milnor invariant* arises. For the case of spheres, see [FQ90, Sections 10.5 and 10.8], as corrected by Stong [Sto94]. Our extension of this invariant to other surfaces is new.

Conventions. All manifolds are assumed to be based, in order to define homotopy groups and equivariant intersection numbers. Topological embeddings are always assumed to be locally flat.

Outline. In Section 2 we list some of the occurrences of the geometrically dual spheres from [FQ90, Theorem 5.1A and Corollary 5.1B] in the literature. In Section 3 we discuss generic immersions, show that every continuous map of a surface is homotopic to a generic immersion, and prove Theorem 1.2 ([FQ90, Proposition 1.6]).

In Section 4, we review the objects and geometric constructions required for our proofs. In Section 5 we recall some further results needed for the proof of the disc embedding theorem.

Section 6 contains the main technical results needed for the construction of geometrically dual spheres. These results are then applied in Section 7 to prove the disc embedding theorem (Theorem A). Section 8 deduces a variant of the disc embedding theorem ([FQ90, Corollary 5.1B]) and the sphere embedding theorem (Theorem B), and proves Corollary 1.4.

Finally, in Section 9 we construct the sphere in Example 1.5 and prove its properties.

Acknowledgements. We are indebted to the 4-manifolds semester at the Max Planck Institute for Mathematics in Bonn in 2013, and in particular to lectures of Michael Freedman, Frank Quinn, and Robert Edwards given as part of the semester. Much of the work leading to this paper took place courtesy of a trimester in Bonn, this time at the Hausdorff Institute for Mathematics in 2016, or during subsequent visits by MP to the MPIM, or by AR to Durham University. We thank these institutions for their hospitality. We thank Stefan Friedl, Daniel Kasprowski, Allison N. Miller, and Patrick Orson for their insights and valuable discussions. We warmly thank Allison for creating many of the pictures within this paper, and Michael Freedman for encouraging us to write this paper.

Note. Certain parts of the proofs given in this paper also appear in the forthcoming book [BKKPR].

2. Geometrically dual spheres in the literature

We discuss where the geometrically dual spheres in the outcome of the disc embedding theorem appear in the literature. We focus first on the book [FQ90], which has generally replaced the original papers of Casson [Cas86], Freedman [Fre82, Fre84], and Quinn [Qui82, Qui88] as the canonical source material on the results and ramifications of Freedman's work on topological 4-manifolds.

2.1. **The Freedman-Quinn book** [FQ90]**.** The geometrically dual spheres in the conclusion of Theorem 5.1A were used to produce geometrically dual spheres in Corollary 5.1B. This in turn was the key theorem cited to prove that the topological surgery sequence for good fundamental groups is exact. In fact, the sphere embedding theorem was needed twice, to define the action of $L_5^s(\mathbb{Z}[\pi_1(X)])$ on the simple structure set of X and to show exactness of the surgery sequence at the normal invariants (see [BKKPR, Chapter 20] for more details). The proof of Theorem 11.3A omitted the observation that geometrically dual spheres are essential for performing surgery in 4-manifolds without inadvertently modifying the fundamental group. We give an explicit example illustrating this point in Section 9.

The geometrically dual spheres in Corollary 5.1B also arose in the proofs of the π_1-negligible embedding (Theorem 10.5A, p. 177), the geometric plus construction (Theorem 11.1A), the $\pi - \pi$ lemma (p. 216), the technical controlled h-cobordism theorem (Theorem 7.2C), and the annulus conjecture (Theorem 8.1A, p.114-6). The annulus conjecture was used to prove topological transversality (Section 9.5), existence and uniqueness of normal bundles (Section 9.3), and smoothing results (Chapter 8).

2.2. **Casson and Freedman papers.** Casson's construction [Cas86, Lecture I] of a Casson handle in a 4-manifold with boundary was closer in spirit to [FQ90, Corollary 5.1B] than [FQ90, Theorem 5.1A] (see Section 8). The construction crucially used the fact that the ambient manifold is simply connected. Casson applied his embedding theorem to represent a hyperbolic pair in the surgery kernel for a simply connected manifold by geometrically dual Casson handles whose base stages are spheres.

In [Fre82, Theorem 1.2], Freedman claimed to prove the exactness of the surgery sequence by embedding half of each hyperbolic pair representing the surgery kernel in a simply connected manifold by an embedded sphere. As shown by the example in Section 9, this is not sufficient: we also need to construct geometrically dual spheres. However, this is possible in the simply connected case by applying Casson's construction of geometrically dual Casson handles mentioned above. This construction was known to Freedman, since he mentioned it in the paragraph immediately following his proof of the exactness of the surgery sequence, in the proof of the Addendum to [Fre82, Theorem 1.2]. Thus the combination of Casson's construction and Freedman's proof that Casson handles are homeomorphic to standard handles provides embedded spheres with geometric duals in the simply connected case.

2.3. **Quinn's paper.** Quinn's proof from [Qui82] that non-compact, connected 4-manifolds admit smooth structures relies on Freedman's proof that Casson handles are standard, but does not rely on the dual spheres constructed in this paper. This also applies to topological transversality (Theorem 3.3) and the existence and uniqueness of normal bundles.

2.4. **Elsewhere.** In a similar manner to classical surgery theory, the use of Kreck's modified surgery [Kre99] to obtain classification results on 4-manifolds, for example in [HK88], [HK93] and [HKT09], needs geometrically dual spheres to avoid changing the fundamental group by surgery on embedded spheres obtained using the sphere embedding theorem.

Finally, in knot concordance geometrically dual spheres are needed to show that the complement of the topological slice disc produced in [GT04] for a knot with Alexander polynomial one has fundamental group $\mathbb{Z}$, and similarly in [FT05] for slice disc exteriors with fundamental group $\mathbb{Z} \ltimes \mathbb{Z}[1/2]$.

3. Generic immersions in smooth and topological 4-manifolds

Recall that for a compact surface Σ and 4-manifold M, a *generic immersion* in the smooth category, written $f\colon \Sigma^2 \looparrowright M^4$, is a smooth map which is an embedding, except for a finite number of transverse double points in the interior. This means that f is an embedding on $\partial\Sigma$ and there are coordinates on Σ and on M such that restricted to the interior, f coincides in local coordinates in M with either the inclusion $\mathbb{R}^2 \times \{0\} \subset \mathbb{R}^4$ or a transverse double point $\mathbb{R}^2 \times \{0\} \cup \{0\} \times \mathbb{R}^2 \subset \mathbb{R}^4$.

In a smooth 4-manifold M, the set of generic immersions is open and dense in the Whitney topology on $C^\infty(\Sigma, M)$ [GG73, Chapter III, Corollary 3.3]. These are exactly the stable maps in the smooth mapping space [GG73, Chapter III, Theorem 3.11] where a map f is said to be *stable* if it has a neighbourhood in $C^\infty(\Sigma, M)$ such that every map in the neighbourhood can be obtained by pre-composing with a diffeomorphism of Σ and post-composing with a diffeomorphism of M, with *both diffeomorphisms isotopic to the identity.* It is well known that any continuous map between smooth manifolds is arbitrarily close to a smooth map (see, for example, [Lee03, Theorem 10.21]). Since moreover the collection of generic immersions is open and dense in $C^\infty(\Sigma, M)$, a smooth map can be further perturbed to a (smooth) generic immersion. Since the perturbations may be chosen to be small, a proper continuous map may be perturbed to a proper smooth generic immersion (see [Spr72, Lemma 1]).

A continuous map between topological manifolds is called an *immersion* if it is locally an embedding. A continuous map of a surface to a topological 4-manifold with the same local behaviour as a smooth generic immersion will be called a *generic immersion* in the topological category. Note that this implies that the map is a *locally flat* embedding near points with a single inverse image.

A smooth homotopy H between between smooth maps $\Sigma \to M$ is said to be generic if the corresponding map $\Sigma \times [0,1] \to M \times [0,1]$ is a generic smooth map. Whitney's classification of singularities [Whi43, Whi44] of generic maps from 3-manifolds to 5-manifolds implies that the singularities of the track of a generic homotopy H as above consist of finger moves, Whitney moves and cusps. These arise at finitely many times $t \in I$, when $H_t\colon \Sigma \to M$ is not a generic immersion but either:

(i) H_t has a tangency, increasing or decreasing the double point set by a pair with opposite signs, corresponding to a finger/Whitney move, or
(ii) the rank of the derivative of H_t drops at a single point, creating a cusp where one double point appears or disappears.

A topological generic homotopy is defined to be a concatenation of finger moves, Whitney moves, and cusps.

A continuous homotopy between smooth maps $\Sigma \to M$ may be perturbed (rel. boundary) to produce a smooth generic homotopy. Since the perturbation may be chosen to be small, a proper homotopy may be perturbed to a proper smooth generic homotopy. Here is the main technical result of this section.

Proposition 3.1. *Let Σ be a compact surface and let M be a 4-manifold.*

(1) Every continuous map $f\colon (\Sigma, \partial\Sigma) \to (M, \partial M)$ is homotopic to a generic immersion, smooth in $M \setminus \{q\}$ for some point $q \in M$.
(2) Every continuous homotopy $H\colon (\Sigma, \partial\Sigma) \times [0,1] \to (M, \partial M)$ that restricts to a smooth generic immersion on $\Sigma \times \{i\}$ for each $i = 0, 1$, in some structure on M minus a point q, is homotopic rel. $\Sigma \times \{0,1\}$ to a generic homotopy, that is smooth in some structure on $M \setminus \{r\}$ for some point $r \in M$.

Proof. First we prove (1). We may assume without loss of generality that M is connected. Choose a smooth structure on Σ. The complement of a point p in the interior of M is smoothable relative to any fixed chosen smoothing of the boundary of M [FQ90, Theorem 8.2], [Qui82, Qui86]. Choose a smooth structure on $M \setminus \{p\}$.

Homotope the restriction of f to $\partial\Sigma$ to a smooth embedding, and extend this to a homotopy of Σ supported in a collar near the boundary (or use a given embedding to start with and work rel. boundary). Consider the smooth surface $\Sigma_p := \Sigma \smallsetminus f^{-1}(p)$ which comes with a proper map $f|\colon \Sigma_p \to M \smallsetminus \{p\}$ that is properly homotopic to a smooth proper map f'. By Sard's theorem, some point q in the interior of $M \smallsetminus \{p\}$ does not lie in the image of f'. Moreover, because Σ is compact and the homotopy was proper, f' maps each end of Σ_p to p (as did f). Send all of $f^{-1}(p)$ to p to extend f' to a continuous map $\Sigma \to M \smallsetminus \{q\}$, homotopic in M to the original map f. The latter homotopy is produced from the proper homotopy between $f|$ and $f'|$ by mapping every end to p.

Now as explained above, the map f' is homotopic to a smooth generic immersion $f''\colon \Sigma \looparrowright M \smallsetminus \{q\}$. Add q back in and forget the smooth structure to yield a topological generic immersion, noting that as required f'' is smooth in some structure on $M \setminus \{q\}$ by construction.

Now, to prove (2), consider a homotopy $H\colon \Sigma \times [0,1] \to M$ whose restriction $H|\colon \Sigma \times \{i\} \to M$ is a smooth generic immersion in some smooth structure on $M \setminus \{q\}$ for some $q \in M$, for each $i = 0, 1$. We follow a similar strategy as the proof of (1). Use a smoothing of M away from q, and consider

$$(\Sigma \times [0,1])_q := (\Sigma \times [0,1]) \setminus H^{-1}(\{q\}) \subseteq M \smallsetminus \{q\}.$$

The proper map $(\Sigma \times [0,1])_q \to M \setminus \{q\}$ is properly homotopic rel. $\Sigma \times \{0,1\}$ to a proper smooth map $H'\colon (\Sigma \times [0,1])_q \to M \setminus \{q\}$, that by Sard's theorem misses at least one point $r \in M$. Since $\Sigma \times [0,1]$ is compact and the homotopy was proper, H' maps each end of $H^{-1}(q)$ to q. Extend H' to a continuous map $H''\colon \Sigma \times [0,1] \to M \setminus \{r\}$. Choose a smooth structure on $M \setminus \{r\}$, such that $H| = H''|\colon \Sigma \times \{0,1\} \to M$ is a smooth generic immersion. To achieve this, start with the original smooth structure on $M \setminus \{q\}$ restricted to a neighbourhood of $H''(\Sigma \times \{0,1\}) \subseteq M \setminus \{q, r\}$, and extend that structure to a smooth structure on $M \setminus \{r\}$.

Now homotope $H''|_{\Sigma \times [0,1]}$ rel. $\Sigma \times \{0,1\}$ to a smooth generic homotopy in $M \setminus \{r\}$. Add r back in and forget the smooth structure to yield a topological generic homotopy, noting that as required H'' is smooth in some smooth structure on $M \setminus \{r\}$. □

Remark 3.2. Let us discuss some consequences of Proposition 3.1 for normal bundles of generic immersions. A generic immersion $f\colon \Sigma \looparrowright M$ has a linear normal bundle $\nu(f) \to \Sigma$, in both the smooth and topological categories. To see this, by Proposition 3.1 f is generically homotopic to a smooth immersion in some smooth structure on $M \setminus \{q\}$. The map f comes with a map $\nu(f) \to M$ of the total space into M, which is an embedding away from a finite number of plumbings near the double points of f. In the case that Σ has nonempty boundary, assume that $f^{-1}(\partial M) = \partial\Sigma$ and that we are given a fixed framing of the normal bundle restricted to the boundary. We call f *framed* if $\nu(f)$ comes with a trivialisation and *twisted* if the (relative) Stiefel-Whitney class $w_2(f) := w_2(\nu(f)) \in H^2(\Sigma, \partial\Sigma; \mathbb{Z}/2)$ is nonzero. If Σ is oriented, $f|_{\partial\Sigma}$ is framed and f is not twisted, then one can add local cusps to f (corresponding to a non-regular homotopy between generic immersions) until the relative Euler number in $H^2(\Sigma, \partial\Sigma; \mathbb{Z})$ of $\nu(f)$ vanishes and hence $\nu(f)$ becomes trivial and the framing on f induces the given framing on $\partial\Sigma$.

Now we prove Theorem 1.2 about homotopy classes $[\Sigma, M]$ of maps $f\colon \Sigma \to M$, when Σ is a union of spheres or discs. We write $\{\Sigma, M\}_\partial$ for the subspace of all continuous maps that restrict on $\partial\Sigma$ to locally flat embeddings disjoint from the image of the interior of Σ, and $[\Sigma, M]_\partial$ for the set of homotopy classes of such maps. In this theorem we do not assume that $f^{-1}(\partial M) = \partial\Sigma$.

Choose a local orientation of M at the basepoint and assume that Σ comes with a whisker to the basepoint. The proof will use topological transversality, which we state first.

Theorem 3.3 ([Qui82, Qui88] (see also [FQ90, Section 9.5]))**.** *Let Σ_1 and Σ_2 be locally flat proper submanifolds of a topological 4-manifold M that are transverse to ∂M. There is an isotopy of M, supported in any given neighbourhood of $\Sigma_1 \cap \Sigma_2$, taking Σ_1 to a submanifold Σ_1' transverse to Σ_2.*

Theorem 1.2. *Let Σ be a disjoint union of discs or spheres, and let M be a 4-manifold. The subspace of generic immersions in the space of all continuous maps leads to a bijection*

$$\frac{F = \{(f_1, \dots, f_m)\colon \Sigma = \Sigma_1 \sqcup \cdots \sqcup \Sigma_m \looparrowright M \mid \mu(f_i)_1 = 0,\ i = 1, \dots, m\}}{\{\textit{isotopies, finger moves, Whitney moves}\}} \longleftrightarrow [\Sigma, M]_\partial,$$

where $\mu(f_i)_1 \in \mathbb{Z}$ denotes the signed sum of double points of f_i whose double point loops are trivial in $\pi_1(M)$, and $[\Sigma, M]_\partial$ denotes the set of homotopy classes of continuous maps that restrict on $\partial\Sigma$ to locally flat embeddings disjoint from the image of the interior of Σ. Moreover, any such homotopy between generic immersions is homotopic rel. $\Sigma \times \{0,1\}$ to a sequence of isotopies, finger moves and Whitney moves.

