OXFORD

THE DISC EMBEDDING THEOREM

Edited by

STEFAN BEHRENS
BOLDIZSÁR KALMÁR
MIN HOON KIM
MARK POWELL
ARUNIMA RAY

THE DISC EMBEDDING THEOREM

The Disc Embedding Theorem

Edited by

STEFAN BEHRENS, BOLDIZSÁR KALMÁR,
MIN HOON KIM, MARK POWELL,
AND ARUNIMA RAY

Contributors: Stefan Behrens, Xiaoyi Cui, Christopher W. Davis,
Peter Feller, Boldizsár Kalmár, Daniel Kasprowski, Min Hoon Kim, Duncan McCoy,
Jeffrey Meier, Allison N. Miller, Matthias Nagel, Patrick Orson, JungHwan Park,
Wojciech Politarczyk, Mark Powell, Arunima Ray, Henrik Rüping,
Nathan Sunukjian, Peter Teichner, and Daniele Zuddas

with an afterword by

MICHAEL H. FREEDMAN

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2021
Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2021937963

ISBN 978–0–19–884131–9

DOI: 10.1093/oso/9780198841319.001.0001

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

PREFACE

In 1982, Michael Freedman, building upon the ideas and constructions of Andrew Casson, proved the h-cobordism theorem and the exactness of the simply connected surgery sequence in dimension four, deducing a classification theorem for topological 4-manifolds, a special case of which was the 4-dimensional topological Poincaré conjecture.

The key ingredient in his proof is the *disc embedding theorem*. In manifolds of dimension five and higher, generic maps of discs are embeddings, whereas in dimension four such maps have isolated double points, preventing the high-dimensional proofs from applying. Freedman showed how to embed discs in simply connected 4-manifolds, revealing that in certain situations topological 4-manifolds behave like higher-dimensional manifolds. Contemporaneous results of Simon Donaldson showed that smooth 4-manifolds do not. Indeed, dimension four exhibits a remarkable disparity between the smooth and topological categories, as demonstrated by the existence of exotic smooth structures on $\mathbb{R}^4$, for example.

Freedman and Donaldson both received the Fields Medal in 1986 for their contributions to the understanding of 4-manifolds. Soon after Freedman's work appeared, Frank Quinn expanded on the techniques of Freedman, proving foundational results for topological 4-manifolds, such as transversality and the existence of normal bundles for locally flat submanifolds. The work of Freedman and Quinn was collected in the book [FQ90], which became the canonical source for topological 4-manifolds in the decades that followed.

The Origin of This Book

In January and February of 2013, Freedman gave a series of 12 lectures at the University of California Santa Barbara (UCSB) in the USA with the goal of explaining his proof of the disc embedding theorem. The lectures were broadcast live to the Max-Planck-Institut für Mathematik (MPIM) in Bonn, Germany as part of the *Semester on 4-manifolds and their combinatorial invariants* organized by Matthias Kreck and Peter Teichner, where Quinn and Teichner ran supplementary discussion sessions. Robert Edwards, in the UCSB audience, not only contributed various remarks but also stepped in as a guest lecturer and presented his perspective on a key step of the proof, namely the construction of 'the design'. The lectures were recorded and are currently available online at www.mpim-bonn.mpg.de/FreedmanLectures.

This book began as annotated transcripts of Freedman's lectures typed by Stefan Behrens. In May and June of 2013, the rough draft of the notes was revised and augmented in a collaborative effort of the MPIM audience, coordinated by Behrens and Teichner. The following people were involved in this process: Xiaoyi Cui, Matthew Hogancamp, Daniel Kasprowski, Ju A. Lee, Wojciech Politarczyk, Mark Powell, Henrik Rüping, Nathan Sunukjian, and Daniele Zuddas.

Three years later, in November and December of 2016, Peter Feller and Mark Powell organized a seminar on the disc embedding theorem at the Hausdorff Institute for Mathematics (HIM) in Bonn. This included screenings of Freedman's UCSB lectures on decomposition space theory and a series of talks by Powell, along with guest lectures by Stefan Behrens, Peter Feller, Boldizsár Kalmár, Allison N. Miller, and Daniel Kasprowski on the constructive part of the proof following the approach in the book by Freedman and Quinn [FQ90]. The HIM audience included many of the participants in the Junior Trimester Program in Topology: Christopher W. Davis, Peter Feller, Duncan McCoy, Jeffrey Meier, Allison N. Miller, Matthias Nagel, Patrick Orson, JungHwan Park, Mark Powell, and Arunima Ray. Together, the speakers and the audience revised the structure of the 2013 notes, fleshing out many details and rewriting certain parts from scratch. From 2017 to 2020, Kalmár, Kim, Powell, and Ray synthesized the individual contributions of the authors into the artefact you presently behold. New chapters on good groups, the applications of the disc embedding theorem to surgery and the Poincaré conjecture, the development of topological 4-manifold theory, and remaining open problems were written. During this period, Kim and Miller, in particular, created the many computerized figures appearing throughout the book.

This text follows Freedman's introduction to decomposition space theory in his 2013 lectures in Part I, before giving a complete proof of the disc embedding theorem in Parts II and IV. The latter parts follow the 2016 lectures based on [FQ90], although they are naturally based on the ideas learnt from Freedman's original lectures and the concurrent explanations and guest lectures by Edwards, Quinn, and Teichner. In particular, we give a detailed new description of tower embedding and the design. Part III contains a discussion of major applications and conjectures related to the disc embedding theorem. It describes how to use the disc embedding theorem to prove the s-cobordism theorem, the Poincaré conjecture, the exactness of the surgery sequence in dimension four for good groups, and the topological classification of simply connected closed 4-manifolds.

Since so much of 4-dimensional topological manifold theory rests on the seminal work of Freedman, it has been felt by the community that another independent and rigorous account ought to exist. We hope that this manuscript will make this high point in 4-manifold topology accessible to a wider audience.

Casson Towers

We choose to follow the proof from [FQ90], using gropes, which differs in many respects from Freedman's original proof using Casson towers [Fre82a]. The infinite construction using gropes, which we call a *skyscraper*, simplifies several key steps of the proof, and the known extensions of the theory to the non-simply connected case rely on this approach. Readers interested in Casson towers should refer to the MPIM videos of Freedman's 2013 lectures, where he explained much about Casson towers and their use in the original proof. Apart from [Fre82a], further literature on Casson towers includes [GS84, Biž94, Sie82, CP16]. Moreover, the combination of [Sie82] and the Casson tower embedding theorem from [GS84] gives another account of the original Casson towers proof from [Fre82a].

Differences

We briefly indicate, for the experts, the salient differences between the proof given in this book and that given in [FQ90]. First, there is a slight change in the definition of towers (and therefore of skyscrapers), which we point out precisely in Remark 12.8. With our definition, it is clear that the corresponding decomposition spaces are mixed ramified Bing–Whitehead decompositions. This possibility was mentioned in [FQ90, p. 238].

Additionally, the statement of the disc embedding theorem in [FQ90] asserts that immersed discs, under certain conditions including the existence of framed, algebraically transverse spheres, may be replaced by flat embedded discs with the same boundary and geometrically transverse spheres. The proofs given in [Fre82a, FQ90] produce the embedded discs but not the geometrically transverse spheres. We remedy this omission by modifying the start of the proof given in [FQ90], as in [PRT20]. The geometrically transverse spheres are essential for the sphere embedding theorem, which is the key result used in the application of the disc embedding theorem to surgery for topological 4-manifolds with good fundamental group and the classification of simply connected, closed, topological 4-manifolds, as we describe in Chapter 22. We also observe that the geometrically transverse spheres in the output are homotopic to the algebraically transverse spheres in the input [PRT20]. Besides these points, the proof of the disc embedding theorem given in this book only differs from that in [FQ90] in the increased amount of detail and number of illustrations.

We largely focus on the first few chapters of [FQ90]. In particular, we assume that the ambient 4-manifold is smooth. We do not delve into the work of Quinn on the smoothing theory of noncompact 4-manifolds, the annulus theorem, transversality, or normal bundles for locally flat submanifolds, instead describing these developments broadly in Chapter 1, and in more detail in Chapter 21.

Seminar Organization

The majority of the chapters in this book may be covered in a single seminar talk each. We expect that Parts II and IV, even without going through all the details in Part IV, will require a semester. We therefore suggest the following alternative to the standard approach. After using Chapters 1 and 2 to provide context, work through Parts II and IV alongside group viewings of the videos of Freedman's UCSB lectures 2–5, which discussed the decomposition space theory of Part I. The exposition in Part I of this book should supply enough additional detail to support the lectures, and it adds to the charm of learning this mathematics to watch the man himself explain it. This also allows Parts I and II to be covered simultaneously. In the latter part of the seminar, results from both can be combined for the proof that skyscrapers are standard in Part IV. Part III is not directly applicable to the proof of the disc embedding theorem and may be safely skipped in the first reading.

Credit

This manuscript is the outcome of a collaborative project of many mathematicians, as described earlier. After Freedman, who of course gave the original lectures and proved the disc embedding theorem in the first place, and Stefan Behrens, who typed up the initial draft, many people contributed to improving individual chapters, or in some cases developing them from scratch. We therefore attribute each chapter to those who contributed the bulk of the work towards it, whether through a new lecture that they wrote and delivered, polishing the exposition, creating original pictures, adding new material to fill in details that could not be covered in the lectures, or writing a chapter on their own by combining information from various sources in the literature.

Apart from the authors, the project benefitted from the input of Bob Edwards and Frank Quinn, as well as Jae Choon Cha, Diarmuid Crowley, Jim Davis, Stefan Friedl, Bob Gompf, Chuck Livingston, Michael Klug, Matthias Kreck, Christian Kremer, Slava Krushkal, Andy Putman, Ben Ruppik, and András Stipsicz.

CONTENTS

LIST OF FIGURES

1

Context for the Disc Embedding Theorem

STEFAN BEHRENS, MARK POWELL, AND ARUNIMA RAY

1.1 Before the Disc Embedding Theorem

1.1.1 High-dimensional Surgery Theory

By 1975, classification problems for manifolds of dimension n at least five, be they smooth, piecewise linear (PL), or topological, had been translated into questions in homotopy theory and algebra. For each of these categories, classification problems are typically of two types: the *existence problem* concerns the existence of a manifold within a given homotopy type, while the *uniqueness problem* concerns the number of such manifolds up to isomorphism. The input for such questions is a *Poincaré complex*—roughly speaking a finite cell complex that satisfies n-dimensional Poincaré duality for some n.

Fix the category *CAT* to be either smooth, PL, or topological. Two closed n-manifolds are said to be *h-cobordant* if they cobound an $(n+1)$-manifold such that the inclusion of each boundary component is a homotopy equivalence. The *structure set* of a given Poincaré complex X, denoted by $\mathcal{S}(X)$, is the set of n-dimensional closed manifolds M along with a homotopy equivalence $M \to X$, up to h-cobordism, where the cobordism has a compatible map to X. For Poincaré complexes of dimension at least five, *surgery theory* can decide if $\mathcal{S}(X)$ is nonempty, and if so, can compute it explicitly using algebraic topology, at least in favourable circumstances [Bro72, Nov64, Sul96, Wal99, KS77]. More precisely, the structure set of a Poincaré complex X with dimension at least five is nonempty if and only if (i) a certain spherical fibration over X, called the *Spivak normal fibration*, lifts to a *CAT* bundle, and (ii) an L-theoretic surgery obstruction vanishes. This completely answers, at least in principle, the question of whether X is homotopy equivalent to a *CAT* manifold.

Moreover, if the structure set for a Poincaré complex X of dimension n at least five is nonempty, it fits in the following exact sequence of pointed sets, called the Browder–Novikov–Sullivan–Wall *surgery exact sequence*:

$$L_{n+1}(\mathbb{Z}[\pi_1(X)]) \to \mathcal{S}(X) \to \mathcal{N}(X) \xrightarrow{\sigma} L_n(\mathbb{Z}[\pi_1(X)]).$$

Stefan Behrens, Mark Powell, and Arunima Ray, *Context for the Disc Embedding Theorem* In: *The Disc Embedding Theorem*. Edited by: Stefan Behrens, Boldizsár Kalmár, Min Hoon Kim, Mark Powell, and Arunima Ray, Oxford University Press.
 DOI: 10.1093/oso/9780198841319.003.0001

Here $\mathcal{N}(X)$ denotes the set of *normal invariants* of X, namely bordism classes of degree one maps from some n-manifold to X, together with normal bundle data. Via transversality, this can be computed using homotopy theory. The *L-groups* are purely algebraic and depend only on the group $\pi_1(X)$, the orientation character, and the residue of n modulo 4.

