

# 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. We also prove Proposition 1.6 from the Freedman-Quinn book regarding generic homotopies of discs or spheres in a 4-manifolds, which was not proven there.

## 1. INTRODUCTION

The *disc embedding theorem* [Fre84] combines work of Casson, Freedman, and Quinn. In a topological 4-manifold with *good* fundamental group, the theorem replaces an immersed disc with embedded boundary and a framed algebraically dual sphere by a locally flat embedded disc with the same boundary as the original disc. Consequences include topological 4-dimensional surgery theory and the topological 5-dimensional *s*-cobordism theorem, both for good fundamental groups.

Freedman’s original proof was restricted to the simply connected case and used *Casson handles*, built out of layers of (thickened) immersed discs. 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 expanded on Freedman’s argument by using a generalisation of Casson handles built out of capped gropes, variously called a skyscraper [BKK<sup>+</sup>21], 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 proofs of [FQ90, Theorem 5.1A and Corollary 5.1B] related to geometrically dual spheres. Briefly, 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. The proof in [FQ90] produces algebraically dual spheres in the output, rather than geometrically dual spheres.

Part II of Freedman-Quinn [FQ90] is the principal reference for the tools required to do any nontrivial work with topological 4-manifolds. These include the annulus theorem [FQ90, Theorem 8.1A], smoothing away from a point [FQ90, Theorem 8.2], transversality [FQ90, Theorem 9.5A], and the existence and uniqueness of normal bundles for locally flat submanifolds [FQ90, Section 9.3]. Every theorem in Chapters 7, 8, and 9 of [FQ90] relies on Theorem 5.1A, including the latter’s assertion of geometrically dual spheres. In turn the classification results of Chapters 10 and 11, and indeed all classification results proven since then, make essential use of these tools. Thus it is important to have a complete proof of Theorem 5.1A of [FQ90], from the point of view of both the rest of that book and of much of the rest of the literature on topological 4-manifolds.

In the spirit of addressing foundational omissions from [FQ90], we also prove [FQ90, Proposition 1.6] on homotopies between immersed surfaces in a 4-manifold. This states that a homotopy between generic immersions of a surface in a 4-manifold is homotopic to a composition of homotopies, each of which is a regular homotopy or a cusp homotopy in some ball, or the inverse of a cusp homotopy. This important proposition was stated but not proven in the Freedman-Quinn book. We will discuss the details in Section 1.3.

One might wonder whether our concern about the existence of geometrically dual spheres is valid, and whether, for example, the algebraically dual spheres produced in the proof in [FQ90] might suffice for the above mentioned applications. However, 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. Therefore, many applications hinge on providing geometric duals together with the embedded spheres. We also note that a 2-sphere in a 4-manifold with an algebraically dual sphere need not admit a geometrically dual sphere, a phenomenon not seen in higher dimensions due to the Whitney trick. In dimension four, geometric duals cannot be found *post hoc*. An example of a sphere in  $S^2 \times S^2$  with an algebraically dual sphere but no geometrically dual sphere, was produced in [Sat91, Section 3; Sat89, Example 4.1] cf. [KM19, Figure 9]. In the construction one begins with a 2-knot  $\Sigma \subseteq S^4$  and performs surgery on a simple closed curve in  $S^4 \setminus \Sigma$  which is homologous to the meridian. The result is an embedded sphere  $S$  in  $S^2 \times S^2$ . By a judicious choice of  $\Sigma$ , one may ensure that  $\pi_1((S^2 \times S^2) \setminus S)$  is a nontrivial perfect group, implying that  $S$  does not admit a geometrically dual sphere.

**1.1. The disc embedding theorem.** The *equivariant intersection form* of a connected 4-manifold  $M$  is a pairing

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

As well as having nonempty boundary, the topological 4-manifold  $M$  may be nonorientable or noncompact. For topological 4-manifolds, we will explain the *reduced self-intersection number*

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

in Section 1.3. If  $M$  is orientable, the quantity  $\tilde{\mu}$  is determined by  $\lambda$ . In this case, the assumptions  $\tilde{\mu}(g_i) = 0$  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  $\{g_i: S^2 \rightarrow M\}_{i=1}^k$  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, transverse point for  $i = j$  and is empty for  $i \neq j$ . An intersection point between  $f_i$  and  $g_i$  is said to be *transverse* if there are local coordinates  $\mathbb{R}^4 \hookrightarrow M$  in which  $f_i$  and  $g_i$  are linear. Here, and throughout the paper, we conflate the maps  $f_i, g_i$  with their images  $f_i(S^2)/f_i(D^2)$  and  $g_i(S^2)$ , as well as with the corresponding homology/homotopy classes.

We remark that [FQ90] uses the terminology of ‘algebraically transverse spheres’ which by definition have trivial normal bundles. We make the condition on normal bundles more explicit in our nomenclature. Our algebraically dual spheres need not have trivial normal bundles. When the latter condition is required, we state it explicitly.

**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*

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

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

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

*such that  $\bar{F}$  has the same boundary as  $F$  and admits a generically immersed, geometrically dual collection of framed spheres  $\{\bar{g}_i\}_{i=1}^k$ , such that  $\bar{g}_i$  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  $\bar{f}_i$  may then 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 6.3 to see that even the original continuous maps  $f_i$  induce framings modulo 2 on the boundary circles.

**Remark 1.2.** The geometrically dual spheres in the conclusion of the disc embedding theorem imply that inclusion induces an isomorphism  $\pi_1(M) \cong \pi_1(M \setminus \bigcup_i \bar{f}_i)$ . As in [FQ90], we refer to a collection of immersed surfaces whose removal does not change the fundamental group as  $\pi_1$ -negligible.

We discuss topological immersions in Section 1.3, 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 trivialisation, or framing, which is used in the last paragraph of the statement of Theorem A.

We recall the notion of a *good* group, from [FT95, KOPR21], in Definition 3.2. For applications, it suffices to know that the class of good groups is known to contain groups of subexponential growth [FT95, 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 use [FQ90, pp. 86-7, Proposition 2.9 and Lemma 3.3] to 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  $f$ . Then use [FQ90, Proposition 3.8] to show that every 1-storey capped tower with at least four surface stages contains a skyscraper with the same attaching region. Finally, using *decomposition space theory*, [FQ90, Theorem 4.1] shows that every skyscraper is homeomorphic to a handle  $D^2 \times D^2$  relative to its attaching region. The proof given in [FQ90, pp. 86-7] does not mention geometrically dual spheres, although they appear in the theorem statement. However, the claimed geometrically dual spheres in the output are used multiple times in [FQ90], as we indicate in Section 2.

In this article we give a modified proof of the disc embedding theorem, including the construction of the claimed geometrically dual spheres. Specifically, we only modify the first step of the proof, producing 1-storey capped towers equipped with geometrically dual spheres. Our modification utilises dual *capped surfaces* obtained from Clifford tori. Such capped surfaces allow us to create arbitrarily many collections of pairwise disjoint geometrically dual spheres. The key insight in our argument is that the Clifford tori are close to the corresponding double points, which allows us to ensure that the tracks of certain null homotopies do not intersect them. This small amount of extra disjointness is just enough to make the argument work. In Remark 1.4 we also explain an alternative argument suggested by a referee.

**Remark 1.3.** For experts, we explain the problem with the proof in [FQ90] in more detail. That proof begins with immersed discs with geometrically transverse spheres and upgrades the former to capped gropes with arbitrarily many surface stages, still equipped with geometrically transverse spheres  $\{g_i\}$ . The key step is finding a further stage of caps, constructing a 1-storey capped tower. The tower caps come from null homotopies for the double point loops in the caps of the capped grope, which exist due to the assumption on fundamental groups. The null homotopies cannot be easily controlled, and therefore the tower caps may intersect the dual spheres  $\{g_i\}$ , as well as the lower stages of the tower, arbitrarily. The final step in the proof of [FQ90, Lemma 3.3] is to push intersections of the tower caps with the grope caps or surfaces stages down to the base stage and tube into the spheres  $\{g_i\}$ . One sees then that the intersections between  $\{g_i\}$  and the tower caps have not been controlled. These could be pushed down into the surface stages, but in that case the 1-storey capped towers would still only have algebraically dual spheres.