Theorem 1.2 could be rephrased more succinctly, but perhaps less transparently, as the statement that the inclusion of the space of generic immersions into $\{\Sigma, M\}_\partial$ is a 1-connected map. Note that the theorem implies [FQ90, Proposition 1.6].

Proof. The map is well-defined since any isotopy, finger move, or Whitney move is a homotopy. Note that for each i, $\mu(f_i)_1$ can be changed arbitrarily by (non-regular) cusp homotopies. Therefore by Proposition 3.1(1) the map is surjective.

For injectivity, consider generic immersions $F, F'\colon \Sigma \looparrowright M$, which restrict to embeddings on $\partial\Sigma$ that miss the image of the interior, and assume that F and F' are homotopic. Using topological transversality, we can assume that F and F' intersect transversely after performing an isotopy. Choose a point $q \in M$ disjoint from the image of both F and F'. Since F and F' are generic immersions intersecting transversely, there is a neighbourhood, namely the union of the images of the normal bundles, within which F and F' are smooth generic immersions. Extend this smooth structure over $M \smallsetminus \{q\}$.

Now by Proposition 3.1(2), we can replace the homotopy from F to F' by a smooth generic homotopy, by performing a homotopy rel. $\Sigma \times \{0,1\}$. Then we saw earlier that the singularities of the track of this homotopy consists of finger moves, Whitney moves and cusps. If $\mu(f_i)_1 = \mu(f_i')_1$ for every i then the cusps arising in H can be cancelled in pairs [FQ90, p. 23], leading to a regular homotopy, which as desired is a sequence of isotopies, finger moves, and Whitney moves. □

4. Geometric definitions and operations

4.1. **Conventions regarding Whitney discs.** In this paper, we shall assume that a Whitney disc is a generic immersion $W\colon D^2 \looparrowright M$, which in particular implies that the boundary is embedded, with each Whitney arc lying on one of the sheets whose intersection points paired by W. We allow interior self-intersections of W as well as intersection with other surfaces (and other Whitney discs) but assume that these are transverse, using topological transversality.

The *Whitney section* is a preferred section of the normal bundle of W restricted to its boundary, normal to one sheet, along one of the Whitney arcs, and tangent to the other sheet along the other Whitney arc. We call W *framed* if this section extends to a non-vanishing section of the entire normal bundle $\nu(W)$. Note that if such a section exists then it is unique up to isotopy and indeed defines a framing of $\nu(W)$. The relative normal Euler number of W in $\mathbb{Z}$ is by definition the unique obstruction for the existence of a framing of $\nu(W)$ extending the Whitney section and there are two manoeuvres that can alter it [FQ90, Section 1.3].

- An *interior twist* introduces a single self-intersection to W and the new Whitney disc W' comes with a non-regular homotopy to W that has a single cusp singularity. The relative normal Euler number changes by ± 2, depending on the direction of the twist.
- A *boundary twist* introduces a single intersection between W and one of the sheets its boundary lies on. The new Whitney disc W' comes with a non-regular homotopy to W and the normal Euler number changes by ± 1, depending on the direction of the twist.

Given two Whitney discs, the corresponding Whitney circles may *a priori* intersect one another. We can ensure that Whitney circles are disjoint by pushing one Whitney circle along the other, as shown in Figure 1. This is an isotopy of the Whitney disc, and as we see in the figure, leads to new intersections between one Whitney disc and one of the sheets paired by the other Whitney disc. From now on, we will assume that Whitney circles are pairwise disjoint and embedded.

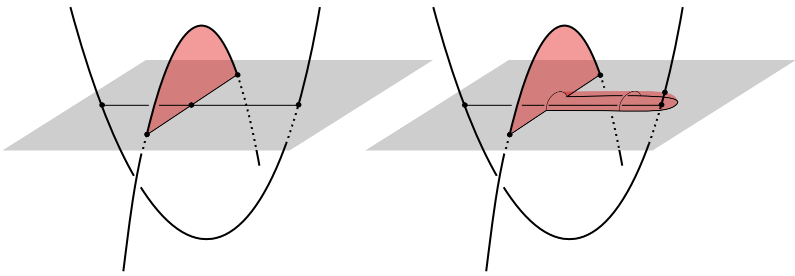


Figure 1. Boundary push-off to ensure Whitney circles are disjoint.

4.2. **Intersection and self-intersection numbers.** Let M be a connected, based 4-manifold, and choose a local orientation of M at its basepoint. Suppose we are given two generic immersions Σ, Σ' from discs or spheres to M, intersecting transversely and only in their interiors, with whiskers $\gamma, \gamma' \colon [0,1] \to M$ joining the base point of M to the base-points of Σ and Σ' respectively, as well as orientations of Σ and Σ'. Then the equivariant intersection number [Wal99, Chapter 5], [FQ90, Section 1.7]

$$\lambda(\Sigma, \Sigma') \in \mathbb{Z}[\pi_1 M]$$

is the sum of contributions $\pm g$ from intersection points between Σ and Σ' with:

- the sign $\pm$ obtained from comparing the local orientation at the intersection point induced by the orientations of Σ and Σ' with the orientation obtained from transporting the orientation at the basepoint of M along γ and along a path in Σ to the intersection point;
- an element g of $\pi_1(M)$ obtained from the concatenation, in order, of γ, a path in Σ to the intersection point, a path in Σ' from the intersection point to the basepoint of Σ', and finally the reverse of γ'.

When $\Sigma = \Sigma'$, set $\lambda(\Sigma, \Sigma) := \lambda(\Sigma, \Sigma^+)$, where Σ^+ is a push-off of Σ along a section of the normal bundle $\nu(\Sigma)$, transverse to the zero section. In the case of a disc, where $\Sigma \colon D^2 \looparrowright M$, this requires a choice of framing of the normal bundle restricted to the boundary. Framings of the trivial 2-dimensional bundle on S^1, up to isotopy, correspond bijectively with $[S^1, SO(2)] \leftrightarrow \mathbb{Z}$. Changing the framing by $n \in \mathbb{Z}$ changes $\lambda(\Sigma, \Sigma^+)$ by $n \cdot 1$ because we may add intersections between Σ and Σ' near the boundary to compensate for the framing change, and these intersections have trivial group element $1 \in \pi_1(M)$.

In this article we use this geometric definition of λ instead of the homological definition, so that we may also apply λ to discs that are not properly immersed, that is whose boundaries do not lie in the boundary of M.

Lemma 4.1. *Let Σ and Σ' be generic immersions from discs or spheres to a connected, based 4-manifold M, equipped with whiskers. Assume that Σ and Σ' intersect transversely and only in their interiors.*

The intersection number $\lambda(\Sigma, \Sigma')$ is unchanged under homotopies that restrict to isotopies on the boundary and that keep the interior of Σ disjoint from $\partial\Sigma'$, and vice versa, at all times. In particular, if both surfaces are spheres, then $\lambda(\Sigma, \Sigma')$ only depends on the equivalence classes $[\Sigma], [\Sigma'] \in \pi_2(M)$.

The requirement that interiors be disjoint from boundaries during the track of the homotopy is necessary, since otherwise one could, for example, push all intersections of Σ with a disc Σ' off the boundary of Σ', assuming the boundary of Σ' is not contained in the boundary of the ambient manifold M.

Proof. We wish to apply Theorem 1.2. Given surfaces Δ, Δ', such that that $\Delta \cup \Delta'$ is homotopic to $\Sigma \cup \Sigma'$ via a homotopy as mentioned in the lemma statement, by applying local cusp homotopies we may assume that the count of double points with trivial group element is zero in the case of both $\Sigma \cup \Sigma'$ and $\Delta \cup \Delta'$. Since cusp homotopies occur on each component separately, they do not affect the value of λ. Now by Theorem 1.2, the surfaces $\Sigma \cup \Sigma'$ and $\Delta \cup \Delta'$ are related by a sequence of isotopies, finger moves, and Whitney moves. Each of these moves preserves λ.

Alternatively, the result follows from the translation into the homological definition of λ in terms of cap products, using the Hurewicz maps $\pi_2(M) \cong H_2(M; \mathbb{Z}[\pi_1 M])$ and $\pi_2(M', \partial M') \to H_2(M', \partial M'; \mathbb{Z}[\pi_1 M])$. As previously mentioned, the general definition in the case of potentially nonorientable or noncompact M involves cohomology with compact support and coefficients twisted using the orientation character $w \colon \pi_1(M) \to \{\pm 1\}$. □

Now we turn to self-intersection numbers. Given the image Σ of a generic immersion of a disc or sphere into M, with a whisker γ, we define the *self-intersection number*

$$\mu(\Sigma) \in \mathbb{Z}[\pi_1 M]/\langle g - w(g)g^{-1} \rangle,$$

as the sum of contributions $\pm g$, summing over the double points of Σ, with $g \in \pi_1(M)$ obtained by the concatenation, in order, of γ, a path in Σ to the intersection point, a path in Σ from the intersection point to the basepoint of Σ, and finally the reverse of γ. The choice of sheet on which to approach a given intersection point gives rise to the indeterminacy $g - w(g)g^{-1}$, where recall that $w \colon \pi_1(M) \to \{\pm 1\}$ is the orientation character of M.

Lemma 4.2. *Let Σ be a generic immersion of a disc or sphere in a connected, based 4-manifold M. The self-intersection number $\mu(\Sigma)$ is unchanged under regular homotopies that restrict to isotopies on the boundary and that keep the interior of Σ disjoint from $\partial\Sigma$ at all times. A non-regular homotopy that runs through a single cusp singularity changes $\mu(\Sigma)$ by ± 1. In particular, if $\Sigma \colon S^2 \looparrowright M$ is a sphere, the* reduced *self-intersection invariant*

$$\widetilde{\mu}(\Sigma) \in \mathbb{Z}[\pi_1 M]/\langle g - w(g)g^{-1}, \mathbb{Z} \cdot 1 \rangle$$

depends only on the homotopy class $[\Sigma] \in \pi_2(M)$.

Proof. The proof is similar to that of Lemma 4.1, except that now cusps become important, in that they change the invariant by ± 1. But this is addressed in the statement. A regular homotopy is by definition a sequence of isotopies, finger moves and Whitney moves, which do not change μ. To see that $\widetilde{\mu}$ only depends on the homotopy class, add cusp homotopies to two representatives Σ and Σ' until $\mu(\Sigma)_1 = \mu(\Sigma')_1 = 0$. By Theorem 1.2, the resulting immersions are regularly homotopic. □

The main property of this self-intersection number lies in the following result.

Lemma 4.3. *Let $f\colon \Sigma \looparrowright M$ be a generic immersion of a disc or sphere. Then $\mu(f) = 0$ if and only if there is a collection of disjointly embedded, framed Whitney discs pairing all the double points of f. If moreover such Whitney discs have interiors disjoint from f, then they guide a sequence of Whitney moves giving rise to a regular homotopy from f to an embedding.*

From this point of view, the reason that $\mu(f) = 0$ does not imply f is regularly homotopic to an embedding is the presence of intersections of the interiors of Whitney discs with f.

Proof. For each double point p of f, we have a choice of first or second sheet in the domain of f, i.e. of a first element in the two element set $f^{-1}(p)$. Given such a choice of sheets at each p, the self-intersection number yields a well-defined sum in $\mathbb{Z}[\pi_1 M]$ that projects to $\mu(f)$. It follows that $\mu(f) = 0$ if and only if we can make choices of sheets in such a way that the double points appear in pairs (p_i^+, p_i^-) with the same group elements but opposite signs. This implies that we can find a mutually disjoint collection of pairs of embedded arcs in Σ, from $f^{-1}(p_i^+)$ to $f^{-1}(p_i^-)$, such that the pair of arcs for a fixed pair $p_i^\pm$ maps to a null homotopic circle under f. These *Whitney circles*, one for each i, are embedded disjointly in the image of f by construction. By construction, they bound a generically immersed collection of Whitney discs in M. Boundary twist to arrange these to be framed. Push the intersections among the Whitney discs across the boundaries, so that the Whitney discs are embedded disjointly in M. □

Next we investigate the relationship between λ and μ. The latter is a *quadratic refinement* of the former. The first relation arises when at least one of the surfaces in question is a sphere. Computing μ of an ambient connected sum, performed along the whiskers yields

$$\mu(\Sigma + \Sigma') = \mu(\Sigma) + \mu(\Sigma') + [\lambda(\Sigma, \Sigma')] \in \mathbb{Z}[\pi_1 M]/\langle g - w(g)g^{-1}\rangle$$

if Σ, Σ' is a generically immersed pair consisting of two spheres or one disc and one sphere. To prove this relation, count the intersections that arise when using connecting tubes that miss the interiors.

Lemma 4.4. *Let Σ be a generic immersion of either a disc with framed boundary or a sphere, in a 4-manifold M. Then*

$$\lambda(\Sigma, \Sigma) = \mu(\Sigma) + \overline{\mu(\Sigma)} + e(\Sigma) \cdot 1,$$

where the involution on $\mathbb{Z}[\pi_1 M]$ is defined by $\overline{g} := w(g)g^{-1} \in \mathbb{Z}[\pi_1 M]$ and $e(\Sigma)$ is the (relative) Euler number of the normal bundle $\nu(\Sigma)$.

Proof. Each double point of Σ contributing $\pm g$ to $\mu(\Sigma)$ gives rise to two intersection points between Σ and a push-off Σ^+ along a generic section of the normal bundle, that is transverse to the zero section. One of these intersections contributes $\pm g$ to $\lambda(\Sigma, \Sigma)$ while the other contributes $\pm w(g)g^{-1}$. The remaining intersections between Σ and Σ^+ come from the zeros of the generic section. They contribute exactly the Euler number times the trivial group element $1 \in \pi_1(M)$. □

Note that in the case of a disc, $\Sigma\colon D^2 \looparrowright M$, the left hand side of the above formula depends on the framing on the boundary and so does the normal Euler number on the right hand side. The case of a sphere is quite different. The left hand side depends only on the homotopy class $[\Sigma] \in \pi_2(M)$, whereas all terms on the right hand side depend on the generic immersion $\Sigma\colon S^2 \looparrowright M$.

The next result gives a general circumstance under which μ is redundant information.