Let us describe the existence programme in more detail. Assuming that the Spivak normal fibration on X lifts to a *CAT*-bundle, a choice of lift gives rise to an element of $\mathcal{N}(X)$, namely a closed manifold N together with a degree one map to X that respects the normal data corresponding to the chosen lift. We wish to improve such an element to a manifold M equipped with a homotopy equivalence to X, at the expense of modifying N by the process of *surgery*. An *elementary surgery* consists of finding an embedded $S^p \times D^q$ within a $(p+q)$-dimensional manifold, cutting out its interior, and gluing in $D^{p+1} \times S^{q-1}$ along its boundary instead. This process kills the homotopy class represented by $S^p \times \{0\}$ and therefore can assist in achieving a given homotopy type. The main theorem of surgery theory says that such a sequence of elementary surgeries on N can produce a manifold homotopy equivalent to X if and only if the obstruction in $L_n(\mathbb{Z}[\pi_1(X)])$ associated with N vanishes. This is encoded by the map σ in the sequence above. In other words, every element of $\sigma^{-1}(0)$ can be modified by surgery to produce an element of the structure set $\mathcal{S}(X)$, namely a closed manifold M equipped with a homotopy equivalence to the Poincaré complex X. This argument shows that, for Poincaré complexes of dimension at least five, we have a procedure for deciding whether the structure set is nonempty—that is, whether the existence problem has a positive resolution.

Exactness of the surgery sequence can be used to calculate the size of the structure set, which addresses part of the uniqueness problem. In order to fully solve the uniqueness problem, we also need to understand when h-cobordant manifolds are isomorphic in the category *CAT*. The *s-cobordism theorem* [Sma62, Bar63, Maz63, Sta67, KS77] (see also [Mil65, RS72]) states that an h-cobordism between closed manifolds of dimension at least five is a product if and only if its Whitehead torsion vanishes. The theorem, which holds for all smooth, PL, and topological manifolds, allows one to obtain uniqueness results.

Its precursor, the *h-cobordism theorem*, states that every *simply connected* h-cobordism between closed manifolds of dimension at least five is a product. This is a straightforward corollary of the s-cobordism theorem, since a simply connected h-cobordism has Whitehead torsion valued in the Whitehead group of the trivial group, which itself vanishes.

Summarizing, by the early 1970s, armed with the powerful tools of the surgery exact sequence and the s-cobordism theorem, topologists had a deep understanding of both the existence and uniqueness problems for manifolds of dimension at least five.

1.1.2 Attempting 4-dimensional Surgery

By contrast, in the early 1970s very little was known about 4-manifolds. Whitehead [Whi49] and Milnor [Mil58] had shown that the homotopy type of a simply connected 4-dimensional Poincaré complex is determined by its intersection form. More precisely, the homotopy types, together with a choice of fundamental class, are in one to one correspondence with isometry classes of unimodular symmetric integral bilinear forms, or, equivalently, congruence classes $A \sim PAP^T$ of symmetric integral matrices with determinant ± 1. So 4-manifold topologists were interested in determining which of these forms are realized by, smooth (equivalently, PL [HM74; FQ90, Theorem 8.3B]) or topological, closed 4-manifolds; whether homotopy equivalent 4-manifolds are s-cobordant; and whether s-cobordant 4-manifolds are *CAT*-isomorphic. Due to its remarkable success in addressing high-dimensional manifolds, surgery theory seemed like a promising tool. However, the main theorems of surgery were not known to hold in dimension four. Similarly, the h- and s-cobordism theorems for 4-manifolds remained open in all three categories.

Let E_8 denote the even 8×8 integer Cartan matrix of the eponymous exceptional Lie algebra; that is,

$$E_8 := \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 2 \end{pmatrix}.$$

This is a symmetric integral matrix with determinant one, and so by the Milnor–Whitehead classification there is a simply connected, 4-dimensional Poincaré complex with intersection form represented by this matrix. Is there a closed 4-manifold homotopy equivalent to this Poincaré complex?

By Rochlin's theorem [Roc52, Kir89], the intersection form of a smooth, closed, spin 4-manifold must have signature divisible by 16. Since E_8 corresponds to an even intersection form, has signature 8, and any simply connected 4-manifold with even intersection form is spin, there cannot be any smooth, closed, simply connected 4-manifold with E_8 as its intersection form. Nevertheless, the question remained: is there a *topological*, closed, simply connected 4-manifold with E_8 as its intersection form? This was an intractable question in the 1970s (refer to Section 1.6 for the answer).

In order to bypass the obstruction from Rochlin's theorem, let us consider the matrix $E_8 \oplus E_8$, which has signature 16. The following is a strategy for constructing a smooth, closed, simply connected 4-manifold with $E_8 \oplus E_8$ as its intersection form. Start with the simply connected 4-manifold K known as the *K3 surface*, given by the solution set for the quartic $x^4 + y^4 + z^4 + w^4 = 0$ in $\mathbb{C}P^3$. Its intersection form is represented by the matrix

$$E_8 \oplus E_8 \oplus H \oplus H \oplus H,$$

where $H = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ is the hyperbolic matrix corresponding to the intersection form of $S^2 \times S^2$ and $\oplus$ denotes the juxtaposition of blocks down the diagonal.

We have the obvious algebraic projection

$$E_8 \oplus E_8 \oplus H \oplus H \oplus H \longrightarrow E_8 \oplus E_8.$$

We would succeed in constructing the desired manifold if this algebraic projection were realized geometrically. That is, we wish to perform surgery on K with the effect of removing the three hyperbolic pairs from the intersection form, resulting in a closed 4-manifold with intersection form $E_8 \oplus E_8$. Let us attempt to do this in the smooth category, and see where and why we fail.

Since K is smooth and simply connected, we know by the Hurewicz theorem that the elements of $H_2(K;\mathbb{Z})$ corresponding to the hyperbolic pairs in the intersection form can be represented by maps $S^2 \to K$, which we can take to be smooth immersions in general position. Henceforth, immersions will be assumed without further comment to be in general position. A single hyperbolic pair is shown schematically on the left of Figure 1.1. According to the matrix H, the two spheres intersect each other algebraically once, but in general there will be excess intersection points geometrically. Additionally, the spheres may only be assumed to be immersed, with algebraically zero self-intersections. Of course, the spheres corresponding to different hyperbolic pairs might have algebraically trivial but geometrically nontrivial intersections as well, but we ignore those for now. If the hyperbolic pair could be represented by framed, embedded spheres which intersect exactly once, such as on the right of Figure 1.1, we could do surgery on either of the two spheres by cutting out a regular neighbourhood (diffeomorphic to $S^2 \times D^2$) and replacing it with $D^3 \times S^1$, with the effect of removing the corresponding hyperbolic matrix from the intersection form. We say that two spheres in an ambient 4-manifold are *geometrically dual* if they intersect at a single point. The existence of the second sphere, geometrically dual to the first, ensures that this surgery would not change the fundamental group of the ambient manifold. For this the second sphere does not need to be embedded. The situation is entirely symmetric: we could do the surgery on an embedding homotopic to the second sphere, with the same effect on homology and the fundamental group.

This strategy is analogous to the idea behind the classification of closed, orientable 2-manifolds, in which we reduce the genus of any given surface by identifying a dual pair of simple closed curves in given homology classes, cutting out an annular neighbourhood of one of them, and filling in the two resulting boundary components with discs; the classification counts the number of such moves needed to produce a sphere. The obstruction

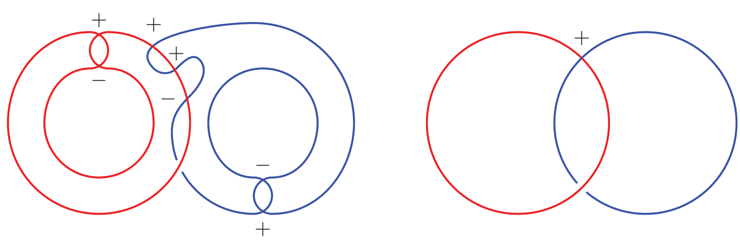

Figure 1.1 Trying to surger a hyperbolic pair. Left: Immersed spheres, depicted schematically, which intersect each other algebraically once but geometrically thrice. Right: The desired situation, where we have embedded spheres which intersect geometrically once.

to carrying out this strategy in dimension four lies in geometrically realizing the algebraic intersection number, passing, as it were, from the left to the right of Figure 1.1. In the smooth category, Donaldson's diagonalization theorem [Don83] (Section 21.2.2) implies that this is a real obstruction, since it shows there is no smooth, closed, simply connected 4-manifold with intersection form $E_8 \oplus E_8$. So we have seen why a naïve attempt to do surgery fails.

For surgery on non-simply connected manifolds, one seeks to remove hyperbolic summands in the equivariant intersection form on $H_2(\widetilde{M})$, the second homology of the universal cover of a closed manifold M, thought of as a module over the group ring $\mathbb{Z}[\pi_1(M)]$. In this context, intersection counts are algebraically trivial if they are trivial over $\mathbb{Z}[\pi_1(M)]$. The principle in such a situation is still the same, namely we wish to represent this algebraic situation geometrically.

1.1.3 Attempting to Prove the s-cobordism Theorem

A similar problem with disjointly embedding 2-spheres occurs when we try to prove the s-cobordism theorem for 5-dimensional cobordisms between 4-manifolds. Let us try to imitate the proof of the high-dimensional smooth s-cobordism theorem, and see what obstructs the strategy from succeeding. Let N be a smooth, compact s-cobordism between two closed 4-manifolds M_0 and M_1; that is, $\partial N = -M_0 \sqcup M_1$, each inclusion $M_i \hookrightarrow N$ is a homotopy equivalence, and the Whitehead torsion $\tau(N, M_0)$ is trivial. Consider a relative handle decomposition of N built on $M_0 \times [0,1]$. Since the Whitehead torsion vanishes, the relative chain complex of finitely generated, free $\mathbb{Z}[\pi_1(N)]$-modules for the pair (N, M_0) can be simplified algebraically so that there are only 2-chains and 3-chains and the boundary map between them is an isomorphism represented by the identity matrix in suitable bases (this might also require some preliminary stabilization in the case of nontrivial fundamental groups). As before, we would like to represent this algebraic situation geometrically.

We find some initial success: since N is connected, we may assume there are no 0-handles or 5-handles, and since N has dimension five and is an h-cobordism, a standard procedure called *handle trading* allows us to trade 1-handles for 3-handles, and 4-handles for 2-handles (see the proof of Theorem 20.1). Thus we see that N is built from $M_0 \times [0,1]$ by attaching only 2-handles and 3-handles, in that order. Since N is an s-cobordism, we arrange by handle slides—possibly after stabilization by adding cancelling 2- and 3-handle pairs—that the 2-handles and 3-handles occur in algebraically cancelling pairs. Let $M_{1/2}$ denote the 4-manifold obtained by attaching the 2-handles to $M_0 \times \{1\} \subseteq M_0 \times [0,1]$. By turning the 3-handles of N upside down, we see that $M_{1/2}$ is also obtained by attaching 2-handles to $M_1 \times \{1\} \subseteq M_1 \times [0,1]$. In other words, $M_{1/2}$ can be obtained from either M_0 or M_1 by a sequence of surgeries on embedded circles. Since the inclusion of M_0 in N induces an isomorphism on fundamental groups, the attaching circles for the 2-handles are null-homotopic in M_0. Similarly, the attaching circles in M_1 are also null-homotopic in M_1. In dimension four, homotopy implies isotopy for loops, and so the surgeries are performed on standard trivial circles. This produces either $S^2 \times S^2$ or $S^2 \widetilde{\times} S^2$ summands in $M_{1/2}$ [Wal99, Lemma 5.5].

The belt spheres $\{0\} \times S^2 \subseteq D^2 \times D^3$ of the 2-handles form a pairwise disjoint collection of framed, embedded 2-spheres in $M_{1/2}$. Each of these spheres has an embedded, geometrically dual sphere coming from pushing the core of the corresponding 2-handle union a null homotopy of the attaching circle into $M_{1/2}$. The latter null homotopy provides an embedded disc, since the attaching circle is trivial. If the framing of the attachment is such that we get an $S^2 \widetilde{\times} S^2$ summand, then this dual sphere need not be framed. Similarly, when we turn the handles upside down, the attaching circles of the 3-handles attached to $M_{1/2}$ become the belt spheres for 2-handles attached to M_1. By the same reasoning as above, the attaching spheres for 3-handles in $M_{1/2}$ form a pairwise disjoint collection of framed, embedded spheres in $M_{1/2}$ equipped with embedded, geometrically dual spheres, which again need not be framed.

Recall that we have arranged that each belt sphere of a 2-handle intersects the attaching sphere of the corresponding 3-handle algebraically once. However, they may intersect multiple times geometrically. A schematic picture for a single pair of a 2-handle belt sphere and 3-handle attaching sphere is shown on the left of Figure 1.2, where, as before, we ignore possible interactions with other pairs. If the 3-handle attaching spheres could be isotoped in $M_{1/2}$ to achieve the situation on the right of the figure for each pair, then the corresponding 2- and 3-handles could be cancelled. Since cancelling all the relative handles of the cobordism (N, M_0) yields the product $M_0 \times [0,1]$, the proof would be complete. However such an isotopy is in general not possible in the smooth category: Donaldson [Don87a] (Section 21.2.2) showed there are h-cobordant, smooth, closed, simply connected 4-manifolds that are not diffeomorphic. So we have seen why imitating the proof of the high-dimensional s-cobordism theorem does not succeed.