It seems to have been tacitly assumed in [FQ90, Proposition 3.3 and pp. 86-7] that [FQ90, Lemma 3.3] provides geometrically dual spheres. In fact, the first paragraph of the proof on page 86 does not mention how to construct transverse spheres, and Lemma 3.3 does not claim to provide them. So [FQ90] gives a correct proof of Theorem 5.1A without the last four words ‘and with transverse spheres’ (known as geometrically dual spheres in our terminology). On the other hand, as discussed above, it is assumed throughout Part II of [FQ90] that Theorem 5.1A constructs geometrically dual spheres.

**1.2. Outline of the proof.** For experts, here is an outline of our argument – more details can be found in Sections 4 and 5. The beginning follows that of [FQ90, Theorem 5.1A]: assume the collections  $\{f_i\}$  and  $\{g_i\}$  are generically immersed, and after cusp homotopies that the intersections and self-intersections of  $\{g_i\}$  are algebraically cancelling. Tube the  $\{f_i\}$  into the  $\{g_i\}$  to ensure that the intersections and self-intersections of  $\{f_i\}$  are algebraically cancelling, and further arrange that  $\{f_i\}$  and  $\{g_i\}$  are geometrically dual by the geometric Casson lemma (Lemma 3.3).

Let  $\{D_j\}$  denote framed, immersed Whitney discs pairing the intersections within and among the  $\{f_i\}$ . Clifford tori at the double points of  $\{f_i\}$  are geometrically dual to  $\{D_j\}$ . These tori can be capped by meridional discs for  $\{f_i\}$ , tubed into  $\{g_i\}$  so that they lie in the complement of  $\{f_i\}$ , so we have dual capped tori  $\{T_j\}$  to the  $\{D_j\}$ .

Now comes our modification. Tube the intersections among  $\{D_j\}$  into the  $\{T_j\}$  to produce capped surfaces  $\{D'_j\}$  with the same framed boundary. The caps of  $\{D'_j\}$  and of the  $\{T_j\}$  can be separated as follows. The intersections between the two sets of caps are paired by framed, immersed Whitney discs. These may be assumed to be disjoint from the Clifford tori since the tori are located in a tubular neighbourhood of  $\{f_i\}$ , and they can be made disjoint from the bodies of  $\{D'_j\}$  by pushing down and tubing into geometrically dual spheres for  $\{D'_j\}$  produced by contracting parallel copies of the  $\{T_j\}$ . Now the geometric Casson lemma, applied to these new Whitney discs, separates the two families of caps, without creating any new undesirable intersections. Next we similarly upgrade the caps of  $\{D'_j\}$  to capped surfaces, with caps disjoint from those of the  $\{T_j\}$ . To do so, push the cap intersections for  $\{D'_j\}$  down and tube into  $\{T_j\}$  to obtain height two capped gropes pairing the intersections among  $\{f_i\}$ . Then separate the caps of these height two capped gropes from the caps of the  $\{T_j\}$  using the separation argument above, so that the final height two capped gropes are geometrically dual to the  $\{T_j\}$ .

The remainder of the proof is standard. By grope height raising and the good group hypothesis replace the height two capped gropes by height four capped gropes with null-homotopic double point loops and the same framed boundary. Let  $\{C_\ell\}$  denote immersed discs bounded by these double point loops, arising from the null homotopies and with the appropriate boundary framing. We may assume that  $\{C_\ell\}$  is disjoint from the Clifford tori (not their caps), since these lie in a tubular neighbourhood of  $\{f_i\}$ . For any intersections of  $\{C_\ell\}$  with the rest of the height four gropes, push down and tube into geometrically dual spheres produced by contracting parallel copies of  $\{T_j\}$ . Now the height four capped gropes can be equipped with  $\{C_\ell\}$  as tower caps to produce 1-storey capped towers  $\{\mathcal{T}_j^c\}$ . Finally, the  $\{T_j\}$  are contracted to spheres  $\{R_j\}$  which are geometrically dual to  $\{\mathcal{T}_j^c\}$ . By tubing  $\{g_i\}$  into  $\{R_j\}$ , we acquire spheres  $\{\bar{g}_i\}$  that are geometrically dual to  $\{f_i\}$  and do not intersect  $\{\mathcal{T}_j^c\}$ . Freedman and Quinn [FQ90, Chapters 3 and 4] (see also [BKK<sup>+</sup>21, Parts II and IV]) showed that 1-storey capped towers with at least four surface stages contain embedded topological discs with the same framed boundary. These can be used to perform the Whitney move on  $\{f_i\}$  to produce the desired embeddings  $\{\bar{f}_i\}$  with geometric duals  $\{\bar{g}_i\}$ .

Finally, in order to see that each  $\bar{g}_i$  is homotopic to  $g_i$ , we note that the latter only changes by homotopy and tubing into the spheres  $\{R_j\}$ . But each sphere  $R_j$  is null-homotopic in  $M$  since it is constructed by contracting a Clifford torus (Lemma 4.4).

**Remark 1.4.** We now describe an alternative method to obtain the geometrically dual spheres in Theorem A suggested by the referee.

As in the outline above, we can arrange that  $\{f_i\}$  and  $\{g_i\}$  are generically immersed, that the intersections and self-intersections of both  $\{f_i\}$  and of  $\{g_i\}$  are algebraically cancelling, and further that  $\{f_i\}$  and  $\{g_j\}$  are geometrically dual. Therefore the intersections and self-intersections of  $\{f_i\}$  are paired by framed, immersed Whitney discs. By tubing  $\{f_i\}$  to itself along one of each pair of Whitney arcs, the collection is upgraded to capped surfaces  $\{f'_i\}$ . Meridional discs and the Whitney discs provide the caps, after tubing into  $\{g_i\}$  to ensure disjointness from the bodies of  $\{f'_i\}$ .

Next, repeat the argument for the caps, as follows. First push all the intersections of  $\{g_i\}$  with the caps of  $\{f'_i\}$  down to the body and then into the (unpushed)  $\{g_i\}$ . This preserves the algebraic intersection conditions on  $\{g_i\}$ , while ensuring that the  $\{g_i\}$  no longer intersect the caps of  $\{f'_i\}$ . Now the intersections among the caps of  $\{f'_i\}$  are pushed down and tubed into the  $\{g_i\}$ , to arrange that the intersections among the caps of  $\{f'_i\}$  are algebraically cancelling. Then we can assume that the cap intersections are paired by framed, immersed Whitney discs. Tube along one of each pair of Whitney arcs, and cap with meridional and Whitney discs, with intersections with  $\{f'_i\}$  pushed down and tubed into  $\{g_i\}$  as needed, to upgrade  $\{f'_i\}$  to a collection of height two capped gropes.

Now apply grope height raising and the good group hypothesis to this collection to produce height four capped gropes with the same framed attaching region as  $\{f_i\}$  and null-homotopic double point loops.

Let  $\{C_\ell\}$  denote immersed discs bounded by these double point loops, arising from the null homotopies and with the appropriate boundary framing. For every intersection of  $\{C_\ell\}$  with the height four capped gropes, push down and tube into parallel copies of  $\{g_i\}$ . Now the height four capped gropes can be equipped with  $\{C_\ell\}$  as tower caps to produce 1-storey capped towers  $\{\mathcal{T}_i^c\}$ . This is the outcome of [FQ90, Lemma 3.3]. Note that  $\{g_i\}$  and the bodies of  $\{\mathcal{T}_i^c\}$  are geometrically dual, but we do not have any control yet on the intersections between  $\{g_i\}$  and  $\{C_\ell\}$ .