Corollary 4.5. *Let Σ be a generic immersion as in Lemma 4.4 but with vanishing (relative) normal Euler number, $e(\Sigma) = 0$. If the orientation character w vanishes on all order two elements of $\pi_1(M)$ then $\lambda(\Sigma, \Sigma)$ uniquely determines $\mu(\Sigma)$ by the formula in Lemma 4.4.*

Proof. Consider the involution $g \mapsto g^{-1}$ on $\pi_1(M)$ and denote the set of orbits by $[\pi_1(M)]$. Since $\mu(\Sigma)$ takes values in $\mathbb{Z}[\pi_1 M]/\langle g - w(g)g^{-1}\rangle$ which is a free abelian group on the set $[\pi_1(M)]$, we may write

$$\mu(\Sigma) = \sum_{[g]\in[\pi_1(M)]} n_{[g]}[g], \text{ with unique } n_{[g]} \in \mathbb{Z}.$$

Choose a representative $r[g] \in \pi_1(M)$ of each orbit $[g]$. Hence for each $g \in \pi_1(M)$ we have $r[g] = g$ or $r[g] = g^{-1}$. Then in $\mathbb{Z}[\pi_1 M]$ we have

$$\begin{aligned}\sum_{g\in\pi_1(M)} a_g g := \lambda(\Sigma,\Sigma) = \mu(\Sigma) + \overline{\mu(\Sigma)} &= \sum_{[g]\in[\pi_1(M)]} n_{[g]} r[g] + n_{[g]} w(g) r[g]^{-1} \\ &= \sum_{[g]\in[\pi_1(M)], g\neq g^{-1}} n_{[g]} r[g] + n_{[g]} w(g) r[g]^{-1} + \sum_{g=g^{-1}\in\pi_1(M)} n_{[g]} g + n_{[g]} w(g) g.\end{aligned}$$

Comparing coefficients, we see that for $g \neq g^{-1}$, we have either $n_{[g]} = a_g$ or $n_{[g]} = a_{g^{-1}}$, depending on whether $r[g] = g$ or $r[g] = g^{-1}$. Using that $w(g) = 1$ if $g = g^{-1}$, we also see that in this case $2n_{[g]} = a_g$, which shows that the $n_{[g]}$ are determined by the a_g for all $[g] \in [\pi_1(M)]$. □

On the other hand, if $w(g) = -1$ and $g^2 = 1$, the last formula shows that $n_{[g]}$ is not determined by a_g, which in this case has to be zero. So the hypothesis of the corollary cannot be removed.

4.3. **Gropes and towers.** In our proofs, we will seek to construct *gropes* and ultimately *towers*, as defined in [BKKPR, Chapter 11] (see also [FQ90, Chapters 2 and 3]). We remark that our definitions match the ones in [BKKPR] but the definition of tower is slightly different from that in [FQ90]. We refer the reader to [BKKPR, Chapter 11] for the complete definitions, but we give a quick summary below.

Gropes and towers are 4-dimensional objects, obtained as thickenings of underlying 2-complexes. When we describe operations on embedded or immersed gropes and towers, such as pushing, tubing, and contracting, we often do so in terms of operations on the 2-complex. Provided that framings are controlled carefully, these operations naturally extend to the 4-dimensional thickenings, and it is these extended operations that we really use.

To define gropes and towers, first we consider model stages. A *surface block* is $\Sigma \times D^2$, where Σ is a compact orientable surface with a single boundary component. A *disc block* is a self-plumbed handle $D^2 \times D^2$, with algebraically zero count of double points. A *stage* is a union of blocks of the same type, that is, a stage is a union of disc blocks or a union of surface blocks. A *generalised tower* $\mathcal{T}$ is a 4-manifold with boundary, which is diffeomorphic to a boundary connected sum of copies of $S^1 \times D^3$, along with the following data.

(i) A decomposition of $\mathcal{T}$ into stages.
(ii) The *attaching region* $\partial_-\mathcal{T}$ of $\mathcal{T}$, which is the image of a given embedding $\phi\colon S^1 \times D^2 \hookrightarrow \partial\mathcal{T}$. The core of the attaching region is called the *attaching circle*.
(iii) *Tip regions*, namely the image of a given embedding $\psi\colon \bigsqcup^i S^1 \times D^2 \hookrightarrow \partial\mathcal{T}$.

The tip regions of a tower may be identified with attaching regions of other towers to form larger towers.

A generalised tower $\mathcal{T}$ is said to be properly embedded in a 4-manifold M is there is a map $(\mathcal{T}, \partial_-\mathcal{T}) \hookrightarrow (M, \partial M)$.

The objects that we build are all special cases of generalised towers, as we now describe. In what follows, tip regions, to which we attach higher stages, are always obtained by taking certain described curves on the underlying 2-complex, pushing them to the boundary, and then thickening.

(1) A *grope* is a generalised tower consisting of only surface stages, each attached to a full symplectic basis of curves for H_1 of the surfaces Σ underlying each block of the previous stage (after pushing the curves to the boundary and thickening). The number of surface stages is called the *height* of the grope. A *sphere-like grope* is a grope where the bottommost stage is a single connected surface with no boundary.
(2) A *capped grope* is a grope with a disc stage attached to the final surface stage, after further plumbings and self-plumbings among the constituent blocks of the final stage. These disc blocks, or often, their underlying discs, are called the *caps* of the capped grope. Note that the caps may intersect themselves and one another arbitrarily. A capped grope with precisely one surface stage will often be referred to as a *capped surface*. A capped grope is *union-of-discs-like*, *union-of-spheres-like*, *sphere-like*, or *disc-like*, according to the topology of the immersed surface obtained by totally contracting (Section 4.4 defines contraction).
(3) A *tower* consists of a sequence of capped gropes, called the *storeys* of the tower, each attached to a (thickening of a) full collection of double point loops (i.e. tip regions) associated with the plumbings of the caps of the previous capped grope. The caps of the constituent capped gropes are only allowed to have self-plumbings, which must further be algebraically cancelling, i.e. the caps form a disc stage.
(4) A *capped tower* is obtained by adding one further disc stage to a tower, along (a thickening of) a full collection of double point loops for the caps of the top storey, followed by further plumbings and self-plumbings among the constituent blocks of the final stage. As before, the discs in the final stage are called the *caps* of the capped tower, and the caps may intersect themselves or one another arbitrarily.

For generalised towers $\mathcal{A}$ and $\mathcal{B}$ such that $\mathcal{A} \subset \mathcal{B}$ and $m \geqslant 1$, we say that the first m stages of $\mathcal{A}$ and $\mathcal{B}$ coincide, or are the same, if the underlying 2-complexes of the first m stages of $\mathcal{A}$ and $\mathcal{B}$ coincide, the inclusion of the first m stages of $\mathcal{A}$ into $\mathcal{B}$ is isotopic to a diffeomorphism with image the first m stages of $\mathcal{B}$, and the attaching regions of $\mathcal{A}$ and $\mathcal{B}$ coincide. If $\mathcal{A}$ and $\mathcal{B}$ are both capped gropes or towers with height m and the first m stages coincide, we say that $\mathcal{A}$ and $\mathcal{B}$ have the same body.

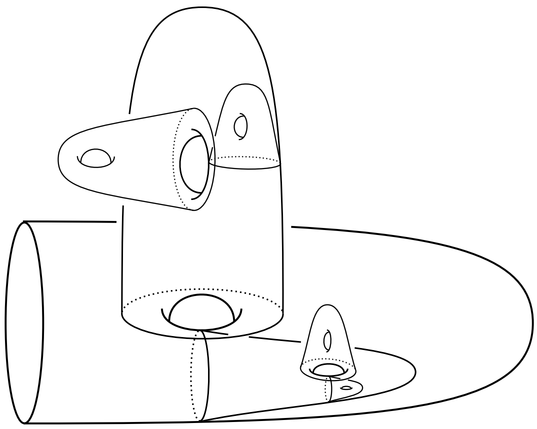


Figure 2. Schematic picture of the 2-dimensional spine of a height three grope.

Within our proofs, at intermediate steps, we will construct generalised towers that do not fit into any of these categories. For example a capped grope obtained by attaching a capped grope of height n to a half-basis of curves for H_1 of a surface Σ, and attaching a capped grope of height m to the symplectically dual basis.

We will also need the following generalisation of the notion of dual spheres in the introduction.

Definition 4.6 (Dual capped gropes and surfaces)**.** Let $\{A_i\}$ be a collection of immersed discs, gropes, or towers, with or without caps. A collection $\{T_i^c\}$ of (sphere-like) capped

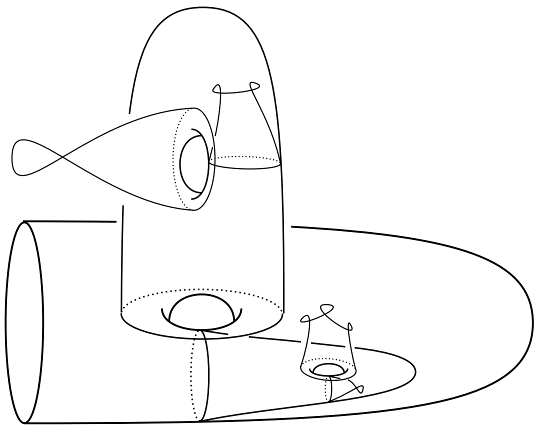


Figure 3. Schematic picture of the 2-dimensional spine of a height two capped grope.

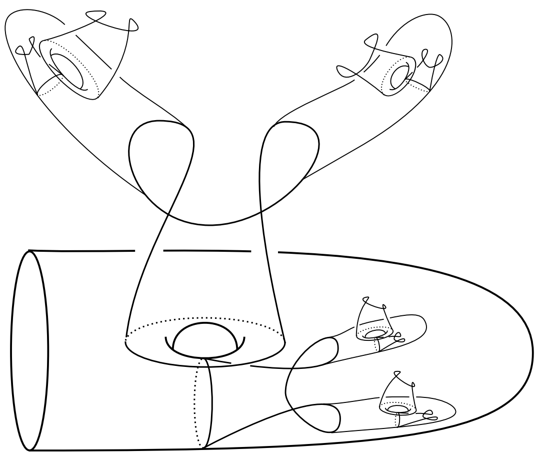


Figure 4. Schematic picture of the 2-dimensional spine of a 2-storey tower with one surface stage in each storey.

gropes is said to be (geometrically) *dual* to $\{A_i\}$ if $T_i^c \pitchfork A_j$ is a single transverse point when $i = j$, located in the bottom stages of T_i^c and A_i, and $T_i^c \pitchfork A_j$ is empty otherwise. For $\{A_i\}$ a generically immersed collection of discs, the *bottom stage* of A_i is simply itself. Note that all the caps of $\{A_i\}$ are required to be disjoint from the caps of $\{T_i^c\}$. Additionally, note that intersections are allowed within the collection $\{T_i^c\}$. If each T_i^c is a capped grope of height one, then we say that $\{T_i^c\}$ is a collection of *dual capped surfaces* for $\{A_i\}$.

4.4. **Contraction.** The process of *contraction and push-off* is introduced in [FQ90, Section 2.3] and [BKKPR, Section 14.1]. *Contraction*, sometimes called *symmetric surgery*, of a capped surface, depicted in Figure 5, converts a capped surface into an immersed disc. As shown by the figure, we start with a symplectic basis of curves on the surface and surger the surface using two parallel copies each cap. One could only surger using one disc per dual pair, but this would not enable the pushing off procedure that we are about to describe in the next paragraph. Given a capped grope, we can iteratively contract caps, to eventually obtain a collection of immersed spheres or discs called the *total contraction*.

After contracting a surface, any other surface that intersected the caps can be pushed off the contraction, as shown in Figure 5. This reduces the number of intersection points between the resulting contraction and the pushed off surfaces. In other words, we gain some disjointness at the expense of converting a capped surface into an immersed disc. An

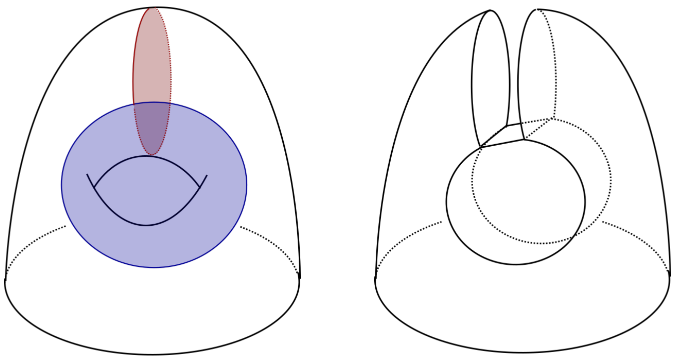
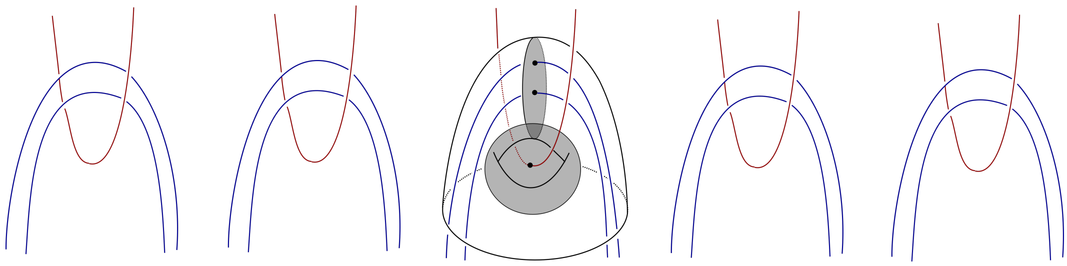
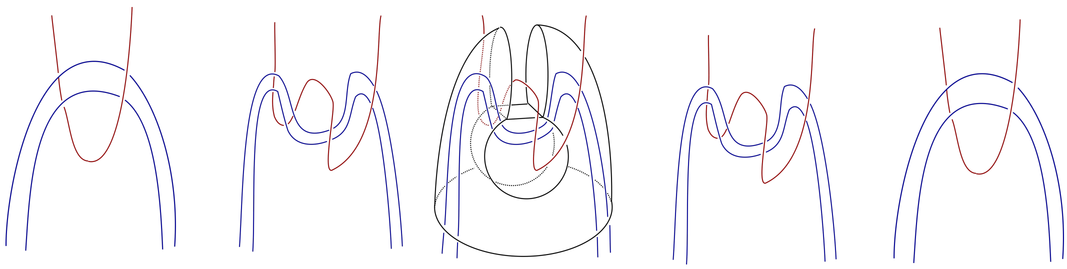


Figure 5. (Symmetric) contraction of a capped surface. Here we show the situation for embedded caps.

additional cost is as follows. Suppose that a surface A intersects a cap of the capped surface, and a surface B intersects a dual cap. Then after pushing both A and B off the contraction, we obtain two intersection points between A and B. The contraction push-off operation is shown in Figure 6.



(a) Before contraction of a surface.



(b) After contraction of a surface, and pushing other surfaces off the caps.

Figure 6. Contraction and push-off. Note the intersections of pushed-off surfaces that occur between diagrams one and two and between diagrams four and five in the bottom row of figures, namely one intersection in the past and one intersection in the future between each pair of surfaces pushed off dual caps.

Lemma 4.7. *The homotopy class of the sphere or disc resulting from symmetric contraction of a fixed surface is independent of the choice of caps, provided the boundaries of the different choices of caps coincide.*

Proof. As explained in [FQ90, Section 2.3], an isotopy in the model induces a homotopy of the immersed models, so the symmetric contraction is homotopic to the result of surgery along one cap per dual pair. Now let $\{C_i, D_i\}_{i=1}^g$ and $\{C_i', D_i'\}_{i=1}^g$ be two sets of caps for a surface of genus g, such that $\partial C_i = \partial C_i'$ and $\partial D_i = \partial D_i'$ are a dual pair of curves on the surface for each i. Then symmetric surgery on $\{C_i, D_i\}_{i=1}^g$ is homotopic to surgery on the $\{C_i\}$, which is homotopic to symmetric surgery on $\{C_i, D_i'\}_{i=1}^g$. This is homotopic to asymmetric surgery on the caps $\{D_i'\}$, which finally is homotopic to the result of symmetric surgery on $\{C_i', D_i'\}_{i=1}^g$, as asserted. □

5. Further ingredients

We will also need to draft in the following results from [Fre82, FQ90]; see also [BKKPR]. The next lemma trades intersections between distinct surfaces for self-intersections.

Lemma 5.1 (Geometric Casson lemma). *Let F and G be transverse generic immersions of compact surfaces in a connected 4-manifold M. Assume that the intersection points $\{p, q\} \subset F \pitchfork G$ are paired by a Whitney disc W. Then there is a regular homotopy from $F \cup G$ to $\overline{F} \cup \overline{G}$ such that $\overline{F} \pitchfork \overline{G} = (F \pitchfork G) \smallsetminus \{p, q\}$, that is the two paired intersections have been removed. The regular homotopy may create many new self-intersections of F and G; however, these are algebraically cancelling.*

Applications of this lemma, proven inductively on the number of intersection points, include the following.