In summary, a key input needed in surgery as well as in the proof of the s-cobordism theorem is the ability to remove pairs of algebraically cancelling intersection points between spheres, and thence geometrically realize algebraic intersection numbers. As mentioned above, this is in general not possible smoothly, but for topological 4-manifolds hope remains. We discuss the surgery problem further in Section 1.3.1, and we return to a discussion of the s-cobordism theorem in Section 1.3.2.

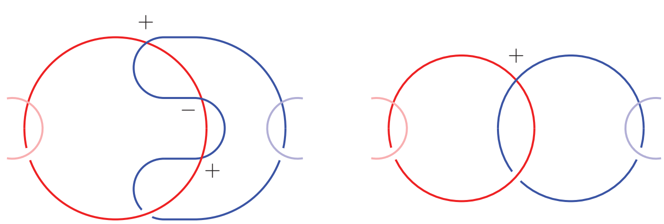

Figure 1.2 An algebraically dual pair consisting of a 2-handle belt sphere (red) and a 3-handle attaching sphere (blue) is shown. The light curves denote the corresponding geometrically dual spheres. Left: The belt sphere and the attaching sphere intersect algebraically once but geometrically thrice. Right: The desired situation where the belt sphere and attaching sphere intersect geometrically once.

1.2 The Whitney Move in Dimension Four

Consider a map of smooth, oriented manifolds $X^d \to Y^{2d}$. In general position, the only singular points are isolated, signed, transverse double points. By inserting local kinks (see Figure 1.3 for a sketch), we can arrange that the sum of the signs of the self-intersection points is zero. In the case of exactly two self-intersection points of opposite sign, the situation is like in the left of Figure 1.4, with two arcs in the image of X joining the two self-intersection points on different sheets. The circle visible in the picture, consisting of two arcs joining the two intersection points, is called a *Whitney circle*. A disc bounded by a Whitney circle is called a *Whitney disc*. Suppose that the Whitney circle bounds an embedded Whitney disc, W, whose interior lies in the exterior of the image of X in Y. Under a condition on the normal bundle of W in Y described in the next paragraph, we can push one sheet of X along W and over the other sheet, as indicated in Figure 1.4, which geometrically cancels the two algebraically cancelling intersection points. This process is called the *Whitney trick* or the *Whitney move* [Whi44].

For $\dim X = d \geq 3$, the Whitney move turns out to be surprisingly simple. If the Whitney circle is null-homotopic in Y, then by general position we can assume it bounds an embedded Whitney disc W whose interior is disjoint from the image of X. Any disc D with boundary a circle C pairing self-intersection points in the image of X determines a $(d-1)$-dimensional sub-bundle of the normal bundle $\nu_{D\subseteq Y}|_C$ of D restricted to C, by requiring that the sub-bundle be normal to one sheet of the image of X and tangent to the other sheet. In order to perform the Whitney move, we need this sub-bundle over the circle C to extend over the entire disc D. Standard bundle theory implies that the

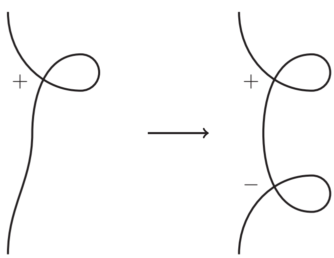

Figure 1.3 Adjusting the algebraic self-intersection number of an immersed submanifold by adding local kinks.

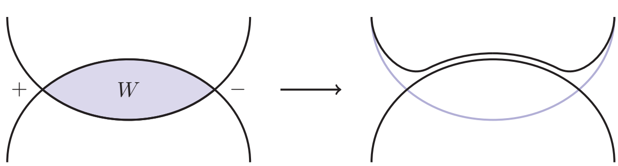

Figure 1.4 The Whitney move. Left: A Whitney disc W is shown in blue. Right: The Whitney move across W removes two intersection points.

sub-bundle extends if and only if it determines the trivial element in $\pi_1(\mathrm{Gr}_{d-1}(\mathbb{R}^{2d-2}))$, where the Grassmannian $\mathrm{Gr}_{d-1}(\mathbb{R}^{2d-2})$ is the space of $(d-1)$-dimensional subspaces in $\mathbb{R}^{2d-2}$. When $d \geq 3$, $\pi_1(\mathrm{Gr}_{d-1}(\mathbb{R}^{2d-2})) \cong \mathbb{Z}/2$ and the nontrivial element corresponds to circles pairing intersection points with the same sign. Since Whitney circles by definition pair intersection points of opposite sign, the sub-bundle in question extends, and we can perform the Whitney move.

Following the strategy outlined in the previous section, in dimensions at least five the availability of the Whitney move is a key ingredient in the proof of the s-cobordism theorem and the efficacy of surgery theory. The Whitney move originated in Whitney's proof [Whi44] of his embedding theorem, which shows that every smooth, compact manifold of dimension d embeds in $\mathbb{R}^{2d}$. The proof finds an immersion of a d-manifold M into $\mathbb{R}^{2d}$ and then improves it to an embedding using the Whitney move. A key step is that the disc guiding the Whitney move can be embedded by general position. This only works for $d \geq 3$, but the Whitney embedding theorem holds for all $d \geq 1$, since compact 1- and 2-dimensional manifolds are classified, and for dimensions 1 and 2 the result can be checked directly.

In contrast to high dimensions, if the ambient dimension is four, even if a Whitney circle is null-homotopic in Y, all we can conclude from general position is that there exists a Whitney disc W whose interior intersects itself and the image of X in isolated points. Moreover, even if an embedded Whitney disc can be found, since $\pi_1(\mathrm{Gr}_1(\mathbb{R}^2)) \cong \mathbb{Z}$, pushing one sheet of X over the other along W may not cancel the intersection points.

Let us investigate the 4-dimensional situation more concretely. Suppose we have two algebraically cancelling intersection points between surfaces P and Q in an ambient 4-manifold. A local model for a transverse intersection between surfaces in a 4-manifold consists of the xy- and the zw-planes meeting at the origin in $\mathbb{R}^4$. A key observation is that the two planes intersect a small 3-sphere around the origin in a Hopf link (see Figure 1.5). A positive (respectively negative) intersection point gives a positive (respectively negative)

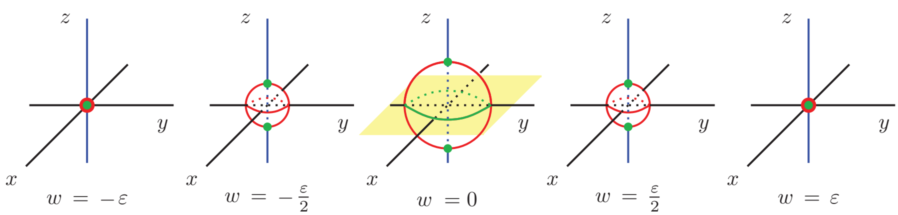

Figure 1.5 The Hopf link at a transverse intersection. Each of the five images above shows the $\mathbb{R}^3$-slice of $\mathbb{R}^4$ corresponding to the w-coordinate, as indicated. Only the central image, where $w = 0$, contains the xy-plane, shown in yellow. The vertical lines in blue trace out the zw-plane, as w is allowed to change. Note that the xy- and zw-planes intersect at the origin. The red spheres, of radius ε in the central image, decrease in radius in either direction until they become points where $w = \pm\varepsilon$. Their union forms a copy of S^3, centred at the origin and of radius ε. The two circles shown in green (one of which appears only as moving points) form a Hopf link, with one component in the xy-plane and the other in the zw-plane. Note that the origin in the far left and far right picture is both red and green.

Hopf link. A neighbourhood of a Whitney circle for P and Q, namely a union of two arcs connecting the algebraically cancelling intersection points, is homeomorphic to $S^1 \times D^3$. The intersection of the boundary of this $S^1 \times D^3$ with P and Q is then the band sum of the two Hopf links corresponding to the two intersection points, where we use one band for each of the two component arcs of the Whitney circle, as shown in Figure 1.6. Note that we have a choice of how many times these bands twist, which corresponds to the choice of framing of the Whitney circle. The correct choice of framing, namely the untwisted framing, yields the *Bing double* of $S^1 \times \{\textit{south pole}\}$ in the solid torus $S^1 \times$ *southern hemisphere* $\subseteq S^1 \times S^2 = \partial S^1 \times D^3$ (see Figure 1.6). Figure 1.7 shows the links we obtain in less than

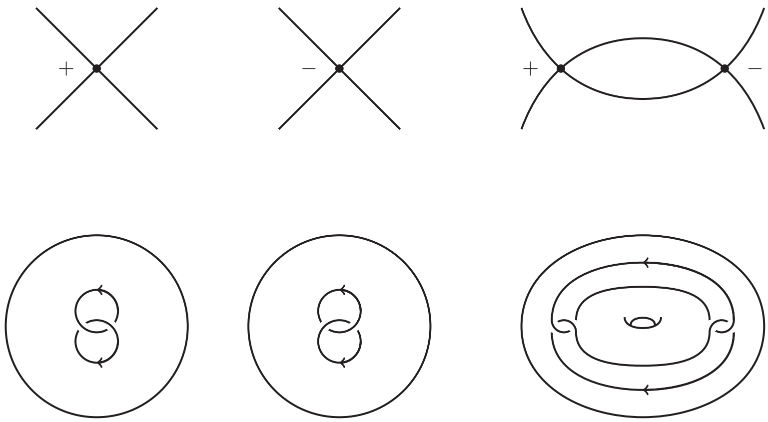

Figure 1.6 Banding together, with untwisted framing, two Hopf links lying in disjoint copies of S^3, around two intersection points produces a Bing double in $S^1 \times D^2 \subseteq S^1 \times S^2$. Each of the three lower pictures corresponds to the picture directly above it.

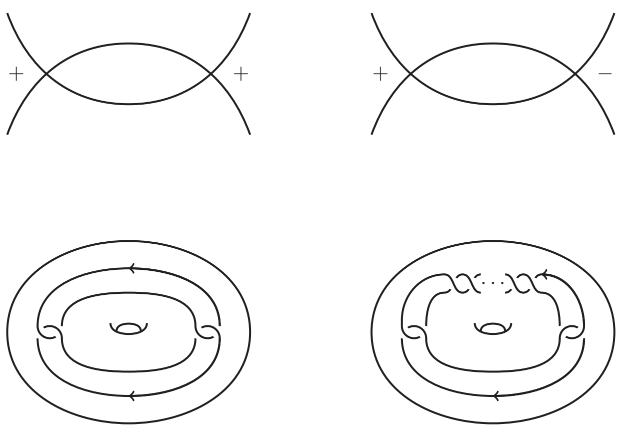

Figure 1.7 Left: When the paired intersection points have the same sign, we get a 2-component link with linking number ± 2. Right: An incorrectly framed Whitney circle produces a twisted Bing double of the Whitney circle.

ideal situations such as when the signs of the intersection points do not cancel or we have the wrong framing; that is, one that does not extend over a Whitney disc.

In the ideal situation, the surfaces P and Q intersect the boundary $S^1 \times S^2$ of a neighbourhood of a Whitney circle in a Bing double of the Whitney circle. If there were an embedded and framed Whitney disc for the Whitney circle, with interior disjoint from P and Q, it would provide a core for an ambient surgery taking $S^1 \times S^2$ to the 3-sphere. Our Bing double would be mapped to a link in this S^3. If this resulting link is the unlink, it is easy to geometrically eliminate the two algebraically cancelling intersection points: cap off the unlink with disjoint embedded discs to obtain disjoint surfaces isotopic to the original ones.

1.3 Casson's Insight: Geometric Duals

We face the problem of finding embedded Whitney discs within 4-manifolds. Indeed, there is no way to locally find such Whitney discs, due to the notion of *slice knots* introduced by Fox and Milnor in the 1950s [FM66], or more accurately, due to the fact that there exist non-slice knots. A knot in $S^3 = \partial D^4$ is said to be topologically (respectively, smoothly) *slice* if it bounds a locally flat (respectively, smoothly) embedded disc in D^4. Slice knots arise, for example, as cross sections of knotted 2-spheres in 4-space. In the 1970s, obstructions to sliceness, for example in terms of the Seifert matrix [Lev69], had already been discovered. At present, a great deal is known about obstruction theory for slice knots; for example, work based on that of Cochran–Orr–Teichner [COT03, COT04], such as [CT07, CHL09, CHL11, Cha14], gives an infinite sequence of obstructions to topological sliceness, building upon the second-order obstructions of Casson and Gordon [CG78]. There also exist several smooth obstructions, including from gauge theory [FS85, HK12], Heegaard–Floer homology [OS03, Hom14, HW16, OSS17], etc. [Ras10, Lob09, LL16, DV16]. Since every knot in S^3 bounds an immersed disc in D^4, but for non-slice knots it is impossible to remove the self-intersection points, we have no hope of removing self-intersections of immersed discs in 4-manifolds in general.