Now we have the modification proposed by the referee. Take parallel push-offs of  $\{g_i\}$ , which we call  $\{g'_i\}$ . For every intersection of  $\{C_\ell\}$  with  $\{g_i\}$ , push  $\{g_i\}$  down to the base surface of the tower, and tube into copies of  $\{g'_i\}$ . Still call the result  $\{g_i\}$ . Note that we have now arranged that the intersections between  $\{C_\ell\}$  and  $\{g_i\}$  are algebraically cancelling, since by the pushing down procedure, each intersection between some element of  $\{C_\ell\}$  and the original set of spheres  $\{g_i\}$  has led to some even number of (algebraically cancelling) tubings into  $\{g'_i\}$ . Let  $\{W_j\}$  denote a set of framed, immersed Whitney discs pairing the intersections between  $\{C_\ell\}$  and  $\{g_i\}$ . Remove any intersections between  $\{W_j\}$  and anything in the 1-storey towers  $\{\mathcal{T}_i^c\}$  below the tower caps by pushing down and tubing into  $\{g'_i\}$ . Now use the geometric Casson lemma (Lemma 3.3) with  $\{W_j\}$  to separate the caps of  $\{\mathcal{T}_i^c\}$  from  $\{g_i\}$ . The result is a new collection of 1-storey capped towers with geometrically dual spheres  $\{g_i\}$ .

The rest of the argument consists of first upgrading the 1-storey capped towers to skyscrapers, also equipped with geometrically dual spheres, and then showing that the homeomorphism from any skyscraper to  $D^2 \times D^2$ , relative to the attaching region, preserves the geometrically dual spheres. One can verify that the proofs of [FQ90, Proposition 3.8 and Theorem 4.1] accomplish this, by moving the intersection point with the dual spheres sufficiently close to the preserved attaching region.

Finally we check that the geometrically dual spheres produced are homotopic to the algebraically dual spheres in the hypotheses. Note that in the construction above each  $g_i$  has been changed by homotopies and by tubing into  $\{g'_i\}$ . Since the tubing occurred after pushing down, each  $g'_j$  is used for these tubing operations an algebraically cancelling number of times and therefore there is no change in the homotopy class of  $g_i$ , as desired.

**1.3. Generic immersions and intersection numbers.** 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. 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.

In order to work effectively in a topological 4-manifold, it is key to be able to assume that a continuous map of a surface can be perturbed to a topological generic immersion [FQ90, Lemma 1.2]. For instance, this is the first step in the proof of the disc embedding theorem in a topological 4-manifold. It also follows that a topological generic immersion  $f: \Sigma \looparrowright M$  has a linear normal bundle  $\nu(f)$ . Similarly, it is also essential to be able to decompose a topological homotopy into a sequence of regular homotopies and cusp homotopies [FQ90, Proposition 1.6]. This latter proposition was stated in [FQ90] without a proof, so we provide one.

Specifically, the combination of [FQ90, Lemma 1.2 and Proposition 1.6] can be stated in the following useful way, which is what we prove in Section 6.

**Theorem 1.5.** *Let  $\Sigma$  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*

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

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  $\Sigma$ . Moreover, any such homotopy



between generic immersions is homotopic rel.  $\Sigma \times \{0, 1\}$  to a sequence of isotopies, finger moves and Whitney moves.

As usual, 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  $f$  [Wal99, Chapter 5; FQ90, Section 1.7]. 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:

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

While  $\mu(f)$  depends on the choice of generically immersed representative  $f: S^2 \looparrowright M$ , the reduced version  $\tilde{\mu}$  only depends on  $[f] \in \pi_2(M)$ . Any two generic immersions homotopic to  $f$  are of course themselves homotopic. By the injectivity in Theorem 1.5 we see that  $\tilde{\mu}$  is well-defined on  $\pi_2(M)$ , since the quantity is evidently unchanged by isotopies, finger moves, and Whitney moves.

For any  $f: S^2 \looparrowright M$ , the invariant  $\mu$  satisfies

$$\lambda(f, f) = \mu(f) + \overline{\mu(f)} + e(f) \cdot 1 \in \mathbb{Z}[\pi_1 M] \quad (1.1)$$

where  $e(f) \in \mathbb{Z}$  is the *Euler number* of the normal bundle  $\nu(f)$  and the involution  $\bar{g} := w(g)g^{-1}$  is extended linearly to the group ring. Apply the augmentation map  $\varepsilon: \mathbb{Z}[\pi_1 M] \rightarrow \mathbb{Z}$  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 Theorem B in the next section. 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  $f: S^2 \looparrowright M$  that can be framed, i.e. whose normal bundle is trivial.

**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 give applications of the disc embedding theorem with geometrically dual spheres from the literature. In particular we recall the sphere embedding theorem, and the *hyperbolic embedding theorem*, which gives us the ability to represent a rank  $2k$  hyperbolic summand of the intersection form by an  $\#_{i=1}^k S^2 \times S^2$  connected summand. In Section 3, we review the objects, geometric constructions, and further results needed for the proof of the disc embedding theorem, then Section 4 contains the main technical results needed for the construction of geometrically dual spheres. These results are applied in Section 5 to prove the disc embedding theorem (Theorem A). Here we also explain an alternative argument kindly suggested by a referee. Finally, in Section 6 we discuss generic immersions and prove Theorem 1.5.

**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 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, Slava Krushkal, Allison N. Miller, Patrick Orson, and Frank Quinn for their insights and valuable discussions. We thank Michael Freedman for encouraging us to write this paper. Finally we are grateful to an anonymous referee for comments that helped improve the exposition, and the alternative argument outlined in Remark 1.4.

MP was partially supported by EPSRC New Investigator grant EP/T028335/2 and EPSRC New Horizons grant EP/V04821X/2.

## 2. APPLICATIONS OF GEOMETRICALLY DUAL SPHERES IN THE LITERATURE

**2.1. The sphere embedding theorem.** A central application of the disc embedding theorem is to prove the *sphere embedding theorem*. The existence and exactness of the topological surgery sequence in dimension four are proven by changing a collection of generically immersed spheres with vanishing intersection and self-intersection numbers, by a regular homotopy, to pairwise disjoint embedded spheres [FQ90, Theorem 11.3A]. The sphere embedding theorem describes precisely when such embedded spheres may be found. The proof of [FQ90, Theorem 11.3A] neglects to

mention that geometrically dual spheres are essential for performing surgery on 4-manifolds without inadvertently modifying the fundamental group. The sphere embedding theorem is also 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 sphere embedding to move from the easier stable classification up to connected sums with copies of  $S^2 \times S^2$  to unstable classification results. We refer to [OPR21, Section 20.3] for a proof of the sphere embedding theorem.

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

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

*satisfying  $\tilde{\mu}(f_i) = 0$  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*

$$\bar{F} = (\bar{f}_1, \dots, \bar{f}_k): (S^2 \sqcup \dots \sqcup S^2) \hookrightarrow M$$

*with  $\bar{F}$  homotopic to  $F$  and with a generically immersed, geometrically dual collection of framed spheres  $\{\bar{g}_i\}_{i=1}^k$ , such that  $\bar{g}_i$  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  $\bar{f}_i$  is regularly homotopic to  $f_i$  if and only if  $e(f_i) = \lambda(f_i, f_i)$ .*

**Remark 2.1.** The assumption  $\tilde{\mu}(f_i) = 0$  implies that  $\lambda(f_i, f_i) \in \mathbb{Z} \cdot 1 \subseteq \mathbb{Z}[\pi_1 M]$ , by (1.1). In this case, (1.1) 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  $\mu(f_i)$ .

**Remark 2.2.** 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  $\{\bar{f}_i\}$  as an output, since then  $w_2(\bar{f}_i) = w_2(f_i) \equiv \varepsilon(\lambda(f_i, f_i)) = 0$  for all  $i$ , by the equality (1.1). As a consequence, we can surger  $M$  along  $\{\bar{f}_i\}$  to obtain a 4-manifold  $M'$ . The existence of the geometric duals  $\bar{g}_i$  implies that  $\pi_1(M') \cong \pi_1(M)$ , but indeed more is true: each intersection (or self-intersection) of some  $\bar{g}_j$  with a fixed  $\bar{g}_i$  can be tubed into the unique intersection point between  $\bar{g}_i$  and (a parallel copy of)  $\bar{f}_i$ , resulting in geometric duals  $h_i$  to  $\bar{f}_i$  that are now disjointly embedded. This means that each pair  $(\bar{f}_i, h_i)$  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  $M$  as a connected sum with copies of  $S^2 \times S^2$ . 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.