(i) Making F and G disjoint, assuming that all intersection points $F \cap G$ are paired by Whitney discs.
(ii) Turning a collection G of algebraically dual spheres for F into geometrically dual spheres $\overline{G}$.

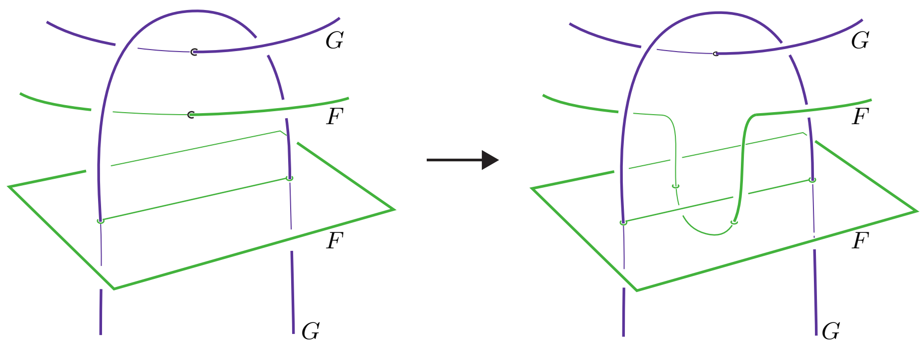


Figure 7. Pushing down a sheet of F into itself.

Proof of Lemma 5.1. First add boundary twists to make W framed. Push each intersection point of F with the interior of W down into F and each intersection point of G with W down into G, as shown in Figure 7. This uses that W pairs intersections between F and G and is a regular homotopy of $F \cup G$, since it is a sequence of finger moves. It introduces many new self-intersections of F and of G. However, doing the immersed Whitney move, another regular homotopy, on the new Whitney disc W' removes the pair of points $\{p, q\}$ as required, while introducing yet more algebraically cancelling intersection points of F or G. □

Proposition 5.2 (Grope height raising). *Given a capped grope $G(m)^c$ of height $m \in \frac{1}{2}\mathbb{Z}$, where m is at least 1.5, and a positive integer $n \geqslant m$, there exists a capped grope $G(n)^c$ of height n embedded in $G(m)^c$ such that the first m stages of $G^c(n)$ and $G^c(m)$ coincide.*

Proof. See [FQ90, Section 2.7] or [BKKPR, Chapter 16] for a proof. □

Lemma 5.3 (Sequential contraction lemma)**.** *Let $G(m+1)^c$ be a height $(m+1)$ capped grope. There exists an embedding of a height m capped grope $G(m)^c$ into $G(m+1)^c$, where all the caps of $G(m)^c$ are mutually disjoint and have algebraically cancelling double points, such that the body of $G(m)^c$ coincides with the first m stages of $G(m+1)^c$.*

Proof. Enumerate the top stage surfaces of $G(m+1)^c$. Iteratively, contract the ith surface, and push the caps of the surfaces numbered greater than i off the contraction. At the end of this procedure, we obtain the desired $G(m)^c$. □

An infinite tower, constructed as a colimit of finite towers, will be denoted $\mathcal{T}_\infty$. We denote the *Freudenthal endpoint compactification* (see [FQ90, Section 3.8] or [BKKPR, Section 11.2]) of such a tower by $\widehat{\mathcal{T}}_\infty$.

Theorem 5.4 (Tower embedding theorem)**.** *Let $\mathcal{T}_1^c$ be a 1-storey capped tower with at least four surface stages. Within $\mathcal{T}_1^c$ there exists an embedding of an infinite compactified tower $\widehat{\mathcal{T}}_\infty$ with the same attaching region as $\mathcal{T}_1^c$, such that the following holds.*

1. *(Replicable) Each storey of $\widehat{\mathcal{T}}_\infty$ has at least four surface stages.*
2. *(Boundary shrinkable) For each connected component of $\widehat{\mathcal{T}}_\infty$, the series $\sum_{j=1}^{\infty} N_j/2^j$ diverges, where N_j is the number of surface stages in the jth storey of the component.*
3. *(Squeezable) For each $n \geqslant 3$, the connected components of the nth storey of $\widehat{\mathcal{T}}_\infty$ lie in arbitrarily small mutually disjoint balls.*

Proof. See [FQ90, Chapter 3] or [BKKPR, Chapter 17] for a proof. □

The names for the properties in Theorem 5.4 are deliberately evocative: a replicable infinite tower contains another such compactified infinite tower within any two consecutive storeys, while a boundary shrinkable tower has boundary homeomorphic to S^3. See [FQ90] or [BKKPR] for more details.

Theorem 5.5. *Every connected compactified infinite tower $\widehat{\mathcal{T}}_\infty$ satisfying the three properties in Theorem 5.4 is homeomorphic, relative to its attaching region $\partial_-\widehat{\mathcal{T}}_\infty$, to the standard handle $(D^2 \times D^2, S^1 \times D^2)$.*

Proof. This is proven in [FQ90, Chapter 4] and [BKKPR, Part IV]. □

Remark 5.6. Combining Theorems 5.4 and 5.5, we see that any disc-like 1-storey capped tower $\mathcal{T}_1^c$ with at least four surface stages contains an embedded disc whose framed boundary coincides with the attaching region of $\mathcal{T}_1^c$.

Definition 5.7 (Good group [FT95b])**.** A group Γ is said to be *good* if for every height 1.5 disc-like capped grope G and for $\phi\colon \pi_1(G) \to \Gamma$ a group homomorphism, there exists an immersed disc $D \looparrowright G$ whose framed boundary coincides with the attaching region of G, such that the double point loops of D, considered as fundamental group elements by making some choice of basing path, are mapped to the identity element of Γ by ϕ.

It is easy to see by contraction that the attaching circle of a capped grope G is null-homotopic in G. Thus, there is always some immersed disc in G with the same framed boundary as G. For a good group Γ we get further control on the double points of such an immersed disc.

Since every homomorphism to the trivial group is trivial, the trivial group is good. We also know that groups of subexponential growth are good [FT95a, KQ00], and that the class of good groups is closed under subgroups, quotients, extensions, and colimits [FQ90, p. 44], [BKKPR, Chapter 18]. In particular, all finite groups, abelian groups, and solvable groups are good.

6. Constructing capped gropes and towers with dual spheres

Our main technical lemma, given below, shows how to upgrade a collection of immersed discs with certain types of dual capped surfaces to a collection of gropes whose attaching regions coincide with the framed boundary of the original discs, as well as the same dual capped surfaces. The requirement on the dual capped surfaces is that the bodies be located close to the boundary of the ambient manifold. This property enables us to find Whitney discs which do not intersect the bodies, at various steps of the argument.

Recall that given two generically immersed transversely intersecting oriented discs C_1 and C_2 in M, if $\lambda(C_1, C_2) = 0$, all the intersection points between C_1 and C_2 may be paired up by immersed Whitney discs in M. If $\lambda(C_1, C_1) = \mu(C_1) = 0$, then the self-intersection points of C_1 may be paired similarly.

Lemma 6.1. *Consider a generically immersed collection of discs in a 4-manifold M*

$$F = (f_1, \ldots, f_k) \colon (D^2 \sqcup \cdots \sqcup D^2, S^1 \sqcup \cdots \sqcup S^1) \looparrowright (M, \partial M).$$

Suppose that $\{\Sigma_i^c\}$ is a dual collection of capped surfaces for the $\{f_i\}$ such that

$$\lambda(C_\ell, C_m) = \mu(C_\ell) = 0$$

for every pair of caps C_ℓ and C_m of $\{\Sigma_i^c\}$. Assume in addition that the body Σ_i of each Σ_i^c is contained in a collar neighbourhood of ∂M.

Then there exists a collection of mutually disjoint capped gropes $\{G_i^c\}$, properly embedded in M, with arbitrarily many surface stages and pairwise disjoint caps with algebraically cancelling double points, such that $\{\Sigma_i^c\}$, after a regular homotopy of the caps, provides a dual collection of capped surfaces for the $\{G_i^c\}$. Moreover, for each i, the attaching region of G_i^c coincides with the framed boundary of f_i.

In the upcoming proof, the capped surfaces will act as dual sphere factories. That is, we will use them to produce a dual sphere for an f_i whenever we need one, taking advantage of the extra power of a dual capped surface versus a dual sphere.

Proof. Convert each f_i to a capped surface by tubing all the intersections among the $\{f_i\}$, including self-intersections, into parallel copies of the dual surfaces $\{\Sigma_i\}$, as shown in Figure 8. This produces surfaces $\{f_i'\}$ with the same framed boundaries as $\{f_i\}$. Since $\{\Sigma_i\}$ is a mutually disjoint collection of embedded surfaces, so is $\{f_i'\}$. Use parallel copies of the caps $\{C_\ell\}$ of the $\{\Sigma_i\}$ to obtain caps $\{C_n'\}$ for the surfaces $\{f_i'\}$. Note that the interiors of the caps $\{C_n'\}$ are disjoint from the surfaces $\{f_i'\}$. Such a capped surface is shown in Figure 9.

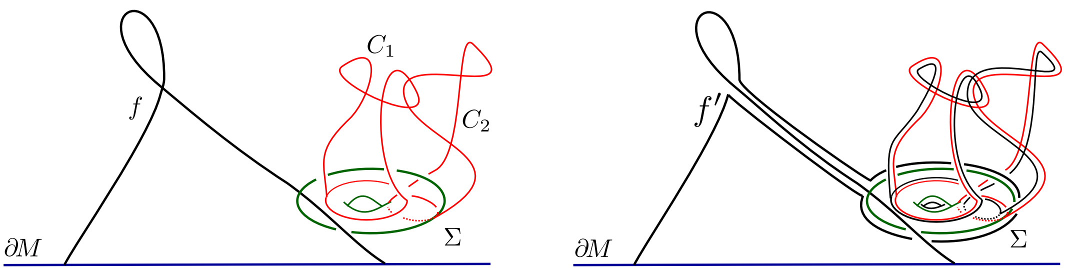


Figure 8. Producing capped surfaces. The self-intersection of the disc f (black) is tubed into the dual capped surface Σ (green). This disc is thus converted into a capped surface, when endowed with parallel copies of the caps for Σ.

Now we will separate the caps $\{C_\ell\}$ of $\{\Sigma_i\}$ from the caps $\{C_n'\}$ for the surfaces $\{f_i'\}$. For this, we will need a dual sphere for each f_i', so we turn to our dual sphere factory: for each i, take a parallel copy of Σ_i along with its caps, and contract to obtain a sphere S_i (we do not push anything off the contraction), as shown in Figure 9. Since $\{\Sigma_i\}$ is a geometrically dual collection of capped surfaces for the $\{f_i\}$, we see that $\{S_i\}$ is a geometrically dual collection of framed immersed spheres for the $\{f_i'\}$. Here we are using the fact that the caps $\{C_\ell\}$ and the surfaces $\{f_i'\}$ do not intersect. However note that the spheres $\{S_i\}$ do have intersections among themselves, coming from the intersections among the caps $\{C_\ell\}$, and they also intersect the caps of the surfaces $\{f_i'\}$ and the caps of the $\{\Sigma_i\}$.

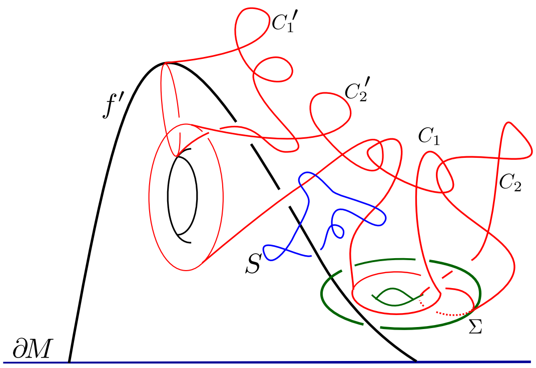


Figure 9. A single capped surface f', whose caps intersect the caps of Σ. The dual sphere S for f' is produced by contraction of Σ.

Recall that $\lambda(C_\ell, C_m) = \mu(C_\ell) = 0$ for every pair of caps C_ℓ and C_m of the dual surfaces $\{\Sigma_i\}$. Since each cap C_n' for some f_i' is obtained as a parallel copy of a cap for some Σ_i, we see that $\lambda(C', C) = 0$ for each cap C' for some f_i' and for each cap C for some Σ_i. Thus the intersection points between C and C' are paired by framed Whitney discs. Let D' be such a Whitney disc. We shall assume that $D' \cap \Sigma_i = \varnothing$ for all i, since D' can be pushed off the collar neighbourhood of the boundary of M containing $\{\Sigma_i\}$. For each intersection of D' with f_i', tube D' into S_i. We obtain a new Whitney disc, which we still call D', having possible intersections only with the caps of $\{\Sigma_i\}$ and the caps of $\{f_i'\}$ (as well as self-intersections).

Now, for each Whitney disc D', by the geometric Casson lemma (Lemma 5.1) there is a regular homotopy of the caps making the interior of D' disjoint from both types of caps at the expense of introducing new (algebraically cancelling) cap intersections among the $\{C_\ell\}$ and among the $\{C_n'\}$. These last two steps are illustrated in Figure 10.

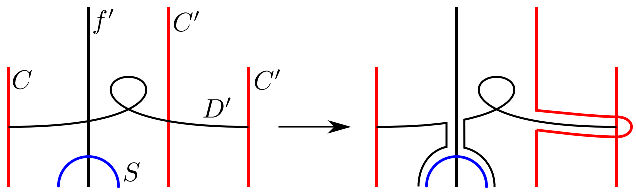


Figure 10. Separating the caps of $\{\Sigma_i\}$ and $\{f_i'\}$.

A subsequent (immersed) Whitney move over D' ensures that C' and C are disjoint. Repeat this process for all the Whitney discs D' to ensure that the caps $\{C_\ell\}$ of the $\{\Sigma_i\}$ are pairwise disjoint from the caps $\{C_n'\}$ for the surfaces $\{f_i'\}$, as shown in Figure 11.

Depending on whether the Whitney move was performed on C or C', there are new intersections between C and $\{S_i\}$, or C' and $\{S_i\}$, due to the intersections and self-intersections of $\{S_i\}$. But we shall forget the $\{S_i\}$ from now on anyway.

Now push the intersections and self-intersections of the caps $\{C_n'\}$ for the surfaces $\{f_i'\}$ down into the surface and then tube into parallel copies of the dual surfaces $\{\Sigma_i\}$, as shown in Figure 12, to promote the collection $\{f_i'\}$ to height two gropes $\{f_i''\}$. Parallel copies of the

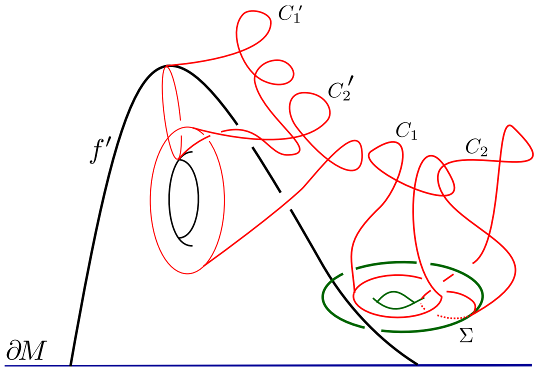


Figure 11. After separating the caps of $\{\Sigma_i\}$ and $\{f_i'\}$.

caps $\{C_\ell\}$ of the $\{\Sigma_i\}$ provide caps $\{C_p''\}$ for the $\{f_i''\}$. Observe that f_i'' has the same framed attaching region as f_i.