However, in 1974, Casson [Cas86] realized that in the surgery and s-cobordism problems there is global information that may be exploited, namely the fact that the spheres come in algebraically dual pairs. Casson also noticed that given two surfaces in a 4-manifold, there is a relationship between their intersections and the fundamental group of their complement. This is exhibited by the local model of a transverse double point shown in Figure 1.5. Consider the complement of the two intersecting planes in the small ball around the origin that is shown in the figure. This complement deformation retracts to a torus, called the *Clifford torus*, which is the common boundary of enlarged tubular neighbourhoods of the two components of the Hopf link in S^3 from Figure 1.5. So the fundamental group of this complement is $\mathbb{Z} \oplus \mathbb{Z}$. On the other hand, the fundamental group of the complement of two disjoint planes in $\mathbb{R}^4$ is $\mathbb{Z} * \mathbb{Z}$. So in principle, the intersection point accounts for adding a relation (precisely, the commutator of the meridians for the two planes) to a presentation of the fundamental group of the complement. This is just the simplest example of a concept we will work with frequently, namely that adding intersection points to our surfaces can improve the fundamental group of the complement; curiously, increasing intersection points aids us in finding embeddings.

How can we use Casson's ideas to help with the surgery and s-cobordism problems? In both situations, our goal is to remove algebraically cancelling pairs of intersection points (possibly self-intersection points) for spheres S and T immersed in a 4-manifold M, by finding pairwise disjoint, embedded, framed Whitney discs with interiors in the complement of S and T.

First we introduce some terminology. Let A be a subset of a 4-manifold M. We say that A is π_1*-negligible* in M if the inclusion induced map $\pi_1(M \setminus A) \to \pi_1(M)$ is an isomorphism. Note that this implies that any curve in $M \setminus A$ that extends to a map of a disc in M also extends to a map of a disc in the complement $M \setminus A$. This condition will enable us to find Whitney discs whose interiors are in the complement of the spheres we wish to embed.

Now suppose that A is the union of a collection of immersed spheres. By the Seifert–van Kampen theorem, A is π_1-negligible if and only if the meridional circle of each sphere is null-homotopic in the complement of A. Geometrically, such a null homotopy creates an immersed disc bounded by the meridional circle in the complement which, together with a meridional disc, gives a sphere intersecting the original immersed sphere in precisely one point. Thus π_1-negligibility for a family of immersed spheres $\{A_i\}$ in M is equivalent to the existence of geometrically dual spheres. That is, there is a family of immersed spheres $\{B_i\}$ such that A_i intersects B_i transversely at a single point for each i, the spheres $\{B_i\}$ may intersect one another nontrivially, and A_i and B_j are disjoint for $i \neq j$.

1.3.1 *Surgery and Geometric Duals*

In the surgery situation, our initial goal is to represent a hyperbolic pair in the intersection form of an ambient 4-manifold M by geometrically dual spheres $\widehat{S}$ and $\widehat{T}$, given representative spheres S and T algebraically dual to one another and with vanishing self-intersection numbers (see Figure 1.1). We will modify S and T by homotopies until they become geometrically dual spheres $\widehat{S}$ and $\widehat{T}$. After that, we will seek to modify $\widehat{S}$ to an embedding.

Suppose there exists an immersed Whitney disc W in M for a pair of algebraically cancelling intersection points between S and T. For example, this holds when the fundamental group $\pi_1(M)$ of the ambient manifold is trivial and the intersection points have opposite sign, or if one counts intersection points algebraically in $\mathbb{Z}[\pi_1(M)]$ instead of in $\mathbb{Z}$. In all likelihood, W meets both S and T. The first step is to push S and T off the interior of W by so-called *finger moves*, as indicated in Figure 1.8, resulting in spheres S' and T'. The spheres S' and T' are homotopic to S and T respectively, and intersect each other geometrically in the same way as S and T, but have been made disjoint from the interior of W at the expense of (possibly) increasing the number of self-intersections. Note that the algebraic self-intersection numbers are still zero, because finger moves do not change them. Next, perform a Whitney move across W on either S' or T' to obtain new spheres that are still homotopic to S and T and have two fewer intersection points but possibly more (algebraically cancelling) self-intersections. Repeating this process finitely many times yields a geometrically dual pair $\widehat{S}$ and $\widehat{T}$. We have obtained our desired geometrically

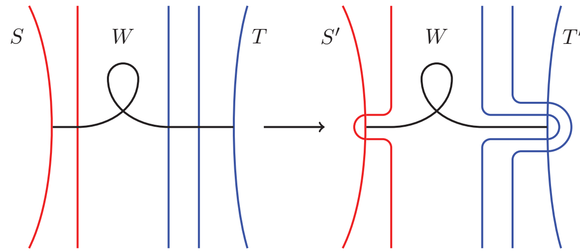

Figure 1.8 Trading intersections for self-intersections. A Whitney disc W (black) pairing algebraically cancelling intersection points between the spheres S (red) and T (blue) is shown in cross section. For each intersection of S with the interior of W, perform a homotopy of S to move the intersection point off W, at the expense of creating two new (algebraically cancelling) self-intersections of S. Do the same for T. This results in the immersed spheres S' (red) and T' (blue) shown on the right.

dual spheres, at the expense of increasing self-intersections. We also know that there are algebraically zero self-intersections for each of $\widehat{S}$ and $\widehat{T}$.

In order to perform a surgery that achieves the desired effect on second homology, we only need to embed one of the spheres, say $\widehat{S}$. We saw above that $\widehat{S}$ has vanishing algebraic self-intersection number. Note that $\widehat{S}$ is π_1-negligible due to the existence of the geometric dual $\widehat{T}$. Thus, using again that the ambient manifold is simply connected or by having counted self-intersections in $\mathbb{Z}[\pi_1(M)]$, there exist Whitney discs pairing the self-intersection points of $\widehat{S}$ whose interiors lie in $M \setminus \widehat{S}$. These Whitney discs are only known to be immersed. If we could instead arrange for pairwise disjoint, embedded, and framed Whitney discs with interiors disjoint from $\widehat{S}$, we would be able to replace $\widehat{S}$ with an embedded sphere. If, in addition, this latter embedded sphere had a geometrically dual sphere, we could do surgery as desired. Note that the geometrically dual sphere ensures that the fundamental group of the ambient manifold remains unchanged after surgery. Such a geometrically dual sphere could come from $\widehat{T}$, depending on how it interacts with the Whitney discs. So if we can find pairwise disjoint, embedded, framed Whitney discs pairing the self-intersection points of $\widehat{S}$ and they can be arranged to have interiors disjoint from both $\widehat{S}$ and $\widehat{T}$, then we will be done. Fortunately, this is exactly what the disc embedding theorem does for us.

While for the purposes of this discussion we have restricted ourselves to a single hyperbolic pair, in reality we will need to embed a collection of spheres disjointly, with a collection of geometrically dual spheres. It is straightforward to extend the argument to the case of several spheres. We explain 4-dimensional surgery in detail in Chapter 22.

1.3.2 The s-cobordism Theorem and Geometric Duals

In the s-cobordism problem the setup is slightly different. Here S and T are not only algebraically dual, but each is embedded and framed and comes with an embedded, possibly unframed, geometric dual: $S^\#$ and $T^\#$, respectively. Again we just consider a single pair $\{S, T\}$ for the purposes of this discussion. The geometrically dual spheres already tell us that S and T are π_1-negligible individually, but they might not be so simultaneously, since, for example, $S^\#$ might intersect T. Let us see how to arrange for $S^\#$ to be disjoint from T with the restriction that we are allowed to move S and T but only by isotopies; this will keep them embedded and also ensure that we are not altering the cobordism we started with. First we arrange that the intersections between $S^\#$ and T cancel algebraically by tubing $S^\#$ into parallel copies of S. That is, we repeatedly perform an ambient connected sum of $S^\#$ and an appropriately oriented copy of S inside M along a suitable arc, as in Figure 1.9. This might increase the number of intersections between $S^\#$ and $T^\#$, or of $S^\#$ with itself, but we do not mind. Now all the intersections between $S^\#$ and T can be paired by Whitney discs in M. Consider some such framed, immersed Whitney disc, W. If we performed the Whitney move on $S^\#$ along W right now, we would be in danger of creating new intersections of $S^\#$ with whatever W intersects, which *a priori* might be any of S, T, $S^\#$, or $T^\#$. However, we do not mind intersections between $S^\#$ and $T^\#$ nor self-intersections of $S^\#$. So the only problems are caused by intersections of W with S or T.

We can remedy the T intersections by tubing W along T into push-offs of $T^\#$, where the push-offs use sections of the normal bundle transverse to the 0-section. This may lead to new intersections of W with S, and also with $T^\#$ if $T^\#$ is not framed. Consequently, the new W only has problematic intersections with S, which, in turn, can be removed by isotoping S off W by finger moves in the direction of T, as shown in Figure 1.10. Since $T^\#$ might not be framed and we have tubed W into it, the framing of the normal bundle of W may no longer agree with the Whitney framing. However, we can correct the framing at the expense of increasing the intersections of W with $S^\#$, by twisting W around its boundary

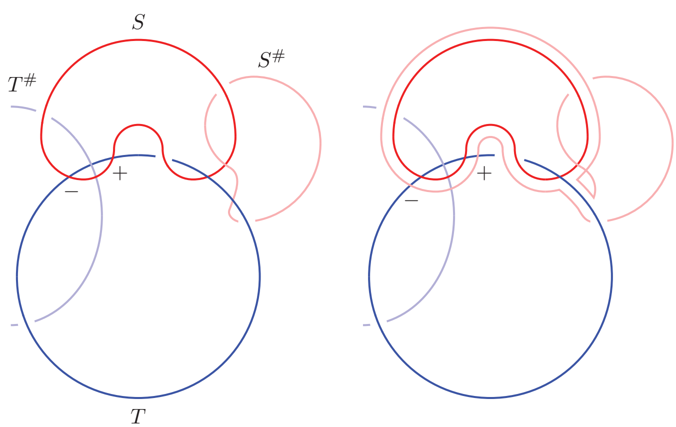

Figure 1.9 Adjusting the intersection number of $S^\#$ (light red) and T (blue) by tubing $S^\#$ into S (red) along a suitable arc. Do this for every point of intersection between $S^\#$ and T. Afterwards, the intersections of $S^\#$ and T are algebraically cancelling. These new intersections are marked on the right with signs.

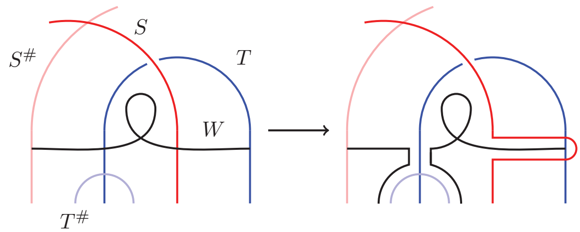

Figure 1.10 Obtaining a Whitney disc for intersections between $S^\#$ and T with interior in the complement of $S \cup T$. We see a Whitney disc W (black) pairing intersection points between $S^\#$ (light red) and T (blue). Remove intersection points of the interior of W with T by tubing W into the dual sphere $T^\#$ (light blue). This might create new intersections (not pictured) of W with S. Remove intersections of W and S by pushing S off W in the direction of T, at the expense of creating a new pair of (algebraically cancelling) intersection points between S and T.

(see Section 15.2.2 for details). At this point, we have possibly made the new Whitney disc more singular (if $T^\#$ meets W, then tubing W into $T^\#$ creates new self-intersections of W) and created new (algebraically cancelling) intersections between S and T, but this does not worry us for now. A Whitney move on $S^\#$ along the new (framed) W produces a (probably immersed) geometric dual for S away from T, as needed.

By applying a similar process, we can upgrade $T^\#$ to a geometric dual for T which does not intersect S, and thus arrange that $S \cup T$ is π_1-negligible. The algebraically cancelling intersection points between S and T may now be paired up with Whitney discs whose interiors lie in the complement of $S \cup T$. However, as in Section 1.3.1, these Whitney discs are only known to be immersed. Note that we found these immersed Whitney discs either by assuming that the ambient manifold is simply connected, or by counting intersection numbers in the group ring $\mathbb{Z}[\pi_1(M)]$.

If we had pairwise disjoint, embedded, and framed Whitney discs in the complement of $S \cup T$ instead, we could use them to perform Whitney moves and obtain a pair of spheres isotopic to S and T and geometrically dual to one another. This would complete the proof of the s-cobordism theorem. Once again, fortunately this is what the disc embedding theorem will provide. As before, we have focused on a single pair of spheres $\{S, T\}$ and their duals $\{S^\#, T^\#\}$, but similar arguments apply to the case of multiple pairs. Further details on the s-cobordism theorem can be found in Chapter 20.

In conclusion, in both the surgery and s-cobordism problems in dimension four, we can use geometrically dual spheres to find immersed Whitney discs with interiors in the complement of the surfaces we are trying to separate. In both cases, we are interested in finding pairwise disjoint, embedded, and framed Whitney discs instead.