0

We record an important special case of the situation described in Remark 2.2 in the following theorem.

**Theorem 2.3** (Hyperbolic embedding theorem). *Let  $M$  be a connected 4-manifold with good fundamental group and let  $H$  be a hyperbolic form in  $(\pi_2(M), \lambda_M, \tilde{\mu}_M)$ , 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*

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

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

Theorem 2.3 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, a simply connected, closed 4-manifold with intersection form  $E_8 \oplus E_8$  can be obtained from the  $K3$  surface by removing a hyperbolic form of rank 6. Since the form  $E_8 \oplus E_8$  is definite but not diagonalisable, it cannot be realised by a closed, smooth, 4-manifold.

*Proof of Theorem 2.3.* Represent the  $a_i$  and  $b_i$  by framed generic immersions, using Proposition 6.2 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 2.2, except that we have the additional information that  $\lambda(b_i, b_j) = 0 = \tilde{\mu}(b_i)$  for all  $i, j$ . This means that the operations of tubing into the  $\{\bar{f}_i\}$  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  $\{b_i\}$  in  $\pi_2(M)$ .  $\square$

**Remark 2.4.** In this paper we prove the disc embedding theorem with geometrically dual spheres, and then use it to deduce the sphere embedding theorem with geometrically dual spheres. Since the former is used throughout [FQ90], our proof fills the gap in that source. However, for applications in surgery theory, one begins with a dual spherical sublagrangian of the intersection form, as in Theorem 2.3. In this case, the following proof of Theorem 2.3 is available, which applies the disc embedding theorem without geometrically dual spheres. The idea is due to Casson [Cas86]; we thank Slava Krushkal for reminding us of it.

Suppose we are given classes  $a_1, \dots, a_k, b_1, \dots, b_k \in \pi_2(M)$  for a topological 4-manifold  $M$  with good fundamental group satisfying

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

Assume  $\{a_i\}$  and  $\{b_i\}$  are represented by immersed spheres  $\{A_i\}$  and  $\{B_i\}$  respectively. Let  $\{A_i^+\}$  and  $\{B_i^+\}$  denote push-offs of  $\{A_i\}$  and  $\{B_i\}$  respectively. For each  $i$ , choose small 4-discs around the unpaired intersection point between  $A_i$  and  $B_i$ , not intersecting  $\{A_i^+\}$  and  $\{B_i^+\}$ . Tube the discs together using embedded 1-handles in  $M$  without intersecting  $\{A_i\} \cup \{A_i^+\} \cup \{B_i\} \cup \{B_i^+\}$ , and call the result  $D$ . The portion of  $\{A_i\}$  and  $\{B_i\}$  lying in  $M \setminus D$  are immersed discs  $\{A_i^0\}$  and  $\{B_i^0\}$  with trivial intersection and self-intersection numbers, equipped with algebraically dual spheres  $\{B_i^+\} \cup \{A_i^+\}$ . Apply local cusp moves to  $\{A_i^0\}$  and  $\{B_i^0\}$  to ensure that the unreduced self-intersection numbers vanish. Apply the version of [FQ90, Corollary 5.1B] without geometrically dual spheres to  $\{A_i^0\}$  and  $\{B_i^0\}$ , to produce regularly homotopic, disjoint, embedded discs with the same framed boundary. Replacing the portions of  $\{A_i\}$  and  $\{B_i\}$  within  $D$  produces the desired geometrically transverse embeddings in the classes  $a_1, \dots, a_k, b_1, \dots, b_k \in \pi_2(M)$ .

It is tempting to attempt a similar strategy to prove the general disc and sphere embedding theorems with geometrically dual spheres. However, this does not work, even for sphere embedding. In the general sphere embedding scenario, one set of spheres may not have the requisite triviality of intersection and self-intersection numbers. This arises in applications, for example in the construction of star partners for 4-manifolds with odd intersection form [FQ90, Section 10.4], [Sto94] [Tei97], [KPR22]. Also, it was essential that we were working with spheres, since ‘dual discs’ are not as helpful.

**2.2. Other instances of geometrically dual spheres.** First we consider the book [FQ90], which is generally regarded as the canonical source for the ramifications of the disc embedding theorem.

- (1) On page 105 of [FQ90], in Part II of the book, within the proof of the technical version of  $h$ -cobordism theorem, it is claimed that 1-storey capped towers with geometrically dual spheres were constructed in Part I. Geometrically dual spheres are used on page 107, where they are key to proving the *negligibility property*.
- (2) The technical version of the  $h$ -cobordism theorem is used in the proof of the technical controlled  $h$ -cobordism theorem ([FQ90, Theorem 7.2C]), which in turn appears in the proofs of the proper  $h$ -cobordism theorem (Corollary 7.3C) and the annulus conjecture (Theorem 8.1A). The negligibility property (called a *regular homotopy property* in the proof of 8.1A) is invoked in an essential way.
- (3) The annulus conjecture is used in [FQ90] to prove topological transversality (Section 9.5), existence and uniqueness of normal bundles (Section 9.3), and smoothing results (Chapter 8).
- (4) The geometrically dual spheres claimed in Theorem B ([FQ90, Corollary 5.1B]) also arise in [FQ90] in the proof of the  $\pi_1$ -negligible embedding theorem (Theorem 10.5A), the plus construction (Theorem 11.1A), and the  $\pi - \pi$  lemma (p. 216).

Next we consider geometrically dual spheres elsewhere in the literature.



- (5) Similar to classical surgery theory, uses 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.
- (6) For  $M$  a compact 4-manifold with  $\pi_1(M)$  good, every element of the Whitehead group  $\text{Wh}(\pi_1(M))$  is realised as the Whitehead torsion of an  $h$ -cobordism  $(W; M, M')$  based on  $M$ . To prove this, one builds a cobordism with 2- and 3-handles. When attaching the 3-handles, one can find smoothly embedded spheres in the desired homotopy classes, but it is not known how to find such embeddings smoothly that also come with geometrically dual immersed spheres. Without the geometric duals, the inclusion induced map  $\pi_1(M') \rightarrow \pi_1(W)$  need not be an isomorphism. Thus the sphere embedding theorem is needed to obtain topologically locally flat embeddings for the 3-handle attachments, with geometrically dual spheres. See [KPR22, Theorem 3.5].
- (7) 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  $\mathbb{Z} \ltimes \mathbb{Z}[1/2]$ .

The papers of Casson [Cas86] and Freedman [Fre82, Fre84] do not have problems relating to geometrically dual spheres. A minor point is that 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, however this can be easily fixed using the method of Remark 2.4 instead.

### 3. DEFINITIONS AND OPERATIONS

We shall assume that a Whitney disc is a generic immersion  $W: D^2 \looparrowright M$ , which in particular implies that the boundary is embedded, with each Whitney arc lying on one of the sheets whose intersection points are paired by  $W$ . We allow interior self-intersections of  $W$  as well as intersections with other surfaces (and other Whitney discs) but assume that these are transverse, using topological transversality. As usual, a Whitney disc  $W$  is said to be *framed* if the Whitney section of the normal bundle  $\nu(W)$  of  $W$  restricted to its boundary extends to a non-vanishing section 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. The framing of a Whitney disc can be altered by interior twists and boundary twists; see [FQ90, Section 1.3].

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. From now on, we will assume that Whitney circles are pairwise disjoint and embedded.

In our proofs, we will construct *groves* and ultimately *towers*, as defined in [PR21c] (see also [FQ90, Chapters 2 and 3]). We assume the reader is familiar with these definitions. We will need the following generalisation of the notion of dual spheres.