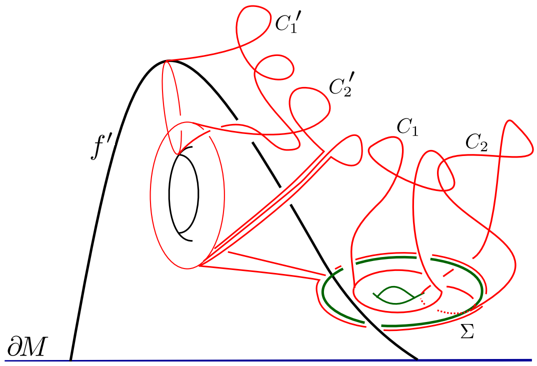


Figure 12. Push down the cap intersections of f' and tube into the dual capped surface.

Remark 6.2. We will soon apply grope height raising to each member of the collection $\{f_i''\}$ (using their caps), and use the good fundamental group hypothesis to upgrade to a collection of 1-storey capped towers. Some extra care will be needed, as we now describe, since we want to produce dual spheres as well. As noted earlier, the $\{S_i\}$ have some unknown number of geometric intersections with caps of $\{f_i'\}$ or caps of $\{\Sigma_i\}$. By the construction of $\{f_i''\}$, in either case, there is now an unknown number of geometric intersections among the members of $\{f_i''\}$ and $\{S_i\}$. In other words, $\{S_i\}$ is not a geometric dual collection for the gropes $\{f_i''\}$. By our construction, the collection $\{S_i\}$ is *algebraically* dual to the collection of gropes $\{f_i''\}$. However, our usual technique for upgrading algebraic duals to geometric duals, namely the geometric Casson lemma, produces intersections within the family $\{f_i''\}$, and these are undesirable because we wish to construct embedded capped gropes. This is why we need a dual sphere factory (namely the dual surfaces $\{\Sigma_i\}$ along with their caps $\{C_\ell\}$) rather than a single dual sphere – at multiple points of the proof we produce a dual sphere and then use it up, so we need the ability to find a new one each time.

With the remark above in mind, the next step is to separate our dual sphere factory from the gropes $\{f_i''\}$. To achieve this, we only need to ensure that there are no intersections between the caps $\{C_\ell\}$ of the $\{\Sigma_i\}$ and the caps $\{C_p''\}$ of the gropes $\{f_i''\}$. As before, for each i, take a parallel copy of Σ_i along with its caps, and contract to obtain a sphere S_i', as shown in Figure 13 (as before, do not push anything off the contraction). Observe as above that

$\{S'_i\}$ is a geometrically dual collection of spheres for $\{f''_i\}$. Here we are using the fact that the caps $\{C_\ell\}$ and the gropes $\{f''_i\}$ do not intersect. However note that the spheres $\{S'_i\}$ do have intersections among themselves, coming from the intersections among the caps $\{C_\ell\}$, and they do intersect the caps of the gropes $\{f''_i\}$ and the caps of the $\{\Sigma_i\}$.

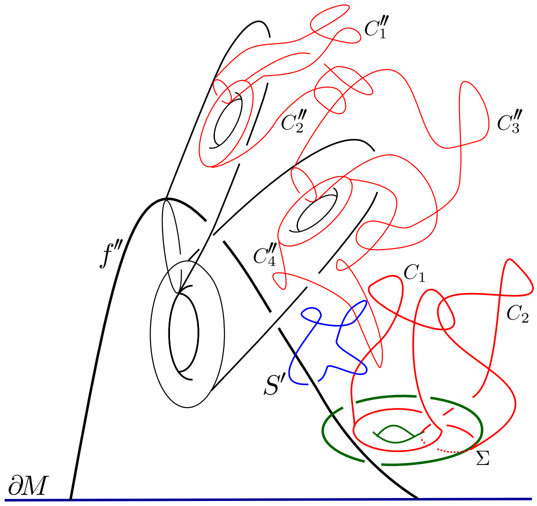


Figure 13. A dual sphere S' for f'' has been produced.

Since the grope caps for $\{f''_i\}$ arose as parallel copies of the caps for the dual surfaces $\{\Sigma_i\}$, we see that $\lambda(C'', C) = 0$ for every cap C'' for some f''_i and every cap C for some Σ_i. Thus the intersection points between C and C'' are paired by Whitney discs. Let D'' be such a Whitney disc. As before, $D'' \cap \Sigma_i = \varnothing$ for all i. For each intersection of D'' with some f''_i, tube D'' into the dual sphere S'_i. We obtain a new Whitney disc, which we still call D'', having possible intersections only with the caps of $\{\Sigma_i\}$ and the caps of $\{f''_i\}$. Do this for all the Whitney discs $\{D''_t\}$, so they are all disjoint from the $\{f''_i\}$.

Now, for each Whitney disc D', by the geometric Casson lemma (Lemma 5.1) there is a regular homotopy making the interior of D'' disjoint from both types of caps, at the expense of creating more cap intersections among the $\{C_\ell\}$ and among the $\{C''_p\}$ i.e. push the caps off D'' over the correct part of the boundary of D''. A subsequent (immersed) Whitney move over D'' reduces the number of intersections between the collections of grope caps $\{C''_p\}$ for $\{f''_i\}$ and the caps $\{C_\ell\}$ for the dual surfaces $\{\Sigma_i\}$. Repeat this process for all the Whitney discs D'' pairing intersection points of C and C'', to ensure that the grope caps for $\{f''_i\}$ and the caps $\{C_\ell\}$ for the dual surfaces $\{\Sigma_i\}$ are pairwise disjoint.

Apply grope height raising (Proposition 5.2) followed by the sequential contraction lemma (Lemma 5.3) to the capped gropes $\{f''^c_i\}$, where each f''^c_i is the union of f''_i with its latest caps. We obtain a collection of capped gropes $\{G^c_i\}$ with the same framed attaching region as the $\{f''_i\}$, arbitrarily many surface stages, and pairwise disjoint caps with algebraically cancelling double points, as shown in Figure 14. Additionally, the surfaces $\{\Sigma_i\}$ are geometrically dual to $\{G^c_i\}$. Moreover, the first two stages of each G^c_i coincide with the first two stages of f''^c_i, for each i, and the caps of $\{G^c_i\}$ do not intersect the caps of $\{\Sigma_i\}$, since the caps $\{C''_p\}$ for $\{f''_i\}$ and the caps $\{C_\ell\}$ for $\{\Sigma_i\}$ are pairwise disjoint. This completes the proof of Lemma 6.1. □

Next, we show that a collection of capped gropes with certain types of dual capped surfaces, such as those produced by the previous lemma, can be replaced by a collection of 1-storey capped towers, with the same framed attaching region as the original capped gropes, and with geometrically dual spheres. The following lemma is the only point in this paper that uses the hypothesis that the fundamental group of the ambient manifold be good.

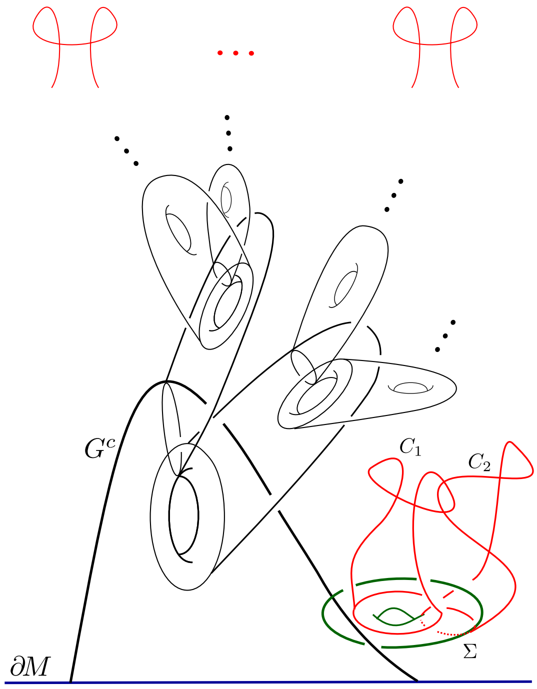


Figure 14. The result of grope height raising followed by the sequential contraction lemma. The caps of Σ may have changed by a regular homotopy of the interior.

Recall that, roughly speaking, a 1-storey capped tower can be built from a capped grope, all of whose cap intersections are self-intersections, by adding a second layer of caps to the double point loops of the caps of the grope i.e. to the tip regions on the disc stage of the capped grope.

Lemma 6.3. *Let M be a connected 4-manifold with $\pi_1(M)$ good. Let n be a non-negative integer. Let $\{G_i^c\}$ be a collection of capped gropes with height $n + 2.5$ and mutually disjoint caps, properly embedded in M, with a geometrically dual collection of capped surfaces $\{\Sigma_i^c\}$, such that the body Σ_i of each Σ_i^c is contained in a collar neighbourhood of ∂M.*

Then there exists a collection of 1-storey capped towers $\{\mathcal{T}_i^c\}$, properly embedded in M, where the first storey grope has height n, with a geometrically dual collection of spheres $\{R_i\}$, such that $\mathcal{T}_i^c$ and G_i^c have the same attaching region for each i. Moreover, the first n surface stages of G_i^c and $\mathcal{T}_i^c$ coincide and each R_i is obtained from Σ_i^c by contraction.

Proof. Consider the union of the top 1.5 stages of the gropes $\{G_i^c\}$. Since $\pi_1(M)$ is good and the caps are mutually disjoint, each component contains an immersed disc whose double point loops are null homotopic in M, and whose framed boundary coincides with the attaching region of the old top 1.5 stages. Attach these discs to the lower stages, producing capped gropes $\{\widetilde{G}_i^c\}$ of height $n+1$. Contract the top stage to obtain capped gropes of height n, whose double point loops are still null-homotopic in M, with caps still mutually disjoint, and such that the caps have algebraically cancelling self-intersection points, since they arose from a symmetric contraction. Here we are using the fact that the new double point loops are parallel push-offs of the previous double point loops (see, for example, [BKKPR, Chapter 19]).

Null homotopies for the double point loops produce immersed discs $\{\widetilde{\delta}_\alpha\}$ bounded by the double point loops of the new caps. Note that the new discs may be assumed to not intersect $\{\Sigma_i\}$, since the collection $\{\Sigma_i\}$ is located close to the boundary of M, and any null homotopy may pushed off a collar of ∂M. On the other hand the discs coming from the null homotopies might intersect $\{\widetilde{G}_i^c\}$ arbitrarily, as shown in Figure 15.

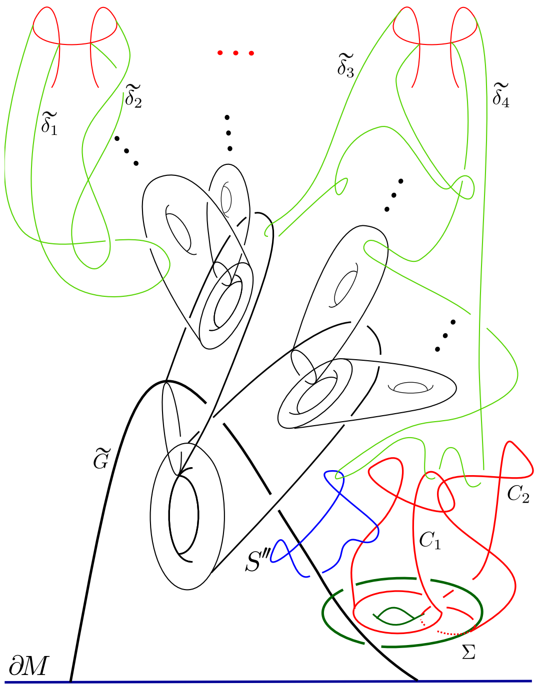


Figure 15. Obtaining discs (light green) from null homotopies of the double point loops. We have produced a dual sphere S'' (blue) for $\widetilde{G}$ (black).

We use our dual sphere factory as before: take a parallel copy of each Σ_i along with its caps and contract, to produce a family of spheres $\{S_i''\}$, as shown in Figure 15 (do not push anything off the contraction). Boundary twist the discs $\{\widetilde{\delta}_\alpha\}$ to achieve the correct framing and then push down and tube into $\{S_i''\}$ to remove any intersections of the resulting discs with $\{\widetilde{G}_i^c\}$. Glue the resulting discs $\{\delta_\alpha\}$ to $\{\widetilde{G}_i^c\}$ to produce the 1-storey capped towers $\{\mathcal{T}_i^c\}$.

Note that at this point the caps for $\{\Sigma_i\}$ only (possibly) intersect $\{\mathcal{T}_i^c\}$ in the tower caps. Destroy the dual sphere factories: contract each Σ_i along its caps and call this family of spheres $\{R_i\}$. Push all intersections with tower caps off the contraction. This produces more intersections among the tower caps of $\{\mathcal{T}_i^c\}$. The family of dual spheres $\{R_i\}$ is then geometrically dual to the resulting 1-storey capped towers $\{\mathcal{T}_i^c\}$ as desired. We show the end result of this procedure in Figure 16. $\square$

The combination of Lemmas 6.1 and 6.3 yields the following proposition, which will be our main tool going forward.

Proposition 6.4. *Let M be a connected 4-manifold with good fundamental group and consider a generically immersed collection of discs*

$$F = (f_1, \dots, f_k)\colon (D^2 \sqcup \cdots \sqcup D^2, S^1 \sqcup \cdots \sqcup S^1) \looparrowright (M, \partial M).$$

Suppose that $\{\Sigma_i^c\}$ is a geometrically dual collection capped surfaces for $\{f_i\}$ such that

$$\lambda(C_\ell, C_m) = \mu(C_\ell) = 0$$

for each pair of caps C_ℓ and C_m of $\{\Sigma_i^c\}$. Assume in addition that the body Σ_i of each Σ_i^c is contained in a collar neighbourhood of ∂M.

Then there exists a collection of 1-storey capped towers $\{\mathcal{T}_i^c\}$, properly embedded in M, with arbitrarily many surface stages, such that the attaching region of $\mathcal{T}_i^c$ coincides with the

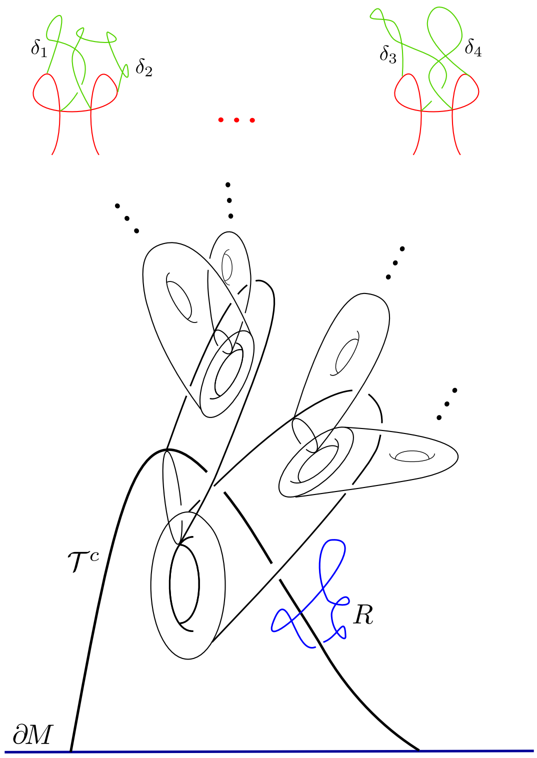


Figure 16. A 1-storey capped tower $\mathcal{T}^c$ with a geometrically dual sphere R.

framed boundary of f_i, with a geometrically dual collection of spheres $\{R_i\}$. Moreover, each R_i is obtained from Σ_i^c by contraction, after changing the caps of $\{\Sigma_i^c\}$ by a regular homotopy in the interior.