1.4 Casson Handles

How can we promote the immersed Whitney discs obtained above to disjointly embedded discs? Assume for this section that the ambient 4-manifold M is simply connected. Note that the singularities of the Whitney discs are isolated double points in the interior. At each such double point, we can find a *double point loop*, that is, a loop that starts at the double point, leaves along one branch, and returns along the other. We may use the same ideas as before to make these loops null-homotopic in the exterior of the base spheres and the Whitney discs and get immersed discs bounded by them away from everything else. If these were pairwise disjoint, embedded, and had the right framing, we could introduce an algebraically cancelling double point for each disc, via adding a local kink, obtain a Whitney disc, and do the Whitney move across this second-level Whitney disc to replace our immersed first-level Whitney discs with embedded ones. This procedure is depicted in Figure 1.11.

The next, and essential, insight of Casson is that we can keep iterating this process by finding layers upon layers of mutually disjoint immersed discs, with each layer attached to the double point loops of the previous layer along the boundary and with interiors disjoint from all previous layers. A closed tubular neighbourhood of the resulting object after any finite number of steps is called a *Casson tower*. See Figure 1.12 for a schematic picture. The base immersed disc in a Casson tower has a circle boundary identified with a Whitney circle of the original immersed spheres. An open tubular neighbourhood of the circle in the boundary of the Casson handle is called the *attaching region*. Take the union of an infinite sequence of inclusions of finite towers, and then remove all the boundary other than the initial attaching region. This is called a *Casson handle*. The attaching region of a Casson handle is the attaching region of any constituent Casson tower, all of which coincide by

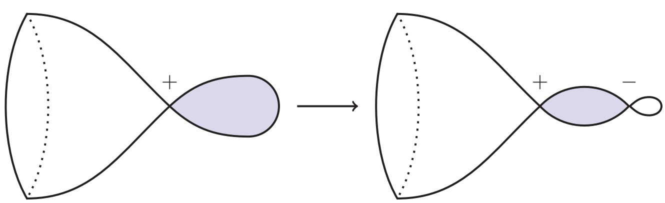

Figure 1.11 Whitney move to resolve a self-intersection. On the left we show a self-intersection point of a disc such that the double point loop bounds an embedded and framed disc whose interior is in the complement of the first disc. When we add a local kink of the opposite sign, a Whitney circle bounding an embedded and framed disc is visible on the right; using this, we may perform the Whitney move to resolve the original self-intersection.

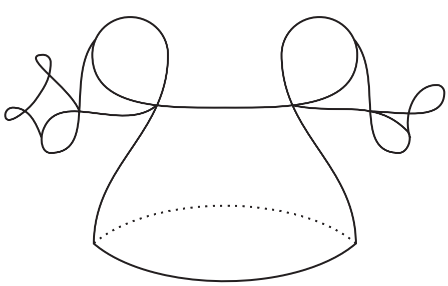

Figure 1.12 Schematic picture of the 2-dimensional spine of a Casson tower of height three.

definition. Thus, in the case of a simply connected ambient manifold, we have now replaced our immersed Whitney discs with disjointly embedded Casson handles.

Note that the fundamental group of a Casson tower is generated by the double point loops at the self-intersections of the final layer of immersed discs, since each successive layer of discs is glued onto a generating set for the fundamental group of the previous stages. Consequently, a Casson handle, informally a Casson tower of infinite height, is simply connected. Casson and Siebenmann proved the following theorem.

Theorem 1.1 (Casson [Cas86, Lecture 1] (see also Siebenmann [Sie80])) *Every Casson handle is properly homotopy equivalent, relative to its attaching region, to the open* 2*-handle* $(D^2 \times \mathring{D}^2, S^1 \times \mathring{D}^2)$.

This is extremely close to what we want. However, to complete our arguments for the surgery and s-cobordism problems, we require not just a proper homotopy equivalence, but rather a homeomorphism to $D^2 \times \mathring{D}^2$, relative to the attaching region. In 1982, Freedman showed exactly this latter fact.

Theorem 1.2 (Freedman [Fre82a]) *Every Casson handle is homeomorphic, relative to its attaching region, to the open* 2*-handle* $(D^2 \times \mathring{D}^2, S^1 \times \mathring{D}^2)$.

One may then consider each Casson handle as one of the topologically embedded, flat, framed Whitney discs that we have been so keen on finding, and perform the Whitney move to delete the offending intersection points.

Based on Casson's constructions outlined in this chapter, in the same paper Freedman used Theorem 1.2 to establish that one can perform surgery on a well chosen smooth 4-manifold to produce a closed topological 4-manifold homotopy equivalent to any given simply connected 4-dimensional Poincaré complex. He also established that every smooth, simply connected h-cobordism between closed 4-manifolds is homeomorphic to a product. Moreover, he proved a more general *proper h-cobordism theorem* which he then used to establish the Poincaré conjecture in dimension four as well as a classification of closed, simply connected, topological 4-manifolds up to homeomorphism, assuming the fact, proved later by Quinn [Qui82b], that every compact, connected topological 4-manifold admits a smooth structure in the complement of a point. There are many more consequences of Freedman's work, which we describe in greater detail in Section 1.6.

1.5 The Disc Embedding Theorem

Casson's construction of Casson handles described above strongly depends on the fact that the ambient manifold is simply connected. For manifolds with more general fundamental groups, there exists a different construction, using layers of surfaces as well as immersed discs, which for certain fundamental groups called *good groups* (discussed below the statement of the theorem) produces what we call a *skyscraper*. Using similar techniques to Freedman in [Fre82a], it can be shown, as first described in [FQ90], that every skyscraper is homeomorphic to the standard 2-handle, relative to the attaching region. This produces the celebrated disc embedding theorem, which we state next and which is the focus of this book.

Below, $\lambda\colon H_2(\widetilde{M},\partial\widetilde{M})\times H_2(\widetilde{M})\to\mathbb{Z}[\pi_1(M)]$ denotes the intersection form, where $\widetilde{M}$ denotes the universal cover of a connected 4-manifold M. We postpone the detailed definition of this, and the precise definition of the self-intersection number μ of an immersed sphere in M, until Chapter 11. Recall that an embedded surface Σ in a 4-manifold M is said to be *flat* if it extends to an embedding $\Sigma\times\mathbb{R}^2\hookrightarrow M$ that restricts to Σ on $\Sigma\times\{0\}$.

Disc embedding theorem ([Fre82a; Fre84; FQ90, Theorem 5.1A; PRT20]) *Let M be a smooth, connected 4-manifold with nonempty boundary and such that $\pi_1(M)$ is a good group. Let*

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

be an immersed collection of discs in M with pairwise disjoint, embedded boundaries. Suppose that F has an immersed collection of framed, algebraically dual 2-spheres

$$G=(g_1,\dots,g_n)\colon S^2\sqcup\cdots\sqcup S^2\looparrowright M;$$

that is, $\lambda(f_i,g_j)=\delta_{ij}$ with $\lambda(g_i,g_j)=0=\mu(g_i)$ for all $i,j=1,\dots,n$.

Then there exists a collection of pairwise disjoint, flat, topologically embedded discs

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

with geometrically dual, framed, immersed spheres

$$\overline{G}=(\overline{g}_1,\dots,\overline{g}_n)\colon S^2\sqcup\cdots\sqcup S^2\looparrowright M,$$

such that, for every i, the discs $\overline{f}_i$ and f_i have the same framed boundary and $\overline{g}_i$ is homotopic to g_i.

Roughly speaking, given a collection of immersed discs with algebraically dual spheres, with some restrictions on the allowed intersections, and in an ambient 4-manifold with nice enough fundamental group, the disc embedding theorem upgrades the immersed discs to embedded discs with tubular neighbourhoods and the same boundary, at the expense of leaving the smooth category.

The hypothesized algebraically dual spheres $\{g_i\}$ in the theorem are needed for π_1-negligibility, and indeed that such dual spheres are required is precisely the reason why the existence of non-slice knots does not contradict the disc embedding theorem.

Good groups will be defined precisely in Chapter 12 and investigated further in Chapter 19, once we have introduced the necessary terms. Briefly, the group needs to satisfy the π_1*-null disc property,* stated in Definition 12.12. For now, it suffices to know that the class of good groups contains groups of subexponential growth and is closed under taking subgroups, quotients, extensions, and colimits. For example, finite groups, abelian groups, and indeed all solvable groups are good. Due to the striking consequences of the disc embedding theorem, the question of which groups are good is one of the most important open questions in 4-manifold topology. Of course, the disc embedding theorem for simply connected 4-manifolds, which holds since the trivial group is good, was itself a ground-breaking result.

Note that the ambient manifold is required to be smooth in the statement of the disc embedding theorem. There exists a category preserving version of the theorem, where 'immersed' discs in a topological manifold are promoted to embedded ones. However, the proof requires the notion of topological transversality and smoothing away from a point (see Section 1.6). These facts, established by Quinn [Qui88; FQ90, Chapters 8 and 9], in turn depend on the disc embedding theorem in a smooth 4-manifold stated above. The fully topological version of the disc embedding theorem is beyond the scope of this book, since we will not discuss Quinn's proof of transversality. We summarize the developments in topological 4-manifold theory that stemmed from the disc embedding theorem in Chapter 21.

The proof of the disc embedding theorem will occupy us for almost the entire book. As outlined in this chapter, the original proof in the simply connected case consisted of building disjoint Casson handles with the same attaching region as the original immersed discs and then showing that every Casson handle is homeomorphic to the standard open handle $D^2 \times \mathring{D}^2$ relative to its attaching region (Theorem 1.2). Freedman's proof of the latter fact consisted of embedding uncountably many compactified Casson handles within the original Casson handle and then applying techniques of *decomposition space theory* and *Kirby calculus*.

The proof in the remainder of this book will not use Casson handles, but rather the alternate infinite tower construction alluded to above, called *skyscrapers*, consisting of layers of both surfaces and immersed discs. Using skyscrapers simplifies both the embedding and decomposition space theory steps of the proof. The proof we shall present is an elaboration of the proof given in the book by Freedman and Quinn [FQ90], using a modification of the constructive step given in [PRT20]. In particular, each skyscraper is compact, and we will show that it is homeomorphic to $D^2 \times D^2$ relative to its attaching region rather than the open 2-handle $D^2 \times \mathring{D}^2$. We direct the reader to the outline of our proof in Chapter 2, where we point out more precisely what is simplified and gained by the skyscraper approach.

Remark 1.3 The geometrically dual spheres $\{\overline{g}_i\}$ in the outcome of the disc embedding theorem were asserted to exist in [FQ90, Theorem 5.1A], but no proof was given. They are also not directly addressed in [Fre82a, Fre84]. They are explicitly constructed in [PRT20] by modifying the constructive part of the proof from [FQ90]. We also include the observation from [PRT20] that each $\overline{g}_i$ is homotopic to g_i. As noted earlier, the geometrically dual spheres are essential when performing surgery to ensure that the fundamental group of the ambient manifold is not altered. We describe the surgery procedure in Chapter 22. In Chapter 20 we also show how to apply the version of the disc embedding theorem without geometrically dual spheres in the outcome to prove the s-cobordism theorem.

1.6 After the Disc Embedding Theorem

The consequences of the disc embedding theorem are many and far reaching, including several foundational results in topological 4-manifold theory. In this section, we list some of the most prominent of the disc embedding theorem's many applications.

1.6.1 Foundational Results

We begin with normal bundles and transversality for submanifolds of topological manifolds. Recall that an embedded surface Σ in a 4-manifold M is said to be *locally flat* if every point in Σ admits a neighbourhood U in M such that $(U, U \cap \Sigma)$ is homeomorphic to $(\mathbb{R}^4, \mathbb{R}^2)$. A normal bundle for a locally flat submanifold $N \subseteq M$ of a topological 4-manifold M is a vector bundle $E \to N$ with an embedding of the total space $E \to M$ such that the 0-section agrees with the inclusion of N and such that E is *extendable*, where the latter term means that if E embeds as the open unit disc bundle of another vector bundle $F \to N$, then the embedding $E \to M$ extends to an embedding $F \to M$ (see [FQ90, p. 137]).

Theorem 1.4 **([FQ90, Section 9.3])** *Every locally flat proper submanifold of a topological 4-manifold has a normal bundle, unique up to ambient isotopy.*

Theorem 1.5 **([Qui88; 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' that is transverse to Σ_2.*

Here, *transverse* means that the points of intersection have coordinate neighbourhoods within which the submanifolds appear as transverse linear subspaces.

It is worth pointing out that in the context of smooth manifolds, transversality and the existence of normal bundles for submanifolds are among the basic results of differential topology. For topological 4-manifolds, the disc embedding theorem is a crucial component of the proofs, and without these results any work with topological submanifolds would be well-nigh impossible.

Freedman's techniques were extended by Quinn to prove the 4-dimensional annulus theorem, stated below.