**Definition 3.1** (Dual capped groves and surfaces). Let  $\{A_i\}$  be a collection of immersed discs, groves, or towers, with or without caps. A collection  $\{T_i^c\}$  of (sphere-like) capped groves 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 grove of height one, then we say that  $\{T_i^c\}$  is a collection of *dual capped surfaces* for  $\{A_i\}$ .

We will use the definition of a good group in the proof of Lemma 4.2, so we recall it here.

**Definition 3.2** ([FT95, KOPR21]). A group  $\Gamma$  is said to be *good* if for every height 1.5 disc-like capped grove  $G^c$ , with some choice of basepoint, and for every group homomorphism  $\phi: \pi_1(G^c) \rightarrow \Gamma$ , there exists an immersed disc  $D \looparrowright G^c$  whose framed boundary coincides with the attaching region of  $G^c$ , 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  $\Gamma$  by  $\phi$ .

We will need the following lemma, to trade intersections between distinct surfaces for self-intersections.

**Lemma 3.3** (Geometric Casson lemma [Fre82, Lemma 3.1] (see also [PR21a, Lemma 15.3])). *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  $\bar{F} \cup \bar{G}$  such that  $\bar{F} \pitchfork \bar{G} = (F \pitchfork G) \setminus \{p, q\}$ , 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. Moreover, the regular homotopy is supported in a small neighbourhood of  $W$ .*

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

- (i) Making  $F$  and  $G$  disjoint, if all intersection points  $F \cap G$  are paired by Whitney discs.
- (ii) Turning algebraically dual spheres  $G$  for  $F$  into geometrically dual spheres  $\bar{G}$ .

The process of *contraction and push-off*, introduced in [FQ90, Section 2.3] (see also [PR21a, Section 15.2.5]) will be important in our proof. *Contraction* converts a capped surface into an immersed disc using two parallel copies of both caps. 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 (but does not necessarily have to) be pushed off the contraction. This reduces the number of intersection points between the resulting contraction and the pushed off surfaces. 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 1.

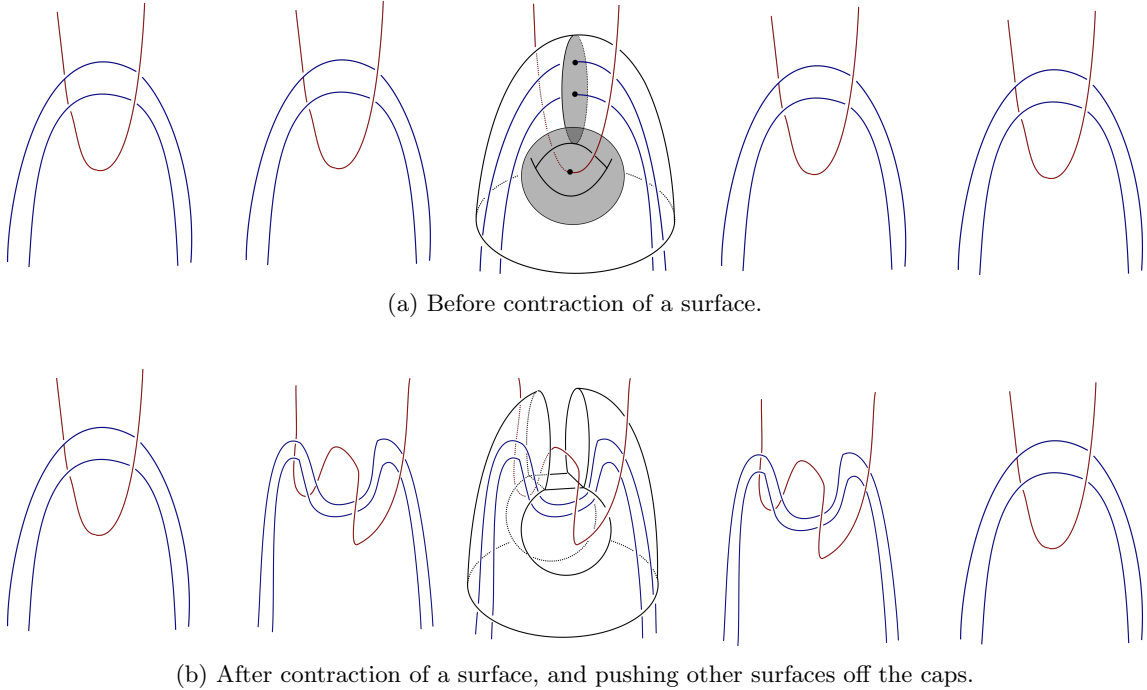


Figure 1. Contraction and push-off. Note the intersections of pushed-off surfaces that occur between diagrams one and two and between diagrams four and five in the bottom row of figures, namely one intersection in the past and one intersection in the future between each pair of surfaces pushed off dual caps.

**Lemma 3.4.** *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.  $\square$

#### 4. 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. The extra power of the dual capped surfaces, compared to dual spheres, is that we can use them to produce dual spheres multiple times in succession.

**Lemma 4.1.** *Consider a generically immersed collection of discs in a 4-manifold  $M$*

$$D = (D_1, \dots, D_k): (D^2 \sqcup \dots \sqcup D^2, S^1 \sqcup \dots \sqcup S^1) \looparrowright (M, \partial M).$$

*Suppose that  $\{T_i^c\}$  is a dual collection of capped surfaces for the  $\{D_i\}$  such that*

$$\lambda(C_\ell, C_m) = \mu(C_\ell) = 0 \quad (4.1)$$

*for every pair of caps  $C_\ell$  and  $C_m$  of  $\{T_i^c\}$ . Assume in addition that the body  $T_i$  of each  $T_i^c$  is contained in a collar neighbourhood of  $\partial M$ , and that otherwise only boundary collars of each of the  $D_i$  and the  $C_\ell$  lie in the collar neighbourhood.*

*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 parallel copies of the  $\{T_i^c\}$ , 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  $D_i$ .*

*Proof.* Parameterise the collar neighbourhood of  $\partial M$  as  $\partial M \times [0, 1] \subseteq M$ , where  $\partial M = \partial M \times \{0\}$ . Assume that initially the bodies  $\{T_i\}$  of the dual capped surfaces lie in  $\partial M \times (5/6, 1)$ . We will take parallel copies of the  $T_i^c$  throughout the proof of this lemma, and also in the proof of Lemma 4.2 below, such that the bodies lie progressively closer to  $\partial M \times \{0\}$ , in regions of the form  $\partial M \times ((6-n)/6, (7-n)/6)$ ,  $n = 1, \dots, 6$ , and caps are obtained as parallel copies of the original caps extended at the boundary collar.

Let  $\{C_\ell\}$  denote the caps of the  $\{T_i^c\}$ . Tube all the intersections and self-intersections among the  $\{D_i\}$  into parallel copies of the dual surfaces  $\{T_i^c\}$ , with bodies  $\{T_i\}$  contained in  $\partial M \times (5/6, 1)$ . This produces a mutually disjoint collection of capped surfaces  $\{D'_i\}$  with the same framed boundaries as  $\{D_i\}$ . Let  $\{C'_n\}$  denote the caps for  $\{D'_i\}$ .

Now we will separate the collections  $\{C_\ell\}$  and  $\{C'_n\}$ . For each  $i$ , take a parallel copy of  $T_i$  in  $\partial M \times (4/6, 5/6)$ , along with its caps, and contract to obtain a sphere  $S_i$ . Push nothing off the contraction. The collection  $\{S_i\}$  is geometrically dual to  $\{D'_i\}$ . By construction and (4.1), all the intersections between  $\{C_\ell\}$  and  $\{C'_n\}$  can be paired by framed Whitney discs  $\{W_k\}$ . Since  $\partial W_k$  lies in  $M \setminus (\partial M \times [0, 4/6])$ , we may and shall assume that  $\{W_k\}$  does not intersect  $\partial M \times [0, 4/6]$ ; this follows since each constituent Whitney disc is obtained as a track of a null homotopy, and because the inclusion map  $M \setminus (\partial M \times [0, 4/6]) \hookrightarrow M$  is a homotopy equivalence. From now on we consider parallel copies of  $\{T_i^c\}$  with each of the bodies  $T_i$  contained in  $\partial M \times [0, 4/6]$ . We continue to denote these parallel copies by  $T_i^c$ .