We need one more lemma, that we shall use to control the homotopy classes of the geometrically dual spheres in the output of the disc embedding theorem.

Lemma 6.5. *Let N be a 4-manifold and let*

$$D = (D_1, \dots, D_m) \colon (D^2 \sqcup \cdots \sqcup D^2, S^1 \sqcup \cdots \sqcup S^1) \to (N, \partial N)$$

be a generic immersion of a collection of discs that admits a geometrically dual, generically immersed collection $\{E_i\}_{i=1}^m$ of framed spheres in N. Let $\Sigma^c \subset N \setminus \bigcup_{i=1}^m \nu D_i$ be a capped surface constructed by taking a Clifford torus corresponding to an intersection point between D_i and D_j, where $i = j$ is permitted, and tubing meridional caps obtained from meridional discs for D_i and D_j into parallel copies of E_i and E_j. Let $S \colon S^2 \to N \setminus \bigcup_{i=1}^m \nu D_i$ be the 2-sphere obtained by contracting Σ^c. Then $\iota_*[S] = 0 \in \pi_2(N)$, where $\iota \colon N \setminus \bigcup_{i=1}^m \nu D_i \to N$ is the inclusion map.

Lemma 6.5 implies in particular that each of the spheres $\{R_i\}$ in Proposition 6.4 is null homotopic in M.

Proof. By Lemma 4.7, and as explained in [FQ90, Section 2.3], the homotopy class of a 2-sphere obtained by contracting a torus along caps is independent of the choice of caps, provided the boundaries of the different choices of caps coincide. Therefore in N we may replace the caps constructed from E_i and E_j by the meridional caps. The sphere S' resulting from contraction along the meridional caps is contained in a D^4 neighbourhood in N of the intersection point giving rise to the Clifford torus. So S' is null homotopic in N. It follows that $\iota \circ S$ is null homotopic as desired. $\square$

7. Proof of the disc embedding theorem

In this section we prove Theorem A, which is [FQ90, Theorem 5.1A]. For the convenience of the reader we recall the theorem, inserting the condition that F be an immersion, which we may do according to Theorem 1.2.

Theorem A ([FQ90, Theorem 5.1A]). *Let M be a connected 4-manifold with good fundamental group. Consider a continuous map*

$$F = (f_1, \dots, f_k)\colon (D^2 \sqcup \cdots \sqcup D^2, S^1 \sqcup \cdots \sqcup S^1) \longrightarrow (M, \partial M)$$

that is a locally flat embedding on the boundary and that admits algebraically dual spheres $\{g_i\}_{i=1}^k$ satisfying $\lambda(g_i, g_j) = 0 = \widetilde{\mu}(g_i)$ for all i, j. Then there exists a locally flat embedding

$$\overline{F} = (\overline{f}_1, \dots, \overline{f}_k)\colon (D^2 \sqcup \cdots \sqcup D^2, S^1 \sqcup \cdots \sqcup S^1) \hookrightarrow (M, \partial M)$$

such that $\overline{F}$ has the same boundary as F and admits a generically immersed, geometrically dual collection of framed spheres $\{\overline{g}_i\}_{i=1}^k$, such that $\overline{g}_i$ is homotopic to g_i for each i.

Moreover, if f_i is a generic immersion, then it induces a framing of the normal bundle of its boundary circle. The embedding $\overline{f}_i$ may be assumed to induce the same framing.

Proof. By Theorem 1.2, we may assume that the collection $\{f_i, g_j\}_{i,j}$ is generically immersed. Since $\lambda(g_i, g_j) = 0 = \widetilde{\mu}(g_i)$ for all $i, j = 1, \dots, k$, all the intersections and self-intersections within $\{g_i\}$ are paired by Whitney discs, after possibly adding cusps to g_i that achieve $\mu(g_i) = 0$. Similarly, since $\lambda(f_i, g_j) = \delta_{ij}$, all but one intersection point between each f_i and g_i are paired by Whitney discs, for each i, and all intersections between f_i and g_j, for $i \neq j$, are paired by Whitney discs.

Tube each intersection and self-intersection within $\{f_i\}$ into $\{g_i\}$ using the unpaired intersection point, as shown in Figure 17. We then obtain a collection of discs, which we still call $\{f_i\}$, where $\lambda(f_i, f_j) = \mu(f_i) = 0$ for each i, j. This follows since now all the intersections within the $\{f_i\}$ are paired by Whitney discs, obtained as parallel copies of the previous Whitney discs. Note also that the framing on the boundary of each f_i is unchanged since all g_j are framed.

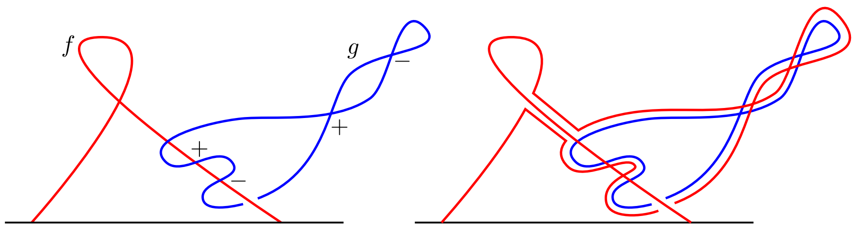


Figure 17. Modify f using g to get $\mu(f) = 0$.

Since the new intersections between $\{f_i\}$ and $\{g_i\}$ arise in algebraically cancelling pairs, we still have $\lambda(f_i, g_j) = \delta_{ij}$. Use the geometric Casson lemma (Lemma 5.1) to make $\{f_i\}$ and $\{g_i\}$ geometrically dual. Since the lemma provides a regular homotopy, the intersection and self-intersection information does not change.

Use that $\lambda(f_i, f_j) = \mu(f_i) = 0$ to pair up all the intersections and self-intersections among the $\{f_i\}$ with Whitney discs. Consider one such a Whitney disc D pairing up intersections between f_i and f_j, where possibly $i = j$. Boundary twist to ensure that the Whitney disc is framed, potentially introducing new intersections of D with the $\{f_i\}$. The resulting disc may intersect itself, the $\{f_i\}$ and the $\{g_i\}$, and any number of other Whitney discs.

For each intersection of D with f_ℓ, for some ℓ, tube D into a parallel push-off of the geometric dual g_ℓ. This introduces potentially many new intersections between D and anything that intersected g_ℓ (including g_ℓ itself), as well as new self-intersections of D coming from the self-intersections of g_ℓ. However, D is still framed and the interior of D no longer intersects $\{f_i\}$, since the collection $\{g_i\}$ is geometrically dual. Repeat this for all the Whitney discs so that the resulting discs have interiors in the complement of the $\{f_i\}$. Call this collection of framed Whitney discs $\{D'_j\}$. Note that if these $\{D'_j\}$ were disjointly embedded, we could use them to perform Whitney moves on $\{f_i\}$ and obtain the embedded discs $\overline{f}_i$ we seek.

We wish to apply Proposition 6.4 to the complement of the $\{f_i\}$. First we check that the fundamental group of the exterior of the $\{f_i\}$ is good. This follows by observing that for each i the meridian of f_i is null homotopic in the exterior of f_i, via the geometric dual sphere g_i punctured at the intersection point with f_i (see [FQ90, p. 26]). That is, the collection $\{f_i\}$ is π_1-negligible, which means that the inclusion induces an isomorphism $\pi_1(M\setminus\bigcup f_i) \cong \pi_1(M)$. Since $\pi_1(M)$ is good, $\pi_1(M\setminus\bigcup f_i)$ is also good, as desired. This group is also the fundamental group of the 4-manifold $M' := M\setminus\bigcup \nu(f_i)$ in which the following constructions take place.

Now, we want to find appropriate dual capped surfaces for the $\{D'_j\}$ in M'. Let Σ_j be the Clifford torus at one of the double points paired by D'_j. Observe that Σ_j intersects D'_j exactly once, and the collection of such Clifford tori are embedded, framed, and disjoint. Cap each Σ_j with meridional discs to $\{f_i\}$, i.e. the embedded discs described in Figure 18. Each cap has a unique intersection point with $\{f_i\}$ and no other intersections. Tube these intersections of the meridional caps with $\{f_i\}$ into parallel copies of the dual spheres $\{g_i\}$. The resulting capped surfaces $\{\Sigma_j^c\}$ then lie in M' as desired. Let C_ℓ and C_m be two caps of $\{\Sigma_j^c\}$. Since $\lambda(g_i, g_j) = \mu(g_i) = 0$ for all i, j and the fundamental group does not change from M to M', we see that $\lambda(C_\ell, C_m) = \mu(C_\ell) = 0$.

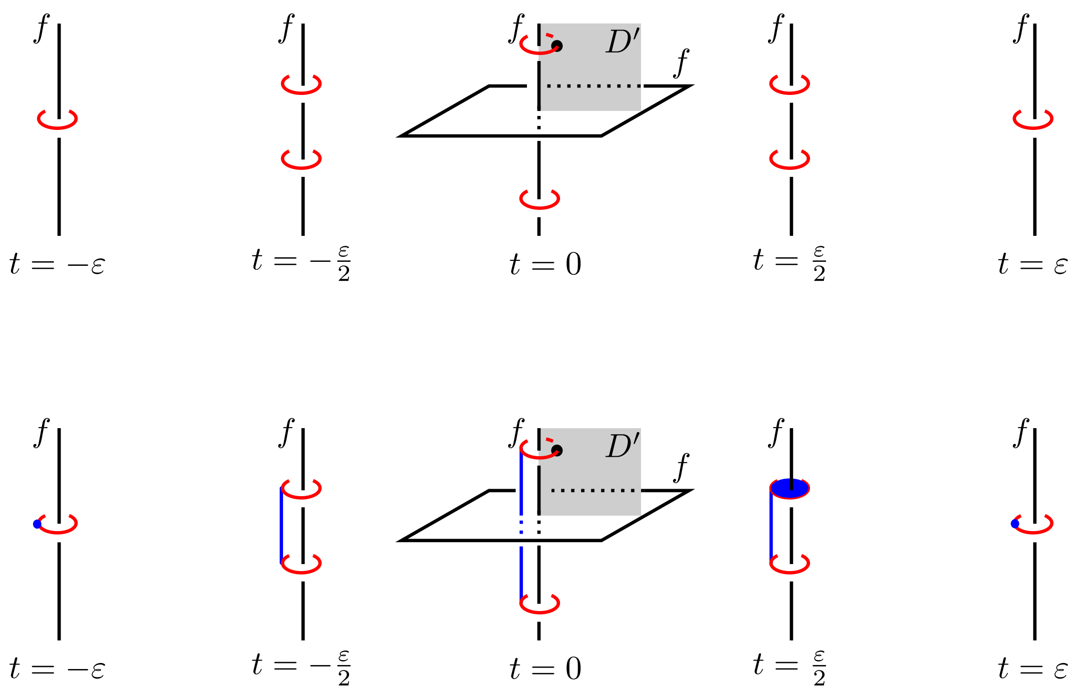


Figure 18. The top shows a (red) Clifford torus with a single intersection point with the (grey) Whitney disc D' pairing intersection points of f (black). In the bottom we see that each (blue) meridional disc intersects exactly one f_i once.

Note that in M', the caps of $\{\Sigma_j^c\}$ may intersect $\{D'_j\}$, due to the original intersections between $\{D'_j\}$ and $\{g_i\}$. Contract a parallel copy of each Σ_j^c to produce a dual sphere S_j for D'_j in M', and push the collection $\{D'_j\}$ off the contraction (do not push off anything else). This results in a collection of Whitney discs $\{D''_j\}$, with more self-intersections than the $\{D'_j\}$ but with $\{S_j\}$ geometrically dual spheres for $\{D''_j\}$.

By construction, each S_j is disjoint from the bodies $\{\Sigma_j\}$ but may intersect the caps of $\{\Sigma_j^c\}$. Tube all intersections of these caps with $\{D''_j\}$ into parallel copies of $\{S_j\}$. This gives new caps for Σ_i with more self-intersections but we have ensured that this new collection $\{\Sigma_j^c\}$ of capped surfaces in M' is geometrically dual to $\{D''_j\}$. They are also framed since the Clifford tori and the $\{g_i\}$ were framed. Moreover, the Clifford tori are located close to the original $\{f_i\}$, that is within a collar neighbourhood of the boundary of M'. Additionally, the latest caps for $\{\Sigma_j\}$ have vanishing intersection and self-intersection numbers.

Therefore we can apply Proposition 6.4 to M', to replace the discs $\{D''_j\}$ with 1-storey capped towers $\{\mathcal{T}_j^c\}$ whose framed attaching regions coincide with the framed boundary of $\{D''_j\}$ and that have geometrically dual spheres $\{R_j\}$, as shown in Figure 19.

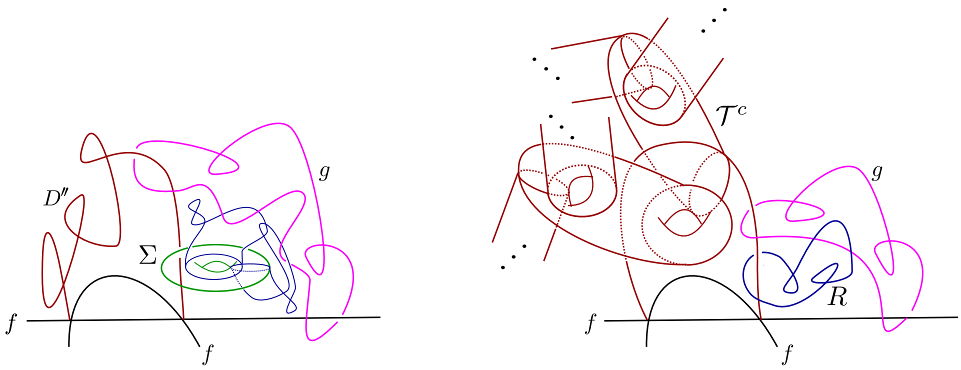


Figure 19. Before (left) and after (right) applying Proposition 6.4 to $M \setminus \bigcup \nu f_i =: M'$.

If some g_i intersects some $\mathcal{T}_j^c$, push down the intersection into the base surface and tube into R_j, as shown in Figure 20; note that an intersection between g_i and $\mathcal{T}_j^c$ may occur at any surface, disc, or cap stage of $\mathcal{T}_j^c$, so we may need to push down the intersection several times before reaching the base surface and tube all the resulting intersections into parallel copies of R_j. Call the resulting spheres $\{\overline{g}_i\}$.

Next we argue that $\overline{g}_i$ is homotopic in M to g_i for each i. Observe that to obtain $\overline{g}_i$, we have:

(i) homotoped g_i by an application of the geometric Casson lemma, and then
(ii) tubed into parallel copies of the dual spheres $\{R_j\}$ for the towers $\{\mathcal{T}_j^c\}$.

However, R_j was obtained by contracting Σ_j^c, whose body is a Clifford torus for an intersection point among the $\{f_i\}$. Therefore by Lemma 6.5, R_j is null homotopic in M, i.e. $[R_j] = 0 \in \pi_2(M)$. (Note that R_j is nontrivial in $\pi_2(M')$.) It follows that $\overline{g}_i$ is homotopic in M to g_i, as desired.

Now we return to replacing the $\{f_i\}$ by embedded discs. By construction, the collection $\{\mathcal{T}_j^c\}$ is properly embedded in M' and is disjoint from the dual spheres $\{\overline{g}_i\}$. We also have that the $\{f_i\}$ and the $\{\overline{g}_i\}$ are geometrically dual.