Theorem 1.6 **(4-dimensional annulus theorem [Qui82b])** *Let $f\colon S^3 \to \operatorname{Int} D^4$ be a locally flat embedding. Then the region between $f(S^3)$ and $S^3 = \partial D^4$ is homeomorphic to the annulus $S^3 \times [0,1]$.*

The result implies that connected sum of oriented topological 4-manifolds is well defined, which had not been known previously. To see this, one notes that connected sum of two 4-manifolds M_1 and M_2 depends *a priori* on a choice of embeddings $D^4 \hookrightarrow M_i$ for $i = 1, 2$. Suppose we are given two embeddings of D^4 in M_i. First produce an ambient isotopy of M_i, taking one ball to a proper sub-ball of the other. Then apply the annulus theorem to produce an isotopy, taking the sub-ball to the bigger ball. Since isotopic embeddings of balls produce homeomorphic connected sums and since every orientation preserving homeomorphism of S^3 is isotopic to the identity [Fis60], it follows that the connected sum is well defined. See Section 21.4.4 for more discussion.

In the same paper, Quinn showed that topological 5-manifolds (not necessarily compact) have topological handlebody structures. Combined with the work of Kirby and Siebenmann [KS77, Essay III, Section 2], as well as Bing [Bin59, Theorem 8] and Moise [Moi52a], this shows that a manifold (of any dimension) admits a topological handlebody structure if and only if it is not a non-smoothable 4-manifold.

Quinn also proved the following result about the smoothability of topological 4-manifolds.

Theorem 1.7 ([Qui82b, Corollary 2.2.3; LT84; FQ90, Theorem 8.7; Qui86]) *The natural map $TOP(4)/O(4) \to TOP/O$ is 5-connected. Moreover, every noncompact, connected component of a topological 4-manifold admits a smooth structure.*

In particular, every compact, connected, topological 4-manifold admits a smooth structure in the complement of a point.

Chapter 21 contains a deeper discussion of these foundational results and their implications.

1.6.2 Classification Results

Roughly speaking, the disc embedding theorem implies that in the topological category, 4-manifolds behave much like high-dimensional manifolds. Topological transversality and the existence of topological handlebody structures on 5-manifolds yield the *topological h-cobordism theorem* for 4-manifolds [FQ90, Chapter 7], following the proof outlined earlier in this chapter. This has far-reaching consequences of its own. For example, it implies the topological 4-dimensional Poincaré conjecture.

Theorem 1.8 (Poincaré conjecture [Fre82a]) *Every homotopy 4-sphere is homeomorphic to the 4-sphere.*

We also obtain the topological s-cobordism theorem for 4-manifolds with good fundamental group. Via the strategy pointed out earlier, the disc embedding theorem implies that the surgery strategy applies topologically for good groups. More precisely, this means the following. Let X be a 4-dimensional Poincaré complex with a lift of its Spivak normal fibration to a TOP-bundle. Suppose that $\pi_1(X)$ is a good group. Then there is a topological 4-manifold M homotopy equivalent to X if and only if, up to choosing a different lift, the corresponding surgery obstruction in $L_4(\mathbb{Z}[\pi_1(X)])$ vanishes. Moreover, for such an X, when the structure set is nonempty, the surgery sequence

$$L_5(\mathbb{Z}[\pi_1(X)]) \to \mathcal{S}^{TOP}(X) \to \mathcal{N}^{TOP}(X) \to L_4(\mathbb{Z}[\pi_1(X)])$$

is defined and exact as a sequence of pointed sets. This uses the *sphere embedding theorem* proven in Chapter 20. As we explain in Chapter 22, computing the structure set $\mathcal{S}^{TOP}(X)$ can lead to classifications of manifolds within a fixed homotopy type. By contrast, surgery does not work for smooth 4-manifolds, neither of the h- and s-cobordism theorems hold [MS78, CS85, Don87a], and there is no known definition of a nontrivial action of $L_5(\mathbb{Z}[\pi_1(X)])$ on the smooth structure set $\mathcal{S}^{DIFF}(X)$. The smooth 4-dimensional Poincaré conjecture remains open to date.

We will prove the sphere embedding theorem and discuss the use of the disc embedding theorem in surgery, the s-cobordism theorem, and the Poincaré conjecture in Chapters 20, 21, and 22.

In the strategy mentioned above, in order to see that the topological structure set $\mathcal{S}^{TOP}(X)$ for a given Poincaré complex X is nonempty, we need to build a closed 4-manifold such that the surgery obstruction vanishes. We are able to do so in the simply connected case using the following theorem of Freedman [Fre82a].

Theorem 1.9 ([Fre82a, Theorem 1.4′; FQ90, Corollary 9.3C]) *Every integral homology 3-sphere is the boundary of a contractible, compact, topological 4-manifold, which is unique up to homeomorphism.*

The high-dimensional counterpart of this statement follows from surgery theory [Ker69]. The existence of such contractible 4-manifolds allows us to construct a closed, simply connected, topological 4-manifold with any given nonsingular intersection form, as follows. Let $\lambda\colon \mathbb{Z}^n \times \mathbb{Z}^n \to \mathbb{Z}$ be a unimodular, symmetric, bilinear form. Take the disjoint union $\sqcup_{i=1}^n B_{k_i}$ of disc bundles over the 2-sphere of the form

$$D^2 \to B_{k_i} \to S^2$$

of Euler number k_i, where $k_i = \lambda(e_i, e_i)$ is the ith diagonal entry of the matrix representing λ with respect to the standard basis of $\mathbb{Z}^n$. Plumb these together according to λ to construct a smooth, simply connected, compact 4-manifold with λ as its intersection form and nonempty boundary. Since λ is unimodular, the boundary of this compact manifold is a homology sphere, which we cap off by the (topological) contractible 4-manifold produced by Theorem 1.9. The result is the desired closed, topological 4-manifold.

In arguably the most interesting case, that of λ represented by the E_8 matrix, the topological manifold produced is called the *E_8-manifold*. As noted earlier, by Rochlin's theorem, there is no closed, simply connected, smooth 4-manifold realizing E_8 as its intersection form, and thus the smooth version of Theorem 1.9 does not hold.

The existence of the E_8-manifold is extremely helpful in surgery. Given a lift for the Spivak normal fibration for a simply connected 4-dimensional Poincaré complex X, we obtain a 4-manifold M with a degree one normal map to X, as mentioned earlier. Take the connected sum of M with copies of the E_8-manifold to arrange for the algebraic obstruction to surgery to vanish. Then we can apply the disc embedding theorem to do surgery and establish that the topological structure set $\mathcal{S}^{TOP}(X)$ is nonempty. See Chapter 22 for further details.

Here is another application of Theorem 1.9. Starting with an integral homology 3-sphere Σ and doubling the contractible 4-manifold from Theorem 1.9 with boundary Σ, we obtain a homotopy 4-sphere, which is homeomorphic to S^4 by the topological 4-dimensional Poincaré conjecture. Thus we have the following theorem.

Corollary 1.10 *Every integral homology 3-sphere admits a locally flat topological embedding into S^4.*

Using Theorem 1.9 and a combination of surgery for the trivial group (which we recall is a good group) and the topological h-cobordism theorem, we may upgrade the Milnor–Whitehead homotopy classification of topological 4-manifolds to the following homeomorphism classification [Fre82a] (see also [FQ90] and [CH90]). We say that a 4-manifold M is *stably smoothable* if $M\#k(S^2 \times S^2)$ admits a smooth structure for some k.

Theorem 1.11 (Homeomorphism classification of closed, topological, simply connected 4-manifolds [Fre82a, Theorem 1.5]) *Fix a symmetric, nonsingular, bilinear form $\theta\colon F \times F \to \mathbb{Z}$ on a finitely generated free abelian group F.*

(1) *If θ is even, there exists a closed, topological, simply connected, (spin), oriented 4-manifold, unique up to homeomorphism, whose intersection form is isometric to (F, θ). This 4-manifold is stably smoothable if and only if the signature of θ is divisible by 16.*

(2) *If θ is odd, there are two homeomorphism classes of closed, topological, simply connected, (non-spin), oriented 4-manifolds with intersection form isometric to (F, θ), one of which is stably smoothable and one of which is not.*

Let M and M' be two closed, topological, simply connected, oriented 4-manifolds and suppose that $\phi\colon H_2(M;\mathbb{Z}) \to H_2(M';\mathbb{Z})$ is an isomorphism that induces an isometry between the intersection forms. If the intersection forms are odd, assume in addition that M and M' are either both stably smoothable or both not stably smoothable. Then there is a homeomorphism $G\colon M \to M'$ such that $G_* = \phi\colon H_2(M;\mathbb{Z}) \to H_2(M';\mathbb{Z})$.

In other words, every even, symmetric, integral, matrix with determinant ± 1 is realized as the intersection form of a unique closed, simply connected, oriented, topological 4-manifold. For such matrices which are odd instead, we get two closed, simply connected, oriented, topological 4-manifolds, exactly one of which is stably smoothable.

On the other hand, by Donaldson's Theorem A [Don83, Don87b], the only definite intersection forms realized as the intersection form of a closed, smooth 4-manifold (not necessarily simply connected) are the standard forms that are the intersection forms of connected sums of $\mathbb{CP}^2$ or connected sums of $\overline{\mathbb{CP}^2}$. Thus there is no closed, smooth 4-manifold with intersection form $E_8 \oplus E_8$. But, by Theorem 1.11, the form $E_8 \oplus E_8$ is realized as the intersection form of a closed, simply connected, topological 4-manifold.

It is still an open question exactly which indefinite forms are realized by closed, simply connected, smooth 4-manifolds. However, partial results exist. For example, further work of Donaldson shows that there is no closed, simply connected, smooth 4-manifold with $E_8 \oplus E_8 \oplus H$ or $E_8 \oplus E_8 \oplus H \oplus H$ as its intersection form [Don86]. The $10/8$ theorem of Furuta [Fur01], recently extended to a $10/8 + 4$ theorem by [HLSX18], obstructs the realization of even more bilinear forms as the intersection forms of closed, simply connected, smooth 4-manifolds. The latter theorem, as well as some work of Donaldson, such as [Don87b], applies to certain non-simply connected 4-manifolds.

Uniqueness also fails quite drastically in the smooth category. There are many constructions of pairs of smooth manifolds that are homeomorphic but not diffeomorphic; these are known as *exotic pairs*. For example, there are infinitely many smooth 4-manifolds homeomorphic to the $K3$ surface, but not diffeomorphic to it [FS98]; similar constructions exist for certain blow ups of the complex projective plane [Don87a, FM88, Kot89, Par05, SS05, PSS05, AP08, BK08, AP10]. For noncompact manifolds, the situation is even wilder. There are uncountably many smooth manifolds that are homeomorphic, but not diffeomorphic, to $\mathbb{R}^4$ with its standard smooth structure [Tau87]. Such a manifold is called an *exotic* $\mathbb{R}^4$. Indeed, there is not a single smooth 4-manifold for which we know that only finitely many distinct smooth structures exist.

The classification of closed, simply connected, topological 4-manifolds was stated in terms of being *stably smoothable*. A compact 4-manifold M is stably smoothable if and only if the *Kirby–Siebenmann invariant* $\mathrm{ks}(M) \in \mathbb{Z}/2$ vanishes [FQ90, Sections 8.6 and 10.2B]. More accurately, this is the obstruction for the stable tangent microbundle of M to admit a lift to a PL-bundle. The existence of such a lift implies that M is stably smoothable, which is also equivalent to $M \times \mathbb{R}$ admitting a smooth structure by smoothing theory [KS77, Essay V]. For closed, simply connected, topological 4-manifolds with even intersection form, and more generally, closed, topological, spin 4-manifolds, the Kirby–Siebenmann invariant is congruent mod 2 to $\sigma(M)/8$. Thus, a closed, simply connected, topological 4-manifold with intersection form $E_8 \oplus E_8$ has vanishing Kirby–Siebenmann invariant. However, we saw earlier that Donaldson's Theorem A [Don83] implies that this manifold is not smoothable. As a result, a compact, topological 4-manifold with vanishing Kirby–Siebenmann invariant—that is, a compact, stably smoothable, topological 4-manifold—need not be smoothable. See [FQ90, Section 10.2B; KS77] for more details on the Kirby–Siebenmann invariant.

We saw that there are two homotopy equivalent but non-homeomorphic closed, simply connected, topological 4-manifolds with a given odd intersection form, one manifold for each value of the Kirby–Siebenmann invariant. As we saw, the manifold with vanishing Kirby–Siebenmann invariant is stably smoothable. The simplest example of an odd unimodular intersection form, namely $\langle 1\rangle$, is already interesting. The two manifolds with this intersection form are $\mathbb{CP}^2$ and its *star partner*, $*\mathbb{CP}^2$, sometimes called the *Chern manifold*.

To construct $*\mathbb{CP}^2$, attach a $+1$-framed 2-handle to a knot K in $S^3 = \partial D^4$ with $\mathrm{Arf}(K) = 1$. The resulting 4-manifold $X_1(K)$ has intersection form $\langle 1\rangle$ and boundary an integral homology sphere Σ, namely the result of $+1$-framed surgery on S^3 along K. Cap off this homology sphere with the contractible 4-manifold C with boundary Σ promised by Theorem 1.9, and call the resulting manifold N. To see that N is homeomorphic to $*\mathbb{CP}^2$, it suffices to show that $\mathrm{ks}(N) = 1$. For this, observe that $\mathrm{ks}(N) = \mathrm{ks}(C)$ since the Kirby–Siebenmann invariant is additive for 4-manifolds glued along their boundary and $X_1(K)$ is smooth (see [FNOP19, Section 8] for more on the Kirby–Siebenmann invariant). Since C is contractible, C is a topological spin manifold. By [FQ90, p. 165; GA70], $\mathrm{ks}(C) = \mu(\Sigma) = \mathrm{Arf}(K) = 1$, where $\mu(\Sigma)$ is the Rochlin invariant of Σ.