By tubing the  $\{W_k\}$  into the  $\{S_i\}$  we can also arrange that  $\{W_k\}$  does not intersect  $\{D'_i\}$ . Now the geometric Casson lemma (Lemma 3.3) applied with the Whitney discs  $\{W_k\}$  ensures that there are no intersections between the collections  $\{C_\ell\}$  and  $\{C'_n\}$ . The fact that  $W_k \cap D'_j = \emptyset = W_k \cap T_i$ , for all  $i, j, k$ , implies that no unwanted cap-body intersections are created whenever  $T_i^c$  is a parallel copy whose body lies in  $\partial M \times [0, 4/6]$ .

Push the intersections and self-intersections of the caps  $\{C'_n\}$  down to  $\{D'_i\}$  and tube into parallel copies of  $\{T_i^c\}$ , with bodies in  $\partial M \times (3/6, 4/6)$ , to produce height two capped gropes  $\{D''_i\}$ . Let  $\{C''_p\}$  denote the caps of  $\{D''_i\}$ . Next we separate the collections  $\{C''_p\}$  and  $\{C_\ell\}$ , as we did above.

The procedure is the same: contract parallel copies of  $\{T_j^c\}$  with bodies in  $\partial M \times (2/6, 3/6)$ , pushing nothing off the contraction, to obtain a collection  $\{S_i'\}$  of framed spheres geometrically dual to  $\{f_i''\}$ ; find framed Whitney discs for the intersections between  $\{C_p''\}$  and  $\{C_\ell\}$  which have interiors disjoint from  $\partial M \times [0, 2/6)$ , as well as from  $\{D_i''\}$  by tubing into  $\{S_i'\}$ . Then apply the geometric Casson lemma. Here, since the boundaries of the Whitney discs lie on  $\bigcup \{C_\ell\} \cup \{C_p''\} \subseteq M \setminus (\partial M \times [0, 2/6))$ , we may assume that the Whitney discs created by choosing a null homotopy miss  $\partial M \times [0, 2/6)$ , and hence miss the body  $T_i$  of any parallel copy of  $T_i^c$  whose body lies in  $\partial M \times [0, 2/6)$ . From now on we will only consider such parallel copies of  $T_i^c$ .

Apply grope height raising [FQ90, Proposition 2.7] (see also [FP21, Proposition 17.1 and Lemma 17.7]) to the capped gropes  $\{D_i''\}$ , to produce a collection of capped gropes  $\{G_i^c\}$  with the same framed attaching region as the  $\{D_i''\}$ , arbitrarily many surface stages, and pairwise disjoint caps with algebraically cancelling double points. Additionally, parallel copies of the surfaces  $\{T_i^c\}$  with bodies in  $\partial M \times [0, 2/6)$  are geometrically dual to  $\{G_i^c\}$ . Moreover, the first two stages of each  $G_i^c$  coincide with the first two stages of  $D_i''^c$ , for each  $i$ . This completes the proof.  $\square$

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. 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 4.2.** *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  $\{T_i^c\}$ , such that the body  $T_i$  of each  $T_i^c$  is contained in a collar neighbourhood  $\partial M \times [0, 2/6)$  of  $\partial M$ . Suppose that otherwise only boundary collars of the attaching regions of the  $G_i^c$  and the caps of the  $T_i^c$  lie in  $\partial M \times [0, 2/6)$ .*

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

**Remark 4.3.** At first glance Lemma 4.2 seems remarkably close to [FQ90, Proposition 3.3], which claims to begin with a collection of properly immersed discs in a 4-manifold  $M$  equipped with a collection of  $\pi_1$ -null geometrically transverse capped surfaces, and replace this with 1-storey capped towers with arbitrary grope height, the same framed attaching region, and equipped with geometrically dual spheres. However, the proof of [FQ90, Proposition 3.3] does not construct the claimed geometrically transverse spheres. We also note that the definition of geometrically transverse capped surfaces in [FQ90] allows intersections with the caps, which is different from our notion of dual capped surfaces in Definition 3.1.

*Proof.* Consider the union of the top 1.5 stages of the gropes  $\{G_i^c\}$ . Since  $\pi_1(M)$  is good (Definition 3.2) 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  $\{\tilde{G}_i^c\}$  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.

Null homotopies for the double point loops produce immersed discs  $\{\tilde{\delta}_\alpha\}$  bounded by the double point loops of the new caps. These new discs may be assumed to miss  $\{T_i\}$ , since any null homotopy may be pushed off the collar  $\partial M \times [0, 2/6)$ , and since the boundaries of these discs, the double point loops, are disjoint from  $\partial M \times [0, 2/6)$  because of the hypothesis that  $\tilde{G}_i^c \cap (\partial M \times [0, 2/6))$  consists of a boundary collar of the attaching region of  $\tilde{G}_i^c$ , for each  $i$ . On the other hand the discs coming from the null homotopies might intersect  $\{\tilde{G}_i^c\}$  arbitrarily.

Contract parallel copies of  $\{T_i^c\}$ , with bodies in  $\partial M \times (1/6, 2/6)$ , to produce a family of spheres  $\{S_i''\}$  geometrically dual to  $\{G_i^c\}$ . Boundary twist the discs  $\{\tilde{\delta}_\alpha\}$  to achieve the correct framing and then push down and tube into  $\{S_i''\}$  to remove any intersections of the resulting discs with  $\{\tilde{G}_i^c\}$ . Glue the resulting discs  $\{\delta_\alpha\}$  to  $\{G_i^c\}$  to produce the 1-storey capped towers  $\{\mathcal{T}_i^c\}$ .

Now consider parallel copies of the  $\{T_i^c\}$  with bodies in  $\partial M \times (0, 1/6)$ . At this point the caps of  $\{T_i^c\}$  only (possibly) intersect  $\{\mathcal{T}_i^c\}$  in the tower caps, and the body  $T_i$  is geometrically dual to  $T_i^c$ . Contract each  $T_i^c$  along its caps and call this family of spheres  $\{R_i\}$ . Push all intersections with tower caps off the contraction. The family of dual spheres  $\{R_i\}$  is then geometrically dual to the resulting 1-storey capped towers  $\{\mathcal{T}_i^c\}$ , as desired.  $\square$

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 4.4.** *Let  $N$  be a 4-manifold and let*

$$D = (D_1, \dots, D_m): (D^2 \sqcup \dots \sqcup D^2, S^1 \sqcup \dots \sqcup S^1) \rightarrow (N, \partial N)$$

*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  $T^c \subseteq N \setminus \bigcup_{i=1}^m \nu D_i$  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  $S: S^2 \rightarrow N \setminus \bigcup_{i=1}^m \nu D_i$  be the 2-sphere obtained by contracting  $T^c$ . Then  $\iota_*[S] = 0 \in \pi_2(N)$ , where  $\iota: N \setminus \bigcup_{i=1}^m \nu D_i \rightarrow N$  is the inclusion map.*

*Proof.* By Lemma 3.4, 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.  $\square$

## 5. PROOF OF THE DISC EMBEDDING THEOREM

In this section we prove Theorem A, which we recall below.

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

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

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

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

*such that  $\bar{F}$  has the same boundary as  $F$  and admits a generically immersed, geometrically dual collection of framed spheres  $\{\bar{g}_i\}_{i=1}^k$ , such that  $\bar{g}_i$  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  $\bar{f}_i$  may be assumed to induce the same framing.*

*Proof.* By Theorem 1.5, we may assume that the collections  $\{f_i\}$  and  $\{g_i\}$  are generically immersed and transverse. By adding local cusps if necessary, assume that  $\mu(g_i) = 0$  for each  $i$ . Tube each intersection and self-intersection within  $\{f_i\}$  into  $\{g_i\}$  using the unpaired intersection point (after having chosen a pairing of all but one of the  $f_i$ - $g_i$  intersection points by Whitney discs, which is possible since  $\lambda(f_i, g_i) = 1$ ). We then obtain a collection of discs, which we still call  $\{f_i\}$ , where  $\lambda(f_i, f_j) = \mu(f_i) = 0$  and  $\lambda(f_i, g_j) = \delta_{ij}$  for each  $i, j$ . Note also that the framing on the boundary of each  $f_i$  is unchanged since all  $g_j$  are framed. Apply the geometric Casson lemma (Lemma 3.3) to arrange that  $\{f_i\}$  and  $\{g_i\}$  are geometrically dual, after a regular homotopy.