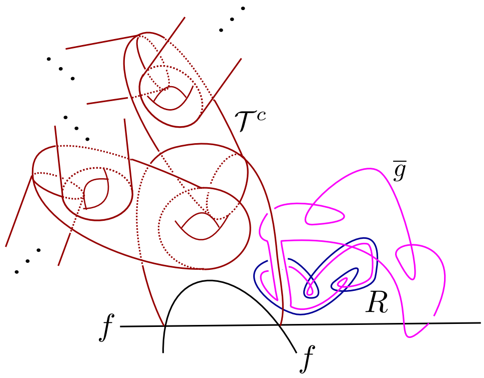


Figure 20. Tube to remove intersections of g with $\mathcal{T}^c$, ensuring that the interior of $\mathcal{T}^c$ lies in the complement of f and $\overline{g}$.

By Remark 5.6, every 1-storey capped tower with at least four surface stages contains a locally flat embedded disc whose framed boundary coincides with the attaching region of the capped tower. This produces mutually disjoint, embedded and framed Whitney discs pairing the intersections and self-intersections of the $\{f_i\}$ away from $\{\overline{g}_i\}$, as shown in Figure 21. Perform Whitney moves guided by these discs to produce embedded discs $\{\overline{f}_i\}$ with corresponding geometrically dual spheres $\{\overline{g}_i\}$. In the case that f_i was initially generically immersed, we obtain the same framing on the boundary of $\overline{f}_i$ since these are regular homotopies in the interior. This completes the proof of the disc embedding theorem with geometrically dual spheres. $\square$

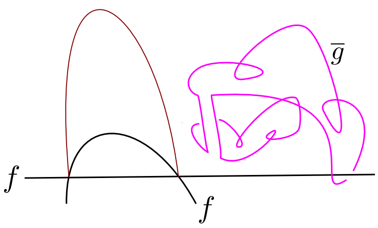


Figure 21. An embedded Whitney disc whose interior is in the complement of f and $\overline{g}$ has been produced.

8. The sphere embedding theorem

Next we prove [FQ90, Theorem 5.1B], whose statement we recall for the convenience of the reader. The main difference from Theorem A ([FQ90, Theorem 5.1A]) is that the intersection conditions are on the $\{f_i\}$ instead of on the $\{g_i\}$, and that we obtain embedded discs $\{f_i'\}$ regularly homotopic to the original immersed discs $\{f_i\}$. Unlike Theorem A, there is no assumption on the intersections among $\{g_i\}$ in the next theorem. The spheres $\{g_i\}$ are still required to be framed.

Note that there is no assumption on $\lambda(f_i, f_i)$, even though $\mu(f_i) = 0$ is required. Indeed in order to define $\lambda(f_i, f_i) = \lambda(f_i, f_i^+)$, we would have to fix a framing of ∂f_i. Then the relative Euler number of the normal bundle of f_i would equal $\lambda(f_i, f_i) \in \mathbb{Z} \subseteq \mathbb{Z}[\pi_1 M]$.

Theorem 8.1 ([FQ90, Theorem 5.1B]). *Let M be a connected 4-manifold with good fundamental group and consider a generically immersed collection of discs*

$$F = (f_1, \dots, f_k) \colon (D^2 \sqcup \cdots \sqcup D^2, S^1 \sqcup \cdots \sqcup S^1) \looparrowright (M, \partial M)$$

satisfying $\mu(f_i) = 0$ for all i and $\lambda(f_i, f_j) = 0$ for all $i \neq j$, with a generically immersed collection of framed algebraically dual spheres $\{g_i\}_{i=1}^k$. Then there exist mutually disjoint locally flat embeddings

$$\overline{F} = \overline{f}_1, \dots, \overline{f}_k \colon (D^2 \sqcup \cdots \sqcup D^2, S^1 \sqcup \cdots \sqcup S^1) \hookrightarrow (M, \partial M)$$

with $\overline{f}_i$ regularly homotopic rel. boundary to f_i for each i, and with a generically immersed, geometrically dual collection of framed spheres $\{\overline{g}_i\}_{i=1}^k$ such that $\overline{g}_i$ is homotopic to g_i for each i.

Proof. Since $\lambda(f_i, g_j) = \delta_{ij}$, we may apply the geometric Casson lemma (Lemma 5.1) to arrange that $\{f_i\}$ and $\{g_i\}$ are geometrically dual (note that the dual collection of spheres $\{g_i\}$ may have any kind of intersections amongst themselves). Since $\lambda(f_i, f_j) = 0$ for all $i \neq j$ and $\mu(f_i) = 0$ for all i, the intersections and self-intersections amongst the $\{f_i\}$ are paired by immersed Whitney discs. Consider one such Whitney disc D pairing up intersections between f_i and f_j, where possibly $i = j$. Boundary twist to ensure that the Whitney disc is framed, potentially introducing new intersections of D with the $\{f_i\}$. Such a resulting disc may intersect itself, any number of the $\{f_i\}$ and the $\{g_i\}$, and any number of other Whitney discs. For each intersection of D with f_ℓ, for some ℓ, tube D into a parallel push-off of the geometric dual g_ℓ. This introduces potentially many new intersections, between D and anything that intersected g_ℓ (including g_ℓ itself), as well as new self-intersections of D coming from the self-intersections of g_ℓ. However, the interior of D no longer intersects any f_i, since g_ℓ intersects exactly one of the $\{f_i\}$, namely f_ℓ, at the intersection point we used for tubing.

Do this for all the Whitney discs pairing intersections among $\{f_i\}$. Now our Whitney discs are more complicated, but their interiors lie in the complement of the $\{f_i\}$. Call this collection of Whitney discs $\{D_j\}$. Note that if these Whitney discs were embedded, we could perform the Whitney trick along them to obtain the embedded discs we seek.

We wish to apply Theorem A to the 4-manifold with boundary $M \setminus \bigcup \nu f_i$. First we check that the fundamental group of $M \setminus \bigcup \nu f_i$ is good. As before, this follows by observing that each f_i has a geometrically dual sphere, and as a result there is an isomorphism $\pi_1(M \setminus \bigcup \nu f_i) \to \pi_1(M)$. Since $\pi_1(M)$ is good, we conclude that $\pi_1(M \setminus \bigcup f_i)$ is also good, as desired.

Next, find algebraically dual spheres for $\{D_j\}$. As before, these will arise from Clifford tori. Let Σ_j be the Clifford torus at one of the two double points paired by some D_j. Recall that Σ_j intersects D_j exactly once, and the collection of such Clifford tori are embedded, framed, and pairwise disjoint. Cap each Σ_j with meridional discs to $\{f_i\}$, i.e. the discs described in Figure 18. Each cap has a unique intersection with $\{f_i\}$, and neither cap intersects $\{D_j\}$. Tube these intersections of the meridional caps with $\{f_i\}$ into parallel copies of the dual spheres $\{g_i\}$. Now contract these capped surfaces to produce algebraically dual spheres $\{h_j\}$ for the discs $\{D_j\}$. Since the collection $\{h_j\}$ is produced by contraction of capped surfaces with disjoint bodies, we see that $\lambda(h_j, h_k) = \mu(h_j) = 0$ for all j, k. Moreover note that by Lemma 6.5, we have $[h_j] = 0 \in \pi_2(M)$.

Apply Theorem A to replace the immersed Whitney discs $\{D_j\}$, with algebraic duals h_j, by disjointly embedded, framed Whitney discs $\{\overline{D}_j\}$ and geometrically duals $\{\overline{h}_j\}$ with $\overline{h}_j$ homotopic to h_j, in $M \setminus \bigcup \nu f_i$. For every intersection of some g_i with $\overline{D}_j$, tube that g_i into $\overline{h}_j$. This transforms the geometrically dual collection $\{g_i\}$ for $\{f_i\}$ to a collection $\{\overline{g}_i\}$, that may have more intersections among themselves, but that are still geometrically dual to the $\{f_i\}$. Since h_j is null homotopic in M for every j, so is $\overline{h}_j$. It follows that $\overline{g}_i$ is homotopic to g_i for each i.

Moreover, the Whitney discs $\{\overline{D}_j\}$ for the intersections among the $\{f_i\}$ now have interiors in $M \setminus (\bigcup f_i \cup \bigcup \overline{g}_i)$. Perform the Whitney move along these Whitney discs to obtain embedded discs $\{\overline{f}_i\}$, regularly homotopic (rel. boundary) to the $\{f_i\}$, with the same framed boundary as the $\{f_i\}$, and with geometrically dual spheres $\{\overline{g}_i\}$. □

Now we show that the above variant of the disc embedding theorem is in fact equivalent to the disc embedding theorem.

Proposition 8.2. *Theorem 8.1 and Theorem A are equivalent.*

Proof. Since we already deduced Theorem 8.1 from Theorem A, it suffices to prove that Theorem 8.1 implies Theorem A. Perform the first steps of the proof of Theorem A, until we tube each intersection and self-intersection within $\{f_i\}$ into $\{g_i\}$ using the unpaired intersection points (see Figure 17). This replaces $\{f_i\}$ with a collection of discs, which we still call $\{f_i\}$ and with the same framed boundaries satisfying $\lambda(f_i, f_j) = \mu(f_i) = 0$ for all i, j. Moreover, we still have that $\lambda(f_i, g_j) = \delta_{ij}$. Apply Theorem 8.1 to achieve the conclusion of Theorem A. □

Next we deduce the sphere embedding theorem from Theorem 8.1.

Theorem B (Sphere embedding theorem with framed duals). *Let M be a connected 4-manifold with good fundamental group and consider a continuous map*

$$F = (f_1, \ldots, f_k) \colon (S^2 \sqcup \cdots \sqcup S^2) \longrightarrow M$$

satisfying $\widetilde{\mu}(f_i) = 0$ for every i and $\lambda(f_i, f_j) = 0$ for $i \neq j$, and with a collection of algebraically dual spheres $\{g_i\}_{i=1}^k$ with $w_2(g_i) = 0$ for each i. Then there is a locally flat embedding

$$\overline{F} = (\overline{f}_1, \ldots, \overline{f}_k) \colon (S^2 \sqcup \cdots \sqcup S^2) \hookrightarrow M$$

with $\overline{F}$ homotopic to F and with a generically immersed, geometrically dual collection of framed spheres $\{\overline{g}_i\}_{i=1}^k$, such that $\overline{g}_i$ is homotopic to g_i for each i.

Moreover, if f_i is a generic immersion and $e(f_i) \in \mathbb{Z}$ is the Euler number of the normal bundle $\nu(f_i)$, then $\overline{f}_i$ is regularly homotopic to f_i if and only if $e(f_i) = \lambda(f_i, f_i)$.

Proof. By Theorem 1.2, we may assume that the entire collection $\{f_i, g_j\}_{i,j}$ is generically immersed, so in particular all intersections among these spheres are transverse. Add local cusps to the f_i if necessary to obtain homotopic generic immersions f_i' with $\mu(f_i') = 0$, or equivalently $e(f_i') = \lambda(f_i', f_i')$ as in the assumption of theorem.

For each i, find a point on f_i' which is far from all intersections and self-intersections of f_i'. Choose a small open ball B_i around this point. Let $M' := \mathrm{cl}(M \setminus \bigcup_i B_i)$. Since $\pi_1(M') \cong \pi_1(M)$ and removing the B_i does not change any intersection and self-intersection numbers, we may apply Theorem 8.1 to M' to find a regular homotopy from F' to an embedding, with geometric duals homotopic to the original duals. □

Remark 8.3. If we assume in addition to the hypotheses of Theorem B that $\lambda(f_i, f_i) = 0$ for all i, then we can get disjointly embedded framed spheres $\overline{f}_i$ as an output, since then $w_2(\overline{f}_i) = w_2(f_i) \equiv \varepsilon(\lambda(f_i, f_i)) = 0$ for all i, by (1) in the discussion just before Section 1.3. As a consequence, we can do surgery on M along the $\overline{f}_i$ to obtain a 4-manifold M'. The existence of the geometric duals $\overline{g}_i$ implies that $\pi_1(M') \cong \pi_1(M)$, but we claim that much more is true: each intersection (or self-intersection) of some $\overline{g}_j$ with a fixed $\overline{g}_i$ can be tubed into the unique intersection point between $\overline{g}_i$ and (a parallel copy of) $\overline{f}_i$. The result are geometric duals h_i to $\overline{f}_i$ that are now disjointly embedded. This means that each pair $(\overline{f}_i, h_i)$ has a regular neighbourhood that is a sphere bundle over S^2, with a 4-ball removed. The sphere bundle is trivial if and only if $w_2(h_i) = w_2(g_i) = 0$.

So in the setting of Theorem B where all $w_2(g_i)$ are assumed to vanish, we get a connected sum decomposition as in Corollary 1.4, which explains the sentence preceding the statement of this corollary in the introduction. Note that the $a_i := [f_i]$ form the first Lagrangian, but the second Lagrangian generated by $b_i := [h_i]$ is formed by linear combinations of the $[g_i]$ and $[f_i]$ induced by the geometric manoeuvres above.

Next we deduce Corollary 1.4, after recalling the statement.

Corollary 1.4. *Let M be a connected 4-manifold with good fundamental group and let H be a hyperbolic form in $(\pi_2(M), \lambda_M, \widetilde{\mu}_M)$, meaning that H is a $\mathbb{Z}[\pi_1 M]$-submodule of $\pi_2(M)$, generated by a hyperbolic basis consisting of classes $a_1, \dots, a_k, b_1, \dots, b_k \in H$ with*

$$\lambda(a_i, b_j) = \delta_{ij},\ \lambda(a_i, a_j) = 0 = \lambda(b_i, b_j) \text{ and } \widetilde{\mu}(a_i) = 0 = \widetilde{\mu}(b_i) \text{ for all } i, j.$$

Then there is a homeomorphism $M \approx (\#_{i=1}^k S^2 \times S^2)\# M'$ with a connected sum that on π_2 sends a_i to $[S_i^2 \times \{\mathrm{pt}_i\}]$ and b_i to $[\{\mathrm{pt}_i\} \times S_i^2]$. In particular, H is an orthogonal summand freely generated by $\{a_i, b_i\}$ and $\pi_2(M) \cong H \perp \pi_2(M')$.

Proof of Corollary 1.4. Represent the a_i and b_i by framed generic immersions, using Proposition 3.1 and the fact that in $\mathbb{Z}/2$ we have $w_2(a_i) \equiv \varepsilon(\lambda(a_i, a_i)) = 0$ and $w_2(b_i) \equiv \varepsilon(\lambda(b_i, b_i)) = 0$ for all i. Apply Theorem B to these framed generic immersions. The output is as in Remark 8.3, except that we have the additional information that $\lambda(b_i, b_j) = 0 = \widetilde{\mu}(b_i)$ for all i, j. This means that the operations of tubing into the $\{\overline{f}_i\}$ occur in such a way that up to homotopy there is no effect, i.e. the resulting disjointly embedded spheres $\{h_i\}$ still represent the $\{b_i\}$. □

9. The necessity of geometrically transverse spheres

We prove that geometrically dual spheres are essential to prove the exactness of the surgery sequence at the normal maps. First we give some background on why spheres are relevant.

To prove the exactness of the topological surgery sequence in dimension 4 for good fundamental groups, one must geometrically realise the vanishing of the L-group algebraic obstruction to a map being a homotopy equivalence [Wal99],[FQ90, Chapter 11]. More precisely, given a 4-dimensional Poincaré complex X and a compact manifold M, along with a 2-connected map $F\colon M \to X$ such that the kernel of the induced map on second homology is represented by hyperbolic pairs of immersed spheres, we wish to surger away the spheres representing this kernel to upgrade F to a homotopy equivalence F' that is bordant to F.