As an alternative construction, consider the connected sum $E_8\#\overline{\mathbb{CP}^2}$, where we abuse notation to let E_8 denote the E_8-manifold. One can compute that $E_8\#\overline{\mathbb{CP}^2}$ has intersection form

$$E_8 \oplus \langle -1\rangle \cong 8\langle 1\rangle \oplus \langle -1\rangle,$$

since these are both indefinite, symmetric, nonsingular, integral, bilinear forms with the same rank, signature, and parity [Ser70, MH73]. The injection

$$7\langle 1\rangle \oplus \langle -1\rangle \hookrightarrow 8\langle 1\rangle \oplus \langle -1\rangle \cong E_8 \oplus \langle -1\rangle$$

on the level of intersection forms produces a connected sum decomposition

$$E_8\#\overline{\mathbb{CP}^2} = 7\mathbb{CP}^2\#\overline{\mathbb{CP}^2}\#N$$

for some closed 4-manifold N with intersection form $\langle 1 \rangle$; this uses the disc embedding theorem, as shown in [FQ90, Section 10.3]. This N has

$$\mathrm{ks}(N) = \mathrm{ks}(7\mathbb{CP}^2 \# \overline{\mathbb{CP}^2} \# N) = \mathrm{ks}(E_8 \# \overline{\mathbb{CP}^2}) = 1,$$

since $\mathrm{ks}(E_8) = 1, \mathrm{ks}(\mathbb{CP}^2) = \mathrm{ks}(\overline{\mathbb{CP}^2}) = 0$, and the Kirby–Siebenmann invariant is additive under connected sum. Then we define $*\mathbb{CP}^2$ to be N.

In general, a star partner of a non-spin 4-manifold W is a manifold $*W$ such that $*W \# \mathbb{CP}^2$ is homeomorphic to $W \# * \mathbb{CP}^2$, via a homeomorphism preserving the decomposition of π_2. For closed, simply connected, non-spin 4-manifolds, this equation, together with [FQ90, Section 10.3], uniquely determines a 4-manifold $*W$, which gives the non-stably smoothable manifolds with odd intersection form from Theorem 1.11. For more general fundamental groups it is not known precisely when star partners of non-spin manifolds exist, nor when the star manifold is unique, should one exist. See [Sto94, Tei97, RS97] for more on this question.

Compact, simply connected, topological 4-manifolds with fixed connected boundaries have also been classified using the disc embedding theorem [Boy86, Boy93, Vog82, Sto93] in terms of the intersection form and the Kirby–Siebenmann invariant. Classification results for closed, topological 4-manifolds exist for other families of good groups as well. For example, closed, topological 4-manifolds with infinite cyclic fundamental group are classified in terms of the spin type, the equivariant intersection form, and the Kirby–Siebenmann invariant [FQ90, Theorem 10.7A, p. 173]. See Section 22.3.4 for further discussion.

As one last example, since free abelian groups are good, the following rigidity, which is a special case of the Borel conjecture, follows from the surgery exact sequence.

Theorem 1.12 *Let M be a closed, topological 4-manifold homotopy equivalent to the torus T^4. Then M is homeomorphic to T^4.*

1.6.3 Knot Theory Results

We end this chapter by giving a few applications of the disc embedding theorem in the realm of knot theory. First, Freedman characterized the unknotted $S^2 \subseteq S^4$ as follows.

Theorem 1.13 ([FQ90, Theorem 11.7A]) *If $S \subseteq S^4$ is a spherical 2-knot with $\pi_1(S^4 \setminus S) \cong \mathbb{Z}$, then S is topologically isotopic to the unknot.*

This is the analogue of the classical result that a 1-knot in S^3 is unknotted if and only if the fundamental group of the complement is $\mathbb{Z}$. The smooth counterpart of Freedman's result for 2-knots remains open.

For classical knots, the following is a central result. We give a sketch of the proof to give a sense of how the disc embedding theorem and its various consequences are necessary.

Theorem 1.14 ([FQ90, Section 11.7; GT04]) *Let $K \subseteq S^3$ be a knot with Alexander polynomial $\Delta_K(t) = \pm t^k$ for some $k \in \mathbb{Z}$. Then K is topologically slice; that is, K bounds a locally flat, embedded disc in D^4.*

Sketch of proof Let M_K denote the result of 0-framed Dehn surgery on S^3 along K. We will construct a compact 4-manifold W with $\partial W = M_K$ such that W is a homology circle whose fundamental group is normally generated by a meridian of K. Given such a 4-manifold, the union of W with a 2-handle glued along a meridian of K produces a homotopy 4-ball with boundary S^3. By the classification of simply connected 4-manifolds, this is homeomorphic to D^4 and the image of the cocore of the attached 2-handle gives the desired locally flat (indeed flat) slice disc for K.

In order to construct W, observe that the spin bordism group $\Omega_3^{spin}(S^1) \cong \Omega_2^{spin} \cong \mathbb{Z}/2$ is detected by the Arf invariant of K. The Arf invariant can be computed from the Alexander polynomial, and so vanishes. Thus there exists a compact, spin 4-manifold V with boundary M_K and a map to S^1 extending the map to S^1 on M_K, corresponding to a generator of $H^1(M_K;\mathbb{Z})$ and sending a positively oriented meridian to 1.

Perform surgery on circles in V to obtain V' with $\pi_1(V') \cong \mathbb{Z}$. The spin condition on V implies that for every element of $\pi_2(V)$ there is a fixed regular homotopy class of immersions of S^2 having trivial normal bundle: the Euler number of the normal bundle can be changed by ± 2 by adding local kinks. The $\mathbb{Z}$-equivariant intersection form on $\pi_2(V')$ is nonsingular and thus defines a surgery obstruction in $L_4(\mathbb{Z}[\mathbb{Z}])$. Here for nonsingularity we use the fact that $H_1(M_K;\mathbb{Z}[\mathbb{Z}]) = 0$, since $\Delta_K(t)$ is a unit in $\mathbb{Z}[\mathbb{Z}]$. Moreover, we are using surgery for manifolds with boundary. It is crucial here that the relevant fundamental group is $\mathbb{Z}$, which is a good group. We have that $L_4(\mathbb{Z}[\mathbb{Z}]) \cong 8\mathbb{Z}$, with generator the E_8 form. Take the connected sum of V' with copies of the E_8-manifold to produce V'' with vanishing surgery obstruction. This implies, by the exactness of the surgery sequence for manifolds with boundary, that there exists a half-basis of $H_2(V'')$ consisting of framed, embedded spheres with geometric duals (see the sphere embedding theorem in Chapter 20) on which we can perform surgery to obtain a 4-manifold W. By construction, W is homotopy equivalent to S^1, and so satisfies the desired conditions. □

Theorem 1.14 shows that there are many topologically slice knots. On the other hand, smooth obstructions can show that many of these are not smoothly slice. For example, the Whitehead double of the right-handed trefoil knot has Alexander polynomial one but is not smoothly slice. In fact, the group of topologically slice knots modulo smoothly slice knots is known to be quite large. It contains an infinite rank summand and a subgroup isomorphic to $(\mathbb{Z}/2)^\infty$ [End95, OSS17, HKL16]. There exists an infinite sequence of obstructions to smooth sliceness for topologically slice knots [CHH13, CK17], similar to those due to Cochran–Orr–Teichner [COT03].

Any knot K that is topologically slice but not smoothly slice can be used to construct an exotic $\mathbb{R}^4$ [Gom85, Lemma 1.1], as follows. Attach a 0-framed 2-handle to the 4-ball D^4 along $K \subseteq \partial D^4$, to obtain the 4-manifold $X_0(K)$. By construction, $X_0(K)$ is a smooth manifold, once we smooth the corners produced by handle addition. Since K is topologically slice, there is a topological locally flat embedding of $X_0(K)$ into $\mathbb{R}^4$, taking the $D^4 \subseteq X_0(K)$ to the unit 4-ball in $\mathbb{R}^4$. The closure of the complement of this embedding

$$U := \overline{\mathbb{R}^4 \setminus X_0(K)}$$

is connected and noncompact and thus admits a smooth structure by Theorem 1.7. The smooth structures on $X_0(K)$ and U glue together to give a smooth structure on $\mathbb{R}^4$, since every homeomorphism of a 3-manifold, in this case $\partial X_0(K)$, is isotopic to a diffeomorphism [Moi52a, Bin59]. Let $\mathcal{R}$ denote $\mathbb{R}^4$ endowed with this smooth structure. Note that the manifold $X_0(K)$ embeds smoothly into $\mathcal{R}$.

Theorem 1.15 *The smooth 4-manifold $\mathcal{R}$ is not diffeomorphic to $\mathbb{R}^4$.*

Proof Suppose, for the sake of a contradiction, that $\mathcal{R}$ is diffeomorphic to $\mathbb{R}^4$ with the standard smooth structure. Then $X_0(K)$ embeds smoothly in the standard $\mathbb{R}^4$ and thus in the standard, smooth 4-sphere S^4, produced by adding a point to $\mathbb{R}^4$. Since any two smooth, oriented embeddings of the standard ball in a connected, smooth manifold are isotopic [RS72, Theorem 3.34], by the isotopy extension theorem, we can assume that $D^4 \subseteq X_0(K)$ is mapped to the lower hemisphere of S^4. Then the closure of the complement of the image of $D^4 \subseteq X_0(K)$ in S^4 is also the standard D^4. The image of the 2-handle in $X_0(K)$ then provides a smooth slice disc for K in this complementary D^4. Since K is not smoothly slice by hypothesis, we have reached a contradiction. This establishes that $\mathcal{R}$ is an exotic $\mathbb{R}^4$. □

We will discuss the use of the disc embedding theorem in surgery, the s-cobordism theorem, the classification of closed, simply connected, topological 4-manifolds, and the Poincaré conjecture in more detail in Chapters 20, 21, and 22. The proofs of the other consequences of the disc embedding theorem presented in this chapter are beyond the purview of this book, and we encourage the reader to study the references given in this section. Primarily, the rest of the book proves the disc embedding theorem in a smooth, connected ambient 4-manifold.

2

Outline of the Upcoming Proof

ARUNIMA RAY

We present an outline of the forthcoming proof of the disc embedding theorem, to orient the reader before we begin. The nonorientable reader is requested to pass to their orientation double cover before continuing. The remainder of this book breaks up the proof into small digestible pieces. The goal of this chapter is to describe how the pieces fit together. This outline is necessarily thin on specifics, and we take a few liberties with the precise definitions and proofs that we will give later, in the interest of providing a general sense of what is to come. We hope that this will be a helpful guide for the reader for when we delve into the proof in earnest. In the course of reading this book, readers might choose to periodically return to this outline to see where they are within the proof.

2.1 Preparation

We work within an ambient 4-manifold M with *good* fundamental group that is additionally assumed to be *smooth*. Thus we will freely discuss *immersions* and *transversality*.

Freedman's disc embedding theorem (Section 1.5) states that, given a collection of properly immersed discs $\{f_i\}$ in such a 4-manifold, with a corresponding collection of framed, algebraically dual, immersed spheres $\{g_i\}$ (that is, $\lambda(f_i, g_j) = \delta_{ij}$, $\lambda(g_i, g_j) = 0$, and $\mu(g_i) = 0$ for all i, j), we can replace $\{f_i\}$ with a collection $\{\overline{f}_i\}$ of flat, disjointly embedded discs such that f_i and $\overline{f}_i$ have the same framed boundary for all i and such that $\{\overline{f}_i\}$ is equipped with a collection $\{\overline{g}_i\}$ of geometrically dual spheres, with $\overline{g}_i$ homotopic to g_i for each i. *Geometrically dual* means that $\overline{f}_i$ and $\overline{g}_j$ are disjoint whenever $i \neq j$, while $\overline{f}_i$ and $\overline{g}_i$ intersect transversely at a single point for each i. The hypothesis that the ambient manifold has good fundamental group is used in a single step of the proof that we shall indicate below.

For the purposes of this outline, we conflate the original immersions $\{f_i\}$ with their image in M. The strategy to promote the original immersed discs $\{f_i\}$ to disjointly embedded discs has two major steps. First, we build a pairwise disjoint collection of complicated 4-dimensional objects called *skyscrapers*, which attempt to approximate a pairwise disjoint collection of embedded, framed Whitney discs for the intersections and self-intersections

Arunima Ray, *Outline of the Upcoming Proof* In: *The Disc Embedding Theorem*. Edited by: Stefan Behrens, Boldizsár Kalmár, Min Hoon Kim, Mark Powell, and Arunima Ray, Oxford University Press. © Oxford University Press 2021.
DOI: 10.1093/oso/9780198841319.003.0002

of $\{f_i\}$. Second, we show that every skyscraper is in fact homeomorphic, relative to its attaching region, to $D^2 \times D^2$. As mentioned in the previous chapter, Freedman's original proof used a different infinite construction, called a *Casson handle*. We will point out below how our proof, which is an elaboration of the proof in [FQ90] incorporating a modification of the first step given in [PRT20], bypasses some of the technical complications of the Casson handle approach.