Since  $\lambda(f_i, f_j) = \mu(f_i) = 0$  for all  $i, j$ , the intersections and self-intersections among the  $\{f_i\}$  can be paired up with framed Whitney discs  $\{D_j\}$ . By tubing into the geometric duals  $\{g_i\}$  we can assume that the interiors of  $\{D_j\}$  lie in the complement of the  $\{f_i\}$ .

Let  $M' := M \setminus \bigcup \nu(f_i)$ . We will apply Lemmas 4.1 and 4.2 in  $M'$ , so we check the hypotheses. The collection  $\{f_i\}$  is  $\pi_1$ -negligible in  $M$  so  $\pi_1(M') \cong \pi_1(M)$  is good. Let  $T_j$  be the Clifford torus



at one of the double points paired by  $D_j$ . Cap each  $T_j$  with meridional discs to  $\{f_i\}$ , then tube the unique intersection point of each cap with  $\{f_i\}$  into parallel copies of the dual spheres  $\{g_i\}$ . The resulting capped surfaces  $\{T_j^c\}$  lie in  $M'$  as desired. The bodies lie in a collar neighbourhood of  $\partial M'$ , and are geometrically dual to  $\{D_j\}$ , while the caps have algebraically cancelling intersections. Contract a parallel copy of each  $T_j^c$  to produce a dual sphere  $S_j$  for  $D_j$  in  $M'$ , and push the collection  $\{D_j\}$  off the contraction. Remove the intersections between the caps of  $\{T_j^c\}$  and  $\{D_j\}$  by tubing into parallel copies of  $\{S_j\}$ . The resulting caps for  $\{T_j^c\}$  no longer intersect  $\{D_j\}$ .

Now apply Lemmas 4.1 and 4.2 to  $M'$ , to replace the discs  $\{D_j\}$  with 1-storey capped towers  $\{T_j^c\}$  whose framed attaching regions coincide with the framed boundary of  $\{D_j\}$  and that have geometrically dual spheres  $\{R_j\}$ . Remove intersections of  $\{g_i\}$  with  $\{T_j^c\}$ , by pushing down into the base surface and tubing into  $\{R_j\}$ . The resulting spheres  $\{\bar{g}_i\}$  are disjoint from  $\{T_j^c\}$  and geometrically dual to  $\{f_i\}$ . By [FQ90, Chapters 3 and 4] (see also [MP21, Chapter 17; BKK<sup>+</sup>21, Part IV]) every disc-like 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. Applied to the collection  $\{T_j^c\}$ , this produces mutually disjoint, embedded and framed Whitney discs pairing the intersections and self-intersections of the  $\{f_i\}$  away from  $\{\bar{g}_i\}$ . Whitney moves guided by these discs produce embedded discs  $\{\bar{f}_i\}$  geometrically dual to  $\{\bar{g}_i\}$ . In the case that  $f_i$  was initially generically immersed, we obtain the same framing on the boundary of  $\bar{f}_i$  since these are regular homotopies in the interior.

It remains only to argue that each  $\bar{g}_i$  is homotopic in  $M$  to  $g_i$  for each  $i$ . Observe that to obtain  $\bar{g}_i$ , 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  $\{T_j^c\}$ .

However,  $R_j$  was obtained by contracting  $T_j^c$ , whose body is a Clifford torus for an intersection point among the  $\{f_i\}$ . Therefore by Lemma 4.4,  $R_j$  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  $\bar{g}_i$  is homotopic in  $M$  to  $g_i$ , as desired. This completes the proof.  $\square$

## 6. GENERIC IMMERSIONS IN SMOOTH AND TOPOLOGICAL 4-MANIFOLDS

The goal of this section is to prove Theorem 1.5, rectifying another omission in [FQ90].

**Remark 6.1.** During the proof we will use results that rely on the smooth-input disc embedding theorem. Therefore the results of this section rely on a version of Theorem A where  $M$  is assumed to be a smooth 4-manifold. Proposition 6.2 was used in the proof of Theorem A with a purely topological input, in order to find generic immersions. The logical dependencies in the development of topological 4-manifold topology are elucidated in [PR21b].

Recall that for a compact surface  $\Sigma$  and 4-manifold  $M$ , a *generic immersion* in the smooth category, written  $f: \Sigma^2 \looparrowright M^4$ , 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  $\Sigma$  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  $C^\infty(\Sigma, M)$ . To see this, apply [Hir76, Theorem 2.2.12] which shows that immersions are dense in the space of smooth maps, and then [GG73, Chapter III, Corollary 3.3] shows that generic immersions are dense in the space of immersions. 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  $\Sigma$  and post-composing with a diffeomorphism of  $M$ , with both diffeomorphisms isotopic to the identity.

It is well known that every continuous map between smooth manifolds is arbitrarily close to a smooth map (see for example [Lee03, Theorem 10.21] or [Hir76, Theorem 2.2.6]). 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 smooth maps  $\Sigma \rightarrow M$  is said to be generic if the corresponding map  $\Sigma \times [0, 1] \rightarrow M \times [0, 1]$  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  $H_t: \Sigma \rightarrow M$  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  $\Sigma \rightarrow M$  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.

Our goal in this section is to state and prove purely topological analogues of these smooth facts. Here is the main technical result.

**Proposition 6.2.** *Let  $\Sigma$  be a compact surface and let  $M$  be a 4-manifold.*

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

*Proof.* First we prove (1), following [FQ90, Corollary 9.5C]. We may assume without loss of generality that  $M$  is connected. Choose a smooth structure on  $\Sigma$ . 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  $M \setminus \{p\}$ .

If the point  $p$  can be chosen disjoint from the image  $f(\Sigma)$  in (1), and disjoint from  $H(\Sigma \times [0, 1])$  in (2), then we have a continuous map, and the result follows from smooth approximation and general position. The aim of the rest of the proof is therefore to arrange that  $f$  (respectively  $H$ ) can be arranged to have image missing a point. The proof given in [FQ90, Corollary 9.5C] assumes that this holds without comment.

Homotope the restriction of  $f$  to  $\partial\Sigma$  to a smooth embedding, and extend this to a homotopy of  $\Sigma$  supported in a collar near the boundary (or use a given embedding to start with and work rel. boundary). Consider the smooth surface  $\Sigma_p := \Sigma \setminus f^{-1}(p)$ , which comes with a proper map  $f|: \Sigma_p \rightarrow M \setminus \{p\}$  that is properly homotopic to a smooth proper map  $f'$ . By Sard's theorem, some point  $q$  in the interior of  $M \setminus \{p\}$  does not lie in the image of  $f'$ . Moreover, as  $\Sigma$  is compact and the homotopy was proper,  $f'$  maps each end of  $\Sigma_p$  to  $p$  (as did  $f$ ). Send all of  $f^{-1}(p)$  to  $p$ , to extend  $f'$  to a continuous map  $\Sigma \rightarrow M \setminus \{q\}$ , 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  $f'': \Sigma \looparrowright M \setminus \{q\}$ . 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  $H: \Sigma \times [0, 1] \rightarrow M$  whose restriction  $H|: \Sigma \times \{i\} \rightarrow M$  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

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

The proper map  $(\Sigma \times [0, 1])_q \rightarrow M \setminus \{q\}$  is properly homotopic rel.  $\Sigma \times \{0, 1\}$  to a proper smooth map  $H': (\Sigma \times [0, 1])_q \rightarrow M \setminus \{q\}$ , 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  $H'': \Sigma \times [0, 1] \rightarrow M \setminus \{r\}$ . Choose a smooth structure on  $M \setminus \{r\}$ , such that  $H| = H''|: \Sigma \times \{0, 1\} \rightarrow M$  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  $M \setminus \{r\}$ .  $\square$

**Remark 6.3.** Let us discuss some consequences of Proposition 6.2 for normal bundles of generic immersions. A generic immersion  $f: \Sigma \looparrowright M$  has a linear normal bundle  $\nu(f) \rightarrow \Sigma$ , in both the smooth and topological categories [FQ90, Section 9.3]. However Proposition 6.2 tells us that  $f$  is generically homotopic to a smooth immersion in some smooth structure on  $M \setminus \{q\}$ , so this gives an easier proof that up to homotopy  $f$  has a linear normal bundle. Then  $f$  comes with a map  $\nu(f) \rightarrow M$  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  $\Sigma$  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  $\Sigma$  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 1.5 about homotopy classes  $[\Sigma, M]$  of maps  $f: \Sigma \rightarrow M$ , when  $\Sigma$  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  $\Sigma$ , 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  $\Sigma$  comes with a whisker to the basepoint. The proof will use topological transversality, which we state first.