In order to perform surgery, we need to represent one sphere of each hyperbolic pair by a framed, embedded sphere. This step is performed in [FQ90, Theorem 11.3A]. We will now show that it is also required for these embedded spheres to be equipped with a geometrically dual sphere each, in order for F' to still induce an isomorphism on fundamental groups. As mentioned in the introduction, in higher dimensions algebraically dual spheres can be promoted to geometrically dual spheres using the Whitney trick, so the analogous concern does not arise in high-dimensional surgery theory. The next proposition gives the promised Example 1.5.

Proposition 9.1. *There exists a framed embedded sphere S in $S^2 \times S^2$ that has an algebraically dual sphere but no geometrically dual sphere.*

Proof. First observe that the Kirby diagram in Figure 22 depicts $S^2 \times S^2$, after adding 3- and 4-handles. To see this, perform a sequence of handle slides and isotopies to produce the diagram in Figure 23. The required handle slides and the 3-/4-handles are indicated in the captions to Figures 22 and 23.

A ribbon disc for the curve L_1 in Figure 22 is produced by the band move shown by a dotted line. We define the announced sphere $S \subset S^2 \times S^2$ to be the union of this ribbon disc and the core of the 2-handle attached along L_1. Call this 2-handle h. Since h is attached with zero framing, the sphere S is framed.

Note that $L_2 \subset S^3$ is an unknot, and an evident unknotted disc bounded by L_2, with the interior pushed into D^4, intersects S geometrically seven times, but algebraically once. Let T be the (framed) sphere formed by gluing this disc bounded by L_2 to the core of the 0-framed 2-handle attached along L_2. The spheres S and T are algebraically dual to one another by construction.

Let M' be the manifold produced from $S^2 \times S^2$ by performing surgery on S. Since $S^2 \times S^2$ is simply connected, the fundamental group of M' is normally generated by a meridian of S.

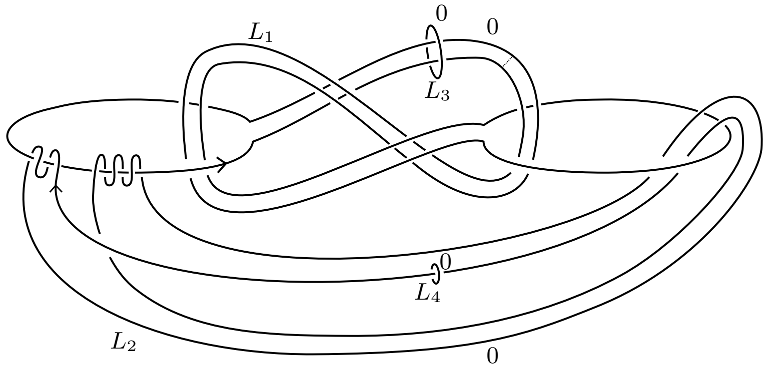


Figure 22. A Kirby diagram for $S^2 \times S^2$. As is customary, the 3- and 4-handles are not pictured (see Figure 23). The curve L_4 allows us to use handle slides to disentangle L_2 from everything else so that (L_2, L_4) becomes a 0-framed Hopf link. Then slide the handle attached to L_1 over L_3 to transform L_1 into an unknot split from everything else.

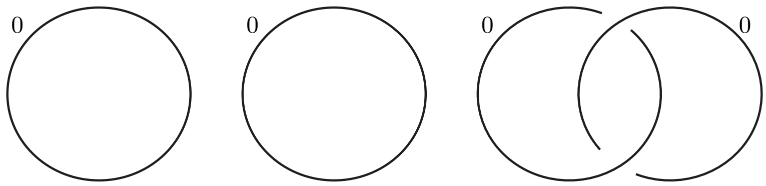


Figure 23. The result of performing a sequence of handle slides and isotopy to the diagram in Figure 22. The 2-handles represented by the unlink on the left are cancelled by the 3-handles. Attaching a single 4-handle produces $S^2 \times S^2$.

If S were equipped with a geometrically dual sphere, the meridian would be null-homotopic in M', and thus M' would also be simply connected. However, when S has only an algebraically dual sphere, the meridian only needs to be null homologous in M', showing that $H_1(M';\mathbb{Z})$ is trivial, so that $\pi_1(M')$ is perfect, i.e. equals its commutator subgroup. To prove that S admits no geometrically dual sphere, we will show in the remainder of the proof that the fundamental group of M' is nontrivial.

First we describe a decomposition of M'. Recall that there is a standard process for obtaining a Kirby diagram for a ribbon disc complement in D^4, where every minimum in the ribbon disc gives rise to a 1-handle in the complement, and every saddle gives rise to a 2-handle [GS99, Section 6.2]. The Kirby diagram for the ribbon disc complement for L_1 used in the construction of S is shown in Figure 24. Let V_1 denote this ribbon disc complement.

Observe that M' is built from V_1 by attaching to it the complement in h of its core, as well as all other 2-, 3-, and 4-handles in the decomposition shown in Figure 22, followed finally by a copy of $D^3 \times S^1$. This final attachment of $D^3 \times S^1$ can be seen as adding a 3-handle and a 4-handle. We know that the addition of 3- and 4-handles does not affect the fundamental group. Let V_2 denote the union of V_1 and the 2-handles in Figure 22 other than h. The submanifold $V_2 \subset M'$ is shown in Figure 24. Then M' is obtained from V_2 by attaching h without its core, in other words, a copy of $D^2 \times S^1 \times [0,1]$, as well as some 3- and 4-handles.

Lemma 9.2. *The fundamental groups of M' and V_2 are isomorphic.*

Proof of Lemma 9.2. Since 3- and 4-handle attachments do not change the fundamental group, we only need to show that the attachment of $D^2 \times S^1 \times [0,1]$ does not change the fundamental group of V_2. Note that this copy of $D^2 \times S^1 \times [0,1]$ is attached along a tubular neighbourhood of the knot L_1 minus its core.

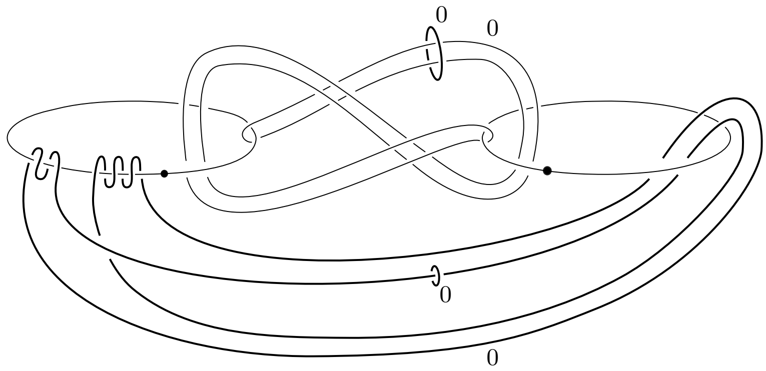


Figure 24. A Kirby diagram for the submanifold $V_2 \subset M'$. The lighter handles correspond to the ribbon disc complement for L_1.

Consider $V_3 = V_2 \cup_{S^1\times S^1\times[0,1]} D^2 \times S^1 \times [0,1]$. The kernel of

$$\pi_1(S^1 \times S^1 \times [0,1]) \cong \mathbb{Z} \oplus \mathbb{Z} \to \pi_1(D^2 \times S^1 \times [0,1]) \cong \mathbb{Z},$$

namely the summand corresponding to the first S^1, is included in the kernel of the map $\pi_1(S^1 \times S^1 \times [0,1]) \to \pi_1(V_2)$, because that S^1 is represented by a push-off of the boundary of a ribbon disc. By the Seifert-van Kampen theorem

$$\pi_1(V_3) \cong \pi_1(V_2) *_{\mathbb{Z}\oplus\mathbb{Z}} \mathbb{Z} \cong \pi_1(V_2) *_{\mathbb{Z}} \mathbb{Z} \cong \pi_1(V_2).$$

This completes the proof of the lemma. □

Returning to the proof of the proposition, we compute from the diagram in Figure 24 that

$$\pi_1(V_2) \cong \langle x, y \mid xyxy^{-1}x^{-1}y^{-1}, x^2yx^{-3}y\rangle.$$

This is known to be a presentation for the fundamental group of the Poincaré homology sphere. Specifically, observe that the presentation includes a standard presentation for the knot group of the right-handed trefoil; one may check that the additional relator describes a +1-framed longitude. Then we know that the above presentation describes the fundamental group of the 3-manifold obtained by performing +1-framed Dehn surgery on S^3 along the right-handed trefoil knot, which is well known to be the Poincaré homology sphere. In particular, $\pi_1(V_2)$ is a nontrivial perfect group and so $\pi_1(M') \cong \pi_1(V_2)$ is nontrivial, as desired. □

The first example of a sphere in $S^2 \times S^2$ with an algebraically dual sphere but no geometrically dual sphere was constructed in [Sat91, Section 3], [Sat89, Example 4.1] (a diagram is provided in [KM19, Figure 9]). Briefly, in Sato's construction one begins with a 2-knot Σ in S^4 and performs surgery on a simple closed curve in $S^4 \smallsetminus \Sigma$ which is homologous to the meridian. The result is an embedded sphere S in $S^2 \times S^2$. By a judicious choice of Σ, one may ensure that $\pi_1((S^2 \times S^2) \smallsetminus S)$ is a nontrivial perfect group as needed.

References

[Bar63] D. Barden. *The structure of manifolds*. PhD thesis, Cambridge University, 1963.

[BKKPR] S. Behrens, B. Kalmár, M. H. Kim, M. Powell, and A. Ray, editors. *The disc embedding theorem*. Oxford University Press, to appear.

[Bro72] W. Browder. *Surgery on simply-connected manifolds*. Springer-Verlag, New York-Heidelberg, 1972. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 65.

[Cas86] A. Casson. Three lectures on new infinite constructions in 4-dimensional manifolds. In *À la recherche de la topologie perdue*, volume 62 of *Progr. Math.*, pages 201–244. Birkhäuser Boston, Boston, MA, 1986. With an appendix by L. Siebenmann.

[Don83] S. K. Donaldson. An application of gauge theory to four-dimensional topology. *J. Differential Geom.*, 18(2):279–315, 1983.

[FQ90] M. Freedman and F. Quinn. *Topology of 4-manifolds*, volume 39 of *Princeton Mathematical Series*. Princeton University Press, 1990.

[Fre82] M. Freedman. The topology of four-dimensional manifolds. *J. Differential Geom.*, 17(3):357–453, 1982.

[Fre84] M. H. Freedman. The disk theorem for four-dimensional manifolds. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983)*, pages 647–663, Warsaw, 1984.

[FT95a] M. H. Freedman and P. Teichner. 4-manifold topology. I. Subexponential groups. *Invent. Math.*, 122(3):509–529, 1995.

[FT95b] M. H. Freedman and P. Teichner. 4-manifold topology. II. Dwyer's filtration and surgery kernels. *Invent. Math.*, 122(3):531–557, 1995.

[FT05] S. Friedl and P. Teichner. New topologically slice knots. *Geom. Topol.*, 9:2129–2158, 2005.

[GG73] M. Golubitsky and V. Guillemin. *Stable mappings and their singularities*. Springer-Verlag, New York-Heidelberg, 1973. Graduate Texts in Mathematics, Vol. 14.

[GS99] R. Gompf and A. Stipsicz. 4*-manifolds and Kirby calculus*, volume 20 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1999.

[GT04] S. Garoufalidis and P. Teichner. On knots with trivial Alexander polynomial. *J. Differential Geom.*, 67(1):167–193, 2004.

[HK88] I. Hambleton and M. Kreck. On the classification of topological 4-manifolds with finite fundamental group. *Math. Ann.*, 280(1):85–104, 1988.

[HK93] I. Hambleton and M. Kreck. Cancellation, elliptic surfaces and the topology of certain four-manifolds. *J. Reine Angew. Math.*, 444:79–100, 1993.

[HKT09] I. Hambleton, M. Kreck, and P. Teichner. Topological 4-manifolds with geometrically two-dimensional fundamental groups. *J. Topol. Anal.*, 1(2):123–151, 2009.

[KM19] M. R. Klug and M. Miller. Concordance of surfaces and the Freedman-Quinn invariant, 2019. arxiv:1912.12286.

[KQ00] V. S. Krushkal and F. Quinn. Subexponential groups in 4-manifold topology. *Geom. Topol.*, 4:407–430 (electronic), 2000.

[Kre99] M. Kreck. Surgery and duality. *Ann. of Math. (2)*, 149(3):707–754, 1999.

[KS77] R. C. Kirby and L. C. Siebenmann. *Foundational essays on topological manifolds, smoothings, and triangulations*. Princeton University Press, Princeton, N.J., 1977. With notes by John Milnor and Michael Atiyah, Annals of Mathematics Studies, No. 88.

[Lee03] J. M. Lee. *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2003.

[Maz63] B. Mazur. Relative neighborhoods and the theorems of Smale. *Ann. of Math. (2)*, 77:232–249, 1963.

[Nov64] S. P. Novikov. Homotopically equivalent smooth manifolds. I. *Izv. Akad. Nauk SSSR Ser. Mat.*, 28:365–474, 1964.

[Qui82] F. Quinn. Ends of maps. III. Dimensions 4 and 5. *J. Differential Geom.*, 17(3):503–521, 1982.

[Qui86] F. Quinn. Isotopy of 4-manifolds. *J. Differential Geom.*, 24(3):343–372, 1986.

[Qui88] F. Quinn. Topological transversality holds in all dimensions. *Bull. Amer. Math. Soc. (N.S.)*, 18(2):145–148, 1988.

[Sat89] Y. Sato. Smooth 2-knots in $S^2 \times S^2$ with simply-connected complements are topologically unique. *Proc. Amer. Math. Soc.*, 105(2):479–485, 1989.

[Sat91] Y. Sato. Locally flat 2-knots in $S^2 \times S^2$ with the same fundamental group. *Trans. Amer. Math. Soc.*, 323(2):911–920, 1991.

[Sma62] S. Smale. On the structure of manifolds. *Amer. J. Math.*, pages 387–399, 1962.

[Spr72] D. Spring. Proper homotopy and immersion theory. *Topology*, 11:295–305, 1972.

[Sta67] J. R. Stallings. *Lectures on polyhedral topology*. Notes by G. Ananda Swarup. Tata Institute of Fundamental Research Lectures on Mathematics, No. 43. Tata Institute of Fundamental Research, Bombay, 1967.

[Sto94] R. Stong. Existence of π_1-negligible embeddings in 4-manifolds. A correction to Theorem 10.5 of Freedmann and Quinn. *Proc. Amer. Math. Soc.*, 120(4):1309–1314, 1994.

[Sul96] D. P. Sullivan. Triangulating and smoothing homotopy equivalences and homeomorphisms. Geometric Topology Seminar Notes. In *The Hauptvermutung book*, volume 1 of *K-Monogr. Math.*, pages 69–103. Kluwer Acad. Publ., Dordrecht, 1996.

[Wal99] C. T. C. Wall. *Surgery on compact manifolds*, volume 69 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, second edition, 1999. Edited and with a foreword by A. A. Ranicki.

[Whi43] H. Whitney. The general type of singularity of a set of $2n - 1$ smooth functions of n variables. *Duke Math. J.*, 10:161–172, 1943.

[Whi44] H. Whitney. The singularities of a smooth n-manifold in $(2n - 1)$-space. *Ann. of Math. (2)*, 45:247–293, 1944.

Durham University, Department of Mathematical Sciences, Stockton Road, Durham, United Kingdom
Email address: mark.a.powell@durham.ac.uk

Max-Planck-Institut für Mathematik, Vivatsgasse 7, 53111 Bonn, Germany
Email address: aruray@mpim-bonn.mpg.de

Max-Planck-Institut für Mathematik, Vivatsgasse 7, 53111 Bonn, Germany
Email address: teichner@mac.com