The techniques from general topology that we will use to show that any skyscraper is homeomorphic to the standard handle come from the realm of *decomposition space theory*, sometimes known as *Bing topology*. We develop the specific results and techniques we need in Part I, exhibiting a proof of the Schoenflies theorem in all dimensions to introduce the theory. Part I also includes an in depth discussion of the Alexander horned sphere. Unsurprisingly, perhaps, since the disc embedding theorem is inherently a topological result (rather than a smooth one, for instance), the techniques from Part I will be essential to the eventual proof of the disc embedding theorem. In Part II, which may be read independently of Part I, we show how to build skyscrapers. The vast majority of the constructions in Parts I and II are direct and hands-on. Part III is an interlude which discusses good groups in greater detail, shows how to apply the disc embedding theorem as well as the techniques from Part II to topological 4-manifolds, and discusses some open questions and conjectures. Part IV completes the proof of the disc embedding theorem by showing that any skyscraper is homeomorphic to the standard handle, relative to the attaching region. We indicate exactly which ingredients we need from Parts I and II at the beginning of Part IV. In contrast to the previous parts, the techniques in Part IV are markedly more abstract and harder to visualize.

Next we sketch the proof of the disc embedding theorem. We reference precise propositions and theorems in the upcoming proof whenever possible, and we use the same notation as in these results.

2.2 Building Skyscrapers

The skyscrapers we build will be the limit of a progression of iterated constructions, such that each finite truncation is roughly speaking an approximation of an embedded 2-handle. An obvious difference between a neighbourhood of an immersed disc and that of an embedded disc is the double point loops traversing the self-intersections. These are essential in the fundamental group of the image of the immersion. In other words, in seeking to approximate an embedded 2-handle, we should aim to construct something simply connected. This is a guiding principle throughout the construction of skyscrapers. Recall that Casson handles were built as a neighbourhood of an infinite tower of immersed discs, with each disc's boundary glued onto a double point loop of a previous disc in the tower. Skyscrapers will be built similarly, except that there will be some surface stages between any two disc stages. Now we begin explaining the construction of skyscrapers performed in Part II.

Step 1 (Modify the base discs until the intersections may be paired by Whitney discs, Proposition 16.1). We start with the initial hypothesized immersed discs $\{f_i\}$ and algebraically dual immersed spheres $\{g_i\}$, where $\{g_i\}$ has trivial intersection and self-intersection numbers by hypothesis. Tube all of the intersections and self-intersections of $\{f_i\}$ into $\{g_i\}$ to arrange that the intersection and self-intersection numbers of the new immersed discs $\{f_i'\}$ are trivial. Note that $\{f_i\}$ and $\{f_i'\}$ have the same framed boundaries. Then we upgrade the algebraically dual spheres $\{g_i\}$ to geometrically dual spheres $\{g_i'\}$ for $\{f_i'\}$, using the ideas of Casson mentioned in Chapter 1. This changes the $\{f_i'\}$ by a regular homotopy, but we still call them $\{f_i'\}$ and the framed boundaries remain the same. Note that g_i and g_i' are regularly homotopic. Since the intersection and self-intersection numbers of the $\{f_i'\}$ are zero, the intersection points are paired by Whitney circles bounding framed, immersed Whitney discs $\{D_k\}$ in the ambient manifold. Tube any intersections of the interiors of these Whitney discs with the $\{f_i'\}$ into the geometrically dual spheres $\{g_i'\}$. Now we have a collection $\{D_k'\}$ of framed (but merely immersed) Whitney discs for the intersections and self-intersections of the $\{f_i'\}$, with interiors in the complement of the $\{f_i'\}$. Note that $\{D_k'\}$ may intersect the spheres $\{g_i'\}$.

To keep the goal in sight, remember that we will eventually replace these immersed Whitney discs $\{D_k'\}$ with pairwise disjoint skyscrapers, which we will later see to be homeomorphic to 2-handles, allowing us to perform the Whitney move on the $\{f_i'\}$.

It would be quite understandable for the reader to be somewhat confused at this point, because it seems that no real progress has been made. We have simply swapped the immersed discs $\{f_i\}$ in the original ambient manifold M for other immersed discs $\{D_k'\}$ in the new ambient manifold $M \setminus \bigcup \nu f_i'$. However, since $\{D_k'\}$ is a collection of Whitney discs, it is equipped with a collection of *transverse capped surfaces*, which will be a key ingredient in the next few steps. Transverse capped surfaces are strictly better than algebraically dual spheres, since they can be used to produce arbitrarily many mutually disjoint geometrically dual spheres at will, as we will indicate soon.

The transverse capped surfaces will be produced from *Clifford tori*. As mentioned in Chapter 1, Clifford tori are found in neighbourhoods of transverse intersections between two surfaces in an ambient 4-manifold. More precisely, the two circle factors in a Clifford torus are each meridians for one of the two intersecting surfaces. The Clifford torus T for either of the intersections between surfaces P and Q paired by a Whitney disc D' intersects D' exactly once, and any meridian (respectively, longitude) of T bounds a disc intersecting P (respectively, Q) exactly once, namely a meridional disc for P (respectively, Q). See Section 15.1 for more details.

A surface equipped with immersed discs bounded by a symplectic basis of curves for its first homology is called a *capped surface*, and the discs are called the *caps*. Capped surfaces have the following key property: they can be transformed into (immersed) spheres by cutting the base surface and gluing on parallel push-offs of their caps to the base surface, as indicated by Figure 2.1. This process is called *contraction*. Since each of the discs is used twice, provided the base surfaces are mutually disjoint, the pairwise intersection and self-intersection numbers of the family of spheres produced by contraction of a collection of capped surfaces are all zero. Moreover, if a surface S intersects a cap of a capped surface, we can perform a regular homotopy of S to ensure that it no longer intersects the sphere produced by contraction. This is called a *push-off* operation. However, if surfaces S and S' intersect dual caps of a capped surface, then after contracting the capped surface and pushing off, the new versions of S and S' will intersect in two (algebraically cancelling) points. See Section 15.2.5 for more details about these operations.

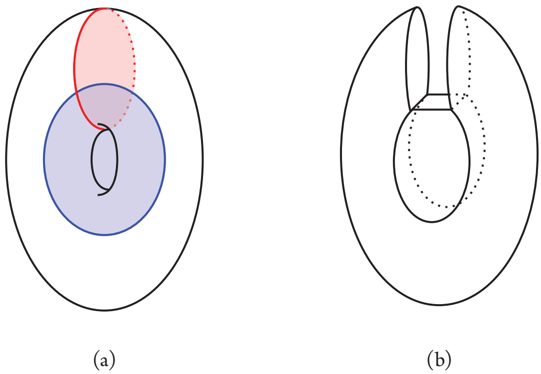

Figure 2.1 (Symmetric) contraction of a capped surface to a sphere. Here we show the situation for embedded caps. Left: A torus with a dual pair of caps. Right: The result of contraction along the pictured caps.

Step 2 (Modify the Whitney discs until they are equipped with transverse capped surfaces, Proposition 16.1). For each Whitney disc D'_k, pick a Clifford torus at either of the two paired intersections. Choose a meridian and longitude of each Clifford torus. These curves are capped by meridional discs for the $\{f'_i\}$, each of which intersects the collection $\{f'_i\}$ exactly once. Tube each such intersection point into the spheres $\{g'_i\}$ to make the discs disjoint from $\{f'_i\}$. Let $\{\Sigma^c_k\}$ denote the resulting collection of Clifford tori, equipped with these caps.

Take a parallel copy of each element of $\{\Sigma^c_k\}$ and contract, then push off all intersections of the discs $\{D'_k\}$ with the caps of this parallel collection. This transforms $\{D'_k\}$ into a collection of immersed discs, $\{W_k\}$, with the same framed boundaries as the $\{D'_k\}$. Let $\{S_k\}$ be the collection of immersed spheres produced by the contraction. Note that this collection is geometrically dual to $\{W_k\}$ by construction. Tube any intersection of the caps of the collection $\{\Sigma^c_k\}$ with $\{W_k\}$ into the collection $\{S_k\}$. We still call the resulting capped surfaces $\{\Sigma^c_k\}$.

Now we summarize the current situation. We have replaced the hypothesized collection of immersed discs $\{f_i\}$ with a collection $\{f'_i\}$ with the same framed boundaries, whose intersections and self-intersections are paired by a collection of framed, immersed Whitney discs $\{W_k\}$ equipped with a collection of geometrically dual capped surfaces $\{\Sigma^c_k\}$. That is, $\Sigma^c_\ell \cap W_k$ is empty whenever $\ell \neq k$ and consists of a single (transverse) intersection in the base surface of Σ^c_ℓ when $\ell = k$. We have also arranged that the capped surfaces $\{\Sigma^c_k\}$ and the interiors of $\{W_k\}$ lie in the complement of $\{f'_i\}$. Moreover, since the caps of $\{\Sigma^c_k\}$ were produced from embedded discs by tubing into parallel copies of the spheres $\{g'_i\}$ and the spheres $\{S_k\}$ were produced by contraction, they have trivial intersection and self-intersection numbers. Additionally, since they were produced from Clifford tori, we know that the tori $\{\Sigma_k\}$ (not including the caps) lie in a regular neighbourhood of the $\{f'_i\}$. The discs $\{f'_i\}$ are also equipped with a collection of geometrically dual spheres $\{g'_i\}$. The details of the construction so far are given in Proposition 16.1 and summarized in Figure 2.2.

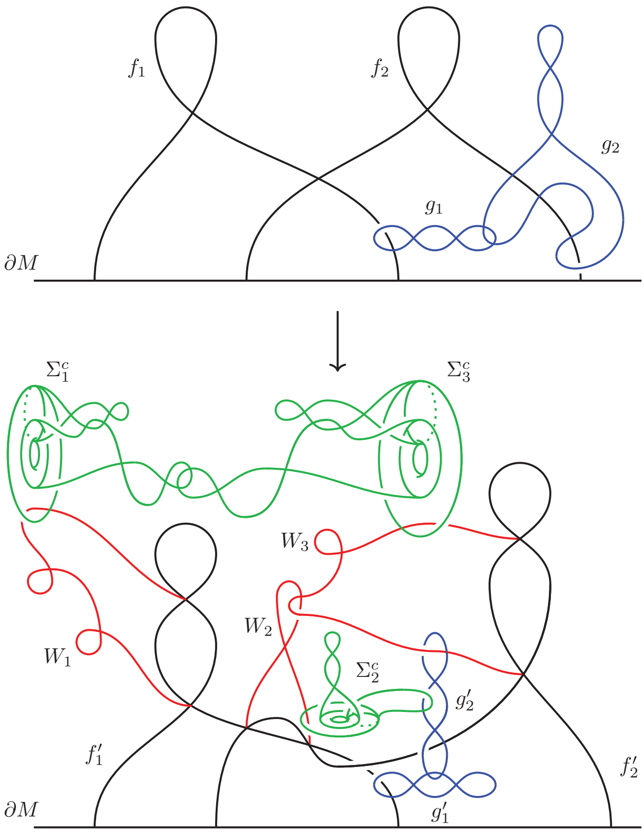

Figure 2.2 Summary of Proposition 16.1. The discs $\{f_i\}$ and $\{f_i'\}$ are in black, the spheres $\{g_i\}$ and $\{g_i'\}$ are in blue, the Whitney discs $\{W_k\}$ are in red, and the transverse capped surfaces $\{\Sigma_k^c\}$ are in green.

Due to the existence of $\{g_i'\}$, we know that $M \smallsetminus \bigcup \nu f_i'$ also has good fundamental group, and the latter will be our ambient manifold from now on. We will use the spheres $\{g_i'\}$ again, but set them aside for now. We will work for a while with the sets $\{W_k\}$ and $\{\Sigma_k^c\}$.

Step 3 (Promote the Whitney discs to capped surfaces, Proposition 16.2). First, we eliminate the intersections and self-intersections of $\{W_k\}$ by tubing them into parallel copies of the transverse capped surfaces $\{\Sigma_k^c\}$. This transforms them into capped surfaces $\{W_k'\}$ with the same framed boundary as $\{W_k\}$. Due to the trivial intersection and self-intersection numbers of the caps of $\{\Sigma_k^c\}$, we can make the $\{W_k'\}$ and the caps of $\{\Sigma_k^c\}$ disjoint, so that the families $\{W_k'\}$ and $\{\Sigma_k^c\}$ remain geometrically transverse. This requires creating another set of geometrically dual spheres by contraction, as before, and tubing into them. The caps of $\{W_k'\}$ and $\{\Sigma_k^c\}$ may change by a regular homotopy in this process, but we keep the same notation. The fact that the bodies of the $\{\Sigma_k^c\}$ lie in a neighbourhood of $\{f_i'\}$