**Theorem 6.4** ([Qui82, Qui88] (see also [FQ90, Section 9.5])). *Let  $\Sigma_1$  and  $\Sigma_2$  be locally flat proper submanifolds of a topological 4-manifold  $M$  that are transverse to  $\partial M$ . There is an isotopy of  $M$ , supported in any given neighbourhood of  $\Sigma_1 \cap \Sigma_2$ , taking  $\Sigma_1$  to a submanifold  $\Sigma'_1$  transverse to  $\Sigma_2$ .*

**Theorem 1.5.** *Let  $\Sigma$  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*

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

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  $\Sigma$ . 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 1.5 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.

*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 6.2(1) the map is surjective.

For injectivity, consider generic immersions  $F, F': \Sigma \looparrowright M$ , 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  $M \setminus \{q\}$ .

Now by Proposition 6.2(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.  $\square$

## REFERENCES

- [BKK<sup>+</sup>21] S. Behrens, B. Kalmár, M. H. Kim, M. Powell, and A. Ray (eds.), *The disc embedding theorem*, Oxford University Press, 2021.
- [Cas86] A. Casson, *Three lectures on new infinite constructions in 4-dimensional manifolds*, à la recherche de la topologie perdue, 1986, pp. 201–244. With an appendix by L. Siebenmann.
- [Don83] S. K. Donaldson, *An application of gauge theory to four-dimensional topology*, J. Differential Geom. **18** (1983), no. 2, 279–315.
- [FP21] P. Feller and M. Powell, *Grope height raising and 1-storey capped towers*, The disc embedding theorem, 2021.
- [FQ90] M. Freedman and F. Quinn, *Topology of 4-manifolds*, Princeton Mathematical Series, vol. 39, Princeton University Press, 1990.
- [Fre82] M. Freedman, *The topology of four-dimensional manifolds*, J. Differential Geom. **17** (1982), no. 3, 357–453.
- [Fre84] M. H. Freedman, *The disk theorem for four-dimensional manifolds*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983), 1984, pp. 647–663.
- [FT05] S. Friedl and P. Teichner, *New topologically slice knots*, Geom. Topol. **9** (2005), 2129–2158.
- [FT95] M. H. Freedman and P. Teichner, *4-manifold topology. I. Subexponential groups*, Invent. Math. **122** (1995), no. 3, 509–529.
- [GG73] M. Golubitsky and V. Guillemin, *Stable mappings and their singularities*, Springer-Verlag, New York-Heidelberg, 1973. Graduate Texts in Mathematics, Vol. 14.
- [GT04] S. Garoufalidis and P. Teichner, *On knots with trivial Alexander polynomial*, J. Differential Geom. **67** (2004), no. 1, 167–193.
- [Hir76] M. W. Hirsch, *Differential topology*, Vol. 33, Springer, Cham, 1976.
- [HK88] I. Hambleton and M. Kreck, *On the classification of topological 4-manifolds with finite fundamental group*, Math. Ann. **280** (1988), no. 1, 85–104.
- [HK93] I. Hambleton and M. Kreck, *Cancellation, elliptic surfaces and the topology of certain four-manifolds*, J. Reine Angew. Math. **444** (1993), 79–100.
- [HKT09] I. Hambleton, M. Kreck, and P. Teichner, *Topological 4-manifolds with geometrically two-dimensional fundamental groups*, J. Topol. Anal. **1** (2009), no. 2, 123–151.
- [KM19] M. R. Klug and M. Miller, *Concordance of surfaces and the Freedman-Quinn invariant*, 2019. arxiv:1912.12286.
- [KOPR21] M. H. Kim, P. Orson, J. Park, and A. Ray, *Good groups*, The disc embedding theorem, 2021.
- [KPR22] D. Kasprowski, M. Powell, and A. Ray, *Counterexamples in 4-manifold topology*, EMS Surv. Math. Sci. **9** (2022), no. 1, 193–249.
- [KQ00] V. S. Krushkal and F. Quinn, *Subexponential groups in 4-manifold topology*, Geom. Topol. **4** (2000), 407–430.
- [Kre99] M. Kreck, *Surgery and duality*, Ann. of Math. (2) **149** (1999), no. 3, 707–754.
- [Lee03] J. M. Lee, *Introduction to smooth manifolds*, Graduate Texts in Mathematics, vol. 218, Springer-Verlag, New York, 2003.
- [MP21] A. N. Miller and M. Powell, *Tower height raising and embedding*, The disc embedding theorem, 2021.
- [OPR21] P. Orson, M. Powell, and A. Ray, *The s-cobordism theorem, the sphere embedding theorem and the poincaré conjecture*, The disc embedding theorem, 2021.
- [PR21a] M. Powell and A. Ray, *Basic geometric constructions*, The disc embedding theorem, 2021.
- [PR21b] M. Powell and A. Ray, *The development of topological 4-manifold theory*, The disc embedding theorem, 2021.
- [PR21c] M. Powell and A. Ray, *Gropes, towers, and skyscrapers*, The disc embedding theorem, 2021.
- [Qui82] F. Quinn, *Ends of maps. III. Dimensions 4 and 5*, J. Differential Geom. **17** (1982), no. 3, 503–521.
- [Qui86] F. Quinn, *Isotopy of 4-manifolds*, J. Differential Geom. **24** (1986), no. 3, 343–372.
- [Qui88] F. Quinn, *Topological transversality holds in all dimensions*, Bull. Amer. Math. Soc. (N.S.) **18** (1988), no. 2, 145–148.
- [Sat89] Y. Sato, *Smooth 2-knots in  $S^2 \times S^2$  with simply-connected complements are topologically unique*, Proc. Amer. Math. Soc. **105** (1989), no. 2, 479–485.
- [Sat91] Y. Sato, *Locally flat 2-knots in  $S^2 \times S^2$  with the same fundamental group*, Trans. Amer. Math. Soc. **323** (1991), no. 2, 911–920.
- [Spr72] D. Spring, *Proper homotopy and immersion theory*, Topology **11** (1972), 295–305.
- [Sto94] R. Stong, *Uniqueness of connected sum decompositions in dimension 4*, Topology and its applications **56** (1994), 277–291.
- [Tei97] P. Teichner, *On the star-construction for topological 4-manifolds*, Geometric topology (Athens, GA, 1993), 1997, pp. 300–312.
- [Wal99] C. T. C. Wall, *Surgery on compact manifolds*, Second, Mathematical Surveys and Monographs, vol. 69, American Mathematical Society, Providence, RI, 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** (1943), 161–172.
- [Whi44] H. Whitney, *The singularities of a smooth  $n$ -manifold in  $(2n - 1)$ -space*, Ann. of Math. (2) **45** (1944), 247–293.

SCHOOL OF MATHEMATICS AND STATISTICS, UNIVERSITY OF GLASGOW, UNITED KINGDOM  
*Email address:* `mark.powell@glasgow.